Electronic Commerce Research and Applications 32 (2018) 23–36

Contents lists available at ScienceDirect

Electronic Commerce Research and Applications

journal homepage: www.elsevier.com/locate/elerap

# Loan evaluation in P2P lending based on Random Forest optimized by genetic algorithm with profit score

Xin Ye, Lu-an Dong, Da Ma

*Faculty of Management and Economics, Dalian University of Technology, Dalian 116024, China*

ARTICLE INFO

*Keywords:*
P2P lending
Loan evaluation
Random Forest
Genetic algorithm
Profit score

ABSTRACT

Loan evaluation is an effective method for credit risk assessment in peer-to-peer (P2P) lending and significantly affects lender investment decisions as well as his/her profits. Besides traditional methods of loan evaluation, machine learning has gained increased attention and has achieved better performance for P2P lending, especially regarding the Random Forest approach. However, the loan evaluation model based on Random Forest aims to improve the overall accuracy, which cannot guarantee that the lender profit is maximized when the overall accuracy is maximized because the profits of each loan are different. To further improve the loan evaluation effect and lender profits, Random Forest optimized using a genetic algorithm with profit score (RFoGAPS) is proposed. First, considering the actual and potential returns and losses, a new profit score is proposed and taken as the optimization objective. Second, the genetic algorithm is used to optimize the combination of decision trees in Random Forest. Then, the dataset of Lending Club is used to evaluate the proposed method. Experimental results show that the RFoGAPS can obtain higher profits for lenders compared with actual profit and traditional methods. Some suggestions are proposed based on experimental results to facilitate the healthy development of P2P lending.

## 1. Introduction

Peer-to-peer (P2P) lending occurs at the intersection of the sharing economy and e-commerce, and has developed into an important form of e-commerce application (Gao et al., 2018). P2P lending is an unsecured financing model between individuals. In the P2P lending process, lenders and borrowers establish credit relationships and complete the transaction procedures through an online platform without financial intermediaries, such as banks. The main features of P2P lending are flexible trading, low access thresholds and short loan periods (Guo et al., 2016; Ma and Wang, 2016; Ma et al., 2017). Both lenders and borrowers can make greater profits through P2P lending than traditional loans provided by financial intermediaries. P2P lending has gradually become one of the major forms of loans for individuals and start-ups. Lending Club, the largest P2P lending platform in the world, has contributed to 1.4 million cases of P2P loans, and the amount of loans has reached 20 billion USD. Lending Club provides lower interest rate loans through a fast and easy online or mobile interface for American borrowers. However, in the developing process of P2P lending, there are some problems, such as the asymmetric information and the imperfect risk control measure (Lei, 2016; Serrano-Cinca et al., 2015). These problems increase the possibility of default and cause a high default rate in P2P lending platforms. Loan defaults are harmful to the profits of lenders and the development of P2P lending platforms. Loan evaluation is an effective tool for credit risk assessment to reduce the rate of loan defaults. The more objective the credit risk evaluation is, the more sensible are the investment decisions that guide lenders, which helps to gain more profits.

With the development of P2P lending, scholars and P2P lending platforms use scorecard and machine learning methods to construct the loan evaluation model. These models are useful to identify defaulters in practical applications of P2P lending. Scholars have found that the performance of Random Forest used in loan evaluation is better than other methods in P2P lending (Malekipirbazari and Aksakalli, 2015). However, these methods are more concerned about maximizing their accuracy or reducing the error rate, which ignores the role of lender profits for model evaluation. To improve the profits of lenders within P2P lending, a new profit score is proposed and used as new model evaluation criteria. The difference between the profit score and actual profit is described as follows. The actual profit only considers the returns and losses of instances that are classified as non-defaulters and ignores the returns and losses of the remaining instances. While for the lender, when a non-defaulter is classified as a defaulter, the lender will lose the returns from the loan, which can be seen as potential losses. Similarly, the correct identification of a defaulter can help the lender avoid losing principal, which can be seen as a potential return. Thus, the profit score takes the potential returns and losses as part of the model evaluation criteria in addition to considering the actual returns and losses. Therefore, the new profit score can better measure and evaluate the effects of the loan evaluation model than the actual profit.

https://doi.org/10.1016/j.elerap.2018.10.004
Received 26 March 2018; Received in revised form 13 July 2018; Accepted 13 October 2018
Available online 16 October 2018
1567-4223/ © 2018 Elsevier B.V. All rights reserved.

Furthermore, an effective improved Random Forest algorithm for loan evaluation is proposed. In the proposed method, a genetic algorithm (GA) is used to optimize the combination of decision trees in a parameter-optimized Random Forest with the objective of maximizing the profit score. The method is called the Random Forest optimized by genetic algorithm with profit score (RFoGAPS). The experimental results on the Lending Club data between January 2014 and December 2016 indicate that the proposed method can help lenders make more profits over the traditional Random Forest. Our research has an important role in improving the performance of loan evaluation in P2P lending and further facilitating the promotion and healthy development of P2P lending.

The rest of this paper is organized as follows. Section 2 provides a literature review of the work that has been conducted on loan evaluation in P2P lending and the improvement of Random Forest. In Section 3, the framework of the RFoGAPS, profit score and GA operations are proposed. Then, experiments are conducted and the results are analyzed in Section 4. Finally, Section 5 concludes the paper.

## 2. Related works

In this section, a literature review is presented from two following aspects. Section 2.1 outlines the loan evaluation in P2P lending, and Section 2.2 discusses the improvement of the Random Forest.

### 2.1. Loan evaluation in P2P lending

Studies and applications of loan evaluation in P2P lending are mainly divided into two primary directions. One is to use a credit scorecard to evaluate the credit risk of the loan, and the other is to transform the loan evaluation into binary-classification problems.

A credit scorecard is the traditional loan evaluation method. Most of the P2P lending platforms build credit scorecards based on their own business needs, e.g., the LC score (Lending Club) and FICO (Ortega and Bell, 2008). A credit scorecard can give each loan a credit score simply and quickly, but it is not a good method to distinguish between non-defaulters and defaulters (Malekipirbazari and Aksakalli, 2015).

To improve the identification accuracy for defaults, P2P lending platforms use machine learning methods to predict whether the loan is overdue, e.g., Logistic Regression (Guo et al., 2016; Savvopoulos, 2010; Wiginton, 1980), decision tree (Feldman and Gross, 2005; Zhang et al., 2016), neural networks (Malhotra and Malhotra, 2002; Bekhet and Eletter, 2014; West, 2000), SVM (support vector machine) (Huang et al., 2004; Moro et al., 2014), Random Forest (Jin and Zhu, 2015; Wang et al., 2018), etc. Scholars have compared the performance of different machine learning methods used on loan evaluation in P2P lending. For example, SVM outperforms the neural network methods in Huang et al. (2004). Malekipirbazari and Aksakalli (2015) indicate that Random Forest has better performance than SVM, KNN (k-Nearest Neighbor) and Logistic Regression based on data from the popular lending platform Lending Club.

The ultimate purpose of loan evaluation is to help lenders make more profits. Eisenbeis and Robert (1977) point out that traditional credit risk evaluation models pay more attention to the minimization of default rates. However, P2P lending platforms should build new model evaluation criteria to maximize the lenders profits rather than minimize the default rates (Thomas, 2000). Lessmann et al. (2015) benchmark state-of-the-art algorithms for both credit scoring and profit scoring. These scholars prove that the most accurate classifier is not always profitable. So et al. (2014) propose a profitability scoring model that is more accurate in estimating the profitability of potential applicants than the standard method in predicting defaults. Verbraken et al. (2014) tailor a new profit-based classification performance measure to the loan evaluation model. These performance measures are all based on the Expected Maximum Profit (EMP) measure and show a better ability to choose a high-profit loan evaluation model. In addition, Xia et al. (2017) incorporate cost-sensitive learning and extreme gradient boosting (XGBoost) to propose a cost-sensitive boosted tree loan evaluation model to enhance the capability of discriminating potential default borrowers.

Through the literature review of loan evaluation in P2P lending, it can be found that Random Forest has a better performance in loan evaluation than other methods, and the profit score is more beneficial for selecting the high-profit loan evaluation model than traditional model evaluation criteria, such as accuracy, AUC, etc. However, there are two shortcomings in the existing research. First, scholars pay more attention to the actual returns and losses of the loan that are classified as non-defaulters and ignore the effects of potential returns and losses of the instance, which is classified as a defaulter. Second, scholars pay more attention to the comparisons between different traditional algorithms but ignore the significant effect that the algorithm improvement may bring to the loan evaluation in P2P lending.

### 2.2. Improvement of Random Forest

Some scholars believe that optimizing key parameters of the Random Forest can achieve higher classification accuracy within the acceptable operating efficiency (Rodriguez-Galiano et al., 2012). In the field of loan evaluation in P2P lending, Malekipirbazari and Aksakalli (2015) investigate the effects of the forest scale, the number of split features and the maximum tree depth on the performance of Random Forest. In the field of computer science, Huang and Boutros (2016) examine the effects of parameter selection on the classification performance using Random Forest on different datasets and find that the parameters are highly correlated with its accuracy and that the optimized parameters are different for different datasets. Although the parameter optimization can improve the performance of Random Forest, the improvement is not very significant because Random Forest is not very sensitive to the choice of parameters (Wang et al., 2017).

Some scholars find that trees in the traditional Random Forest have different contributions towards global accuracy. Some of them may amplify wrong predictions to downgrade the overall predictive performance of the forest (Adnan and Islam, 2016). Zhou et al. (2002) propose a selection strategy to choose a small number of high-quality individual learners to be a new ensemble with a higher performance. Similarly, some scholars propose a greedy algorithm to improve the traditional Random Forest. Lu et al. (2010) sort the trees according to their accuracy and choose the user-defined number of trees according to the order. Martínez-Muñoz and Suárez (2004) use a general climbing strategy and then add or delete a single tree to increase the accuracy or diversity of the forest. However, using the greedy algorithm to optimize the Random Forest readily leads to becoming trapped at local optima. To overcome this problem, some scholars use the GA to optimize the Random Forest where the accuracy is the objective of the optimization in other fields, e.g., cancer prediction, computer science, etc. (Adnan and Islam, 2016; Kim and Oh, 2008). The GA is more likely to obtain the global optimal Random Forest than the greedy algorithm. So the GA improved Random Forest (GA-RF) has a better performance than the traditional Random Forest (Adnan and Islam, 2016). Based on the above research, it is necessary to introduce GA-RF into the loan evaluation in P2P lending. However, the objective of GA-RF is to optimize the accuracy of Random Forest in most existing studies. For the loan evaluation in P2P lending, the high accuracy does not represent a high profit, so GA-RF with the objective of accuracy does not ensure maximizing lender profits. Therefore, a new objective function needs to be defined based on the characteristics of the P2P lending loan evaluation to optimize the Random Forest via the GA.

## 3. Proposed methodology

As mentioned above, compared with other methods, Random Forest has better performance in the loan evaluation in P2P lending

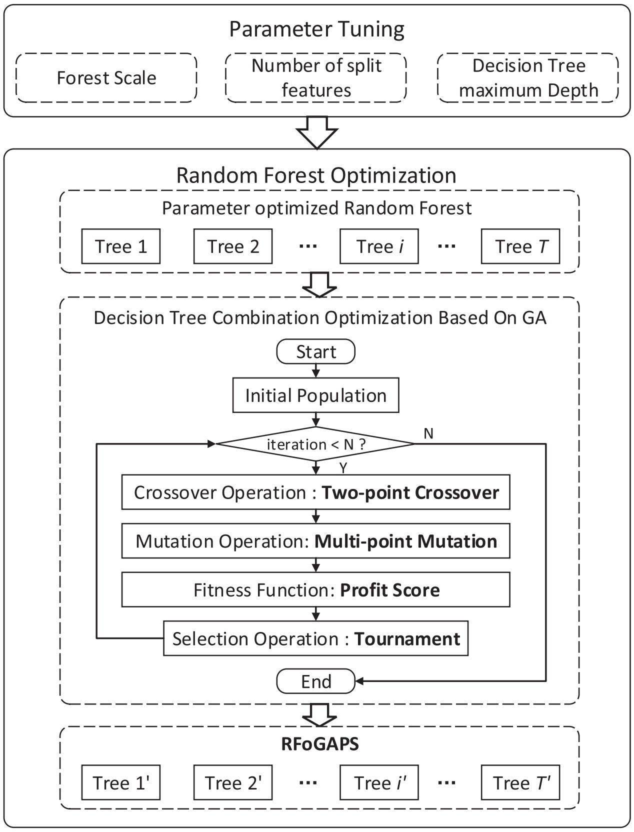



**Fig. 1.** The framework of the RFoGAPS.

(Malekipirbazari and Aksakalli, 2015). However, the objective of the loan evaluation model based on Random Forest is to minimize the overall overdue rate rather than to maximize the lender profits. Although reducing the overall overdue rate can increase the lender's profit, it cannot be guaranteed that the profit is maximized when the overall overdue rate of the loan evaluation model is minimal. Improving the lender profits becomes increasingly valuable under the premise of controlling risk. In addition, some scholars find that decision trees in the traditional Random Forest have different contributions to the global performance and that the optimal combination of decision trees may improve the global performance mentioned in Section 2. Therefore, to further improve the performance of the loan evaluation and the lender profits, the new Random Forest optimized by genetic algorithm with profit score (RFoGAPS) is proposed in this section.

### 3.1. Framework of the Random Forest optimized by genetic algorithm with profit score (RFoGAPS)

The framework of RFoGAPS is shown in Fig. 1. The method mainly consists of two parts: parameter tuning and Random Forest optimization. The parameter tuning mainly identifies the optimal parameters of the Random Forest, such as the forest scale, the number of split features and the maximum decision tree depth. Then, Random Forest optimization is conducted using the GA to optimize the combination of decision trees in the parameter-optimized Random Forest with the objective of profit score maximization considering both actual and potential returns and losses.

### 3.2. Profit score

Loan evaluation is considered as a binary classification problem (Verbraken et al., 2014), where the defaulters and non-defaulters are assigned as class 0 and class 1, respectively. Every instance is classified as one of the two classes according to their characteristics. The outcome of the classifier can be summarized in a confusion matrix, as shown in Table 1.

**Table 1**
Confusion matrix with losses and returns.

| Actual label | Predicted label | |
|---|---|---|
| | Class 0 (defaulter) | Class 1 (non-defaulter) |
| Class 0 (defaulter) | $c_k(0 \mid 0)$ | $c_k(1 \mid 0)$ |
| Class 1 (non-defaulter) | $c_k(0 \mid 1)$ | $c_k(1 \mid 1)$ |

The diagonal of the confused matrix represents the correct predictions, while the off-diagonal indicates instances that are misclassified. $c_k(i|j)$ represents the losses and benefits of the $k$-th instance when the actual label of the instance is class $j$ and the classifier determines it as class $i$ (with $i, j \in \{0, 1\}$), as described as Eq. (1). In the calculation of the losses and returns, the fixed costs are ignored, such as the transaction cost, etc. When the classifier correctly predicts an instance where its actual label is defaulter, the classifier avoids the loss of principal $M$ for the lender. This situation can be treated as a potential return and denoted as $c_k(0|0) = M_k$. In contrast, when the classifier makes a wrong prediction about the defaulter, the lender losses the principal $M_k$ and $c_k(1|0) = -M_k$ is an actual loss. When the classifier makes predictions about an instance that its actual label is a non-defaulter, the correct prediction will give the lender a profit $c_k(1|1) = M_k \times a_k$, which is determined by the principal $M_k$ and interest rate $a_k$. This situation is regarded as an actual return, while the wrong prediction of a non-defaulter makes lender loss the potential return $c_k(0|1) = -M_k \times a_k$.

$$c_k(i|j) = \begin{cases} M_k & actual\ label = class\ 0, \quad predicted\ label\ class\ 0 \\ -M_k & actual\ label = class\ 0, \quad predicted\ label\ class\ 1 \\ M_k \times a_k & actual\ label = class\ 1, \quad predicted\ label\ class\ 1 \\ M_k \times a_k & actual\ label = class\ 1, \quad predicted\ label\ class\ 0 \end{cases} \tag{1}$$

The proposed profit score is the sum of all the profits for all instances, which is calculated as follows.

$$profit\ score = \sum_{k=1}^{N} c_k(i|j) \quad i, j \in \{0, 1\} \tag{2}$$

where $N$ represents the number of instances. When the accuracy of the classifier is 100%, $c_k(1|0)$ and $c_k(0|1)$ will not appear, and the profit score is equal to the actual profit.

### 3.3. Problem description of Random Forest optimization

Random Forest is an ensemble of $T$ decision trees and is usually constructed by building a large number of decision trees to achieve a better performance. Random Forest is described as in Eq. (3).

$$f_{RF}(x_k) = arg \max_{\Phi} \sum_{t=1}^{T} \omega_t h_t(x_k)$$
$$h_t(x_k) = \begin{cases} 0 & instance\ x_k\ is\ classified\ as\ defaulter \\ 1 & instance\ x_k\ is\ classified\ as\ non\text{-}defaulter \end{cases} \tag{3}$$

where $x_k$ is a feature vector of the $k$-th instance, $f_{RF}(x_k)$ and $h_t(x_k)$ represent the prediction results of $x_k$ using Random Forest and the $t$-th decision tree in the Random Forest, respectively, and $\omega_t$ is the weight of the $t$-th decision tree in the Random Forest and has a value of 1 in the traditional Random Forest. $\Phi = \{0, 1\}$ is a discrete class label. In this paper, the voting mechanism of the Random Forest adopts the absolute majority voting method.

Thus, a large number of decision trees cause significant memory consumption and computational complexity. Generally, some trees in the Random Forest have no contribution to improving its performance.

Random Forest Optimization aims to select some trees, which have more contributions to the Random Forest, to compose a new sub Random Forest that has better performance, lower memory consumption and smaller computational complexity than the original Random Forest. The problem of Random Forest Optimization is an NP-hard problem (Adnan and Islam, 2016; Kim and Oh, 2008), and it is described as Eq. (4).

$$\max \sum_{k=1}^{N} c_k(f_{RF}(x_k)|j)$$
$$\text{s. t.}\quad 1 \leqslant \sum_{t=1}^{T} \omega_t \leqslant T$$
$$\omega_t \in \{0,1\} \tag{4}$$

To maximize the lender earnings, maximizing the profit score is adopted as the objective in this problem. Unlike traditional Random Forests, the Random Forest optimization allows the decision tree weight $\omega_t = 0$, which indicates that the corresponding decision tree is removed from the original Random Forest. The remaining decision trees with weight $\omega_t = 1$ in the original Random Forest constitute the new sub Random Forest.

### 3.4. Genetic algorithm (GA)

The GA simulates the process of biological genetic evolution, and it is an efficient method to solve NP-hard problems (Holland, 1992). Each chromosome represents a potential solution. The fitness function is used to evaluate each chromosome to determine the next parent, and then crossover and mutation are used to generate the next generation population. The chromosome with the best solution is reported as the output of the GA. Fig. 1 presents an outline of the GA used in this paper. The chromosome encoding, fitness function and genetic operations (crossover, mutation and selection) are described as follows.

#### 3.4.1. Chromosome encoding

A chromosome represents the potential combination of decision trees in the parameter optimized Random Forest. The Random Forest is considered as a decision tree pool that contains $T$ decision trees ($T$ is the Random Forest scale). In detail, a string with length $T$ is used to represent the Random Forest, and the $t$-th digits of the chromosome represent the $t$-th decision tree in the Random Forest. The value of each digit is 1 or 0, meaning that the corresponding decision tree is selected or removed. For example, a potential combination of decision trees in a Random Forest, which has $T$ decision trees, can be as shown in Fig. 2.

| Decision tree No. | 1 | 2 | 3 | ... | $T$-2 | $T$-1 | $T$ |
|---|---|---|---|---|---|---|---|
| Chromosome | 0 | 1 | 1 | ... | 1 | 0 | 1 |

**Fig. 2.** Chromosome encoding.

#### 3.4.2. Fitness function

The choice of fitness function is very important to the GA, which is related to the convergence speed of the algorithm and the performance of the solution. The purpose of this study is to find a high-profit Random Forest; hence, the fitness function is designed as the profit score (Eq. (2)). The greater the fitness value of the chromosome, the higher are the profits of the corresponding RF for the lender.

#### 3.4.3. Crossover operation

Crossover and mutation are important components of the GA, which are the two basic operations of generating offspring (Mundim et al., 2017). Two-point crossover is adopted in this paper, as shown in Fig. 3. For example, a pair of chromosomes $Cr_1$ and $Cr_2$ is selected from parents, and crossover points 1 and 2 are generated randomly. The two crossover points divide $Cr_1$ and $Cr_2$ into three parts. During the crossover operation, the middle part of the $Cr_1$ genes is interchanged with the intermediate part of the $Cr_2$ genes. After the crossover operation, two new offspring can be obtained. To increase the possibility of diversity of individuals, every parent chromosome performs a crossover operation.

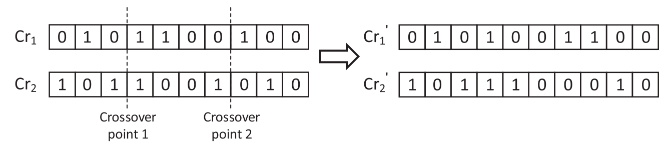



**Fig. 3.** Crossover operation.

#### 3.4.4. Mutation operation

The mutation operation can increase the randomness of the solution and avoid falling into a local optimum (Metawa et al., 2017). The individual is chosen for a mutation operation with a certain probability. Multi-point mutation is used as the mutation operator, as shown in Fig. 4. In the process of mutation, each digit of the chromosome is changed with a certain probability, meaning some decision trees are removed or selected. The mutation operation can cause slight randomness in the search direction and facilitate the algorithm convergence.

#### 3.4.5. Selection operation

Selection operation is the process of selecting the individual with a high fitness from the offspring and parent to generate the next parents. Tournament selection is adopted in this paper, where two individuals in the population are randomly selected for comparison, and the individuals with the best fitness are selected as the parent for the next generation. Tournament selection also enables individuals with a better fitness to have greater probabilities of survival.

## 4. Experimental results and analysis

### 4.1. Dataset and feature

The dataset contains 110 K borrow records with borrower information from the Lending Club[1] data between January 2014 and December 2016. The dataset contains 7 loan statuses including "Current", "Fully paid", "Default", "Charged off", "In Grace Period", "Late (16–30 days)" and "Late (31–120 days)". These loan statuses can be divided into four categories: Current loan, Fully paid loan, Default loan (including "Default" and "Charged off") and Need Attention loan (including "In Grace Period", "Late (16–30 days)" and "Late (31–120 days)") (Jin and Zhu, 2015). In this research, the loan records only belonging to Fully paid loan and Default loan are selected, because the actual and potential returns and losses of these records are stable and computable. After filtering the data, approximately 36 K borrow records are used in the experiment.

Many scholars have used the Lending Club dataset to study loan evaluation, and some features and pretreatment methods are used in the research. According to the reference (Malekipirbazari and Aksakalli, 2015), a total of 17 primitive features are selected. Among them, 14 features are numeric, and 3 features are nominal, including Home ownership (3 levels), Loan amount (14 levels) and Term (2 levels). Selected features and pretreatments are shown in Table 2. The nominal features are binarized, and a total of 3 + 14 + 2 = 19 binarized features are obtained. Finally, 33 features, including 14 numeric features and 19 binarized features, are used in the experiments.

[1] https://www.lendingclub.com/info/statistics.action

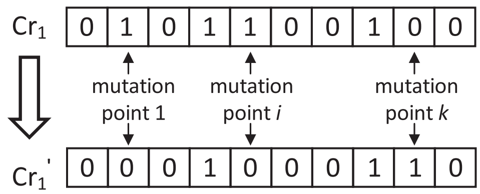



**Fig. 4.** Mutation operation.

### 4.2. Model evaluation and assessment

A 10-fold cross-validation is used to evaluate the generalization ability of the candidate classifier, because this can better evaluate the overall performance of the classifiers (Brown and Mues, 2012). In detail, the dataset is split into 10 equally sized and mutually exclusive subsets, taking 9 subsets for training and 1 subset for testing. The model is trained and tested 10 times, and the mean performance of the results for the 10-fold cross-validation is used to evaluate the classifier. In this paper, we mainly focus on the three performance evaluation criteria for classifier comparison of accuracy, AUC and profit score.

(1) *Accuracy* represents the overall classification accuracy rate of the testing dataset and is shown in Eq. (5).

$$accuracy = \frac{TN + TP}{TN + TP + FN + FP} \tag{5}$$

where *TP* is the number of true positives (non-defaulter predicted as non-defaulter), *TN* is true negatives (defaulter predicted as defaulter), *FP* is false positive (defaulter predicted as non-defaulter), and *FN* is false negatives (non-defaulter predicted as defaulter).

(2) *AUC* represents the usual area under the Receiving Operating Characteristic (ROC) curve in the testing dataset. The ROC is a plot of the proportion of the true positive rate (TPR) against the false positive rate (FPR), where TPR and FPR are the ordinate and abscissa of the ROC curve, respectively.

$$TPR = \frac{TP}{TP + FN}$$
$$FPR = \frac{FP}{TN + TP} \tag{6}$$

(3) *Profit score* is the profit of all instances in the testing dataset considering the actual and potential returns and losses. The profit score was defined in Section 3.2. Compared with the Actual profit, the profit score can more comprehensively measure the lender profits by the classifier. However, the Lending Club dataset does not provide the specific repayment information for the loan records that belong to the Default loan. Therefore, we assume that the borrowers do not pay any money for these loan records.

### 4.3. Cost sensitive analysis

When the dataset is imbalanced, the classifier focuses too much on the majority class, and the predictive accuracy of the minority class declines (Díez-Pastor et al., 2015). There are many methods to handle the problem of data imbalance, such as under-sampling, over-sampling, and cost sensitive analysis. Notably, some research notes that both over-sampling and under-sampling possess significant drawbacks compared to the cost sensitive analysis (Xia et al., 2017; Seiffert et al., 2010). In the experiment, the proportions of non-defaulters and defaulters are 0.76 and 0.24, respectively. The number of non-defaulters (majority class) is far more than the number of defaulters (minority class). To reduce the effect of the imbalanced dataset on the loan evaluation performance, a weighted cost matrix is used to increase the cost of misclassification associated with defaulters (Schebesch and Stecking, 2005; Malekipirbazari and Aksakalli, 2015). The misclassification of a defaulter (as a non-defaulter) is 5 times more costly than the misclassification of a non-defaulter (as a defaulter). Therefore, the classifier will pay more attention to defaulters that are falsely predicted, thus improving the predictive accuracy for defaulters.

### 4.4. Parameter tuning

The main purpose of this section is to analyze the influence of the Random Forest scale, the number of split features and the maximum decision tree depth on the performance of the Random Forest in loan evaluations. First, the scale of the Random Forest is optimized. As referenced in Breiman (2001), the number of split attributes is set to $\log_2(p)$, where $p$ represents the total number of features. In this case, the Random Forest scale increases from 1 to 500 with 10 increments each time. At the same time, the other parameters are set to default values. Figs. 5–7 show the Random Forest performances with respect to the Random Forest scale.

From the macroscopic trend, if the Random Forest scale is smaller than 100, the accuracy, AUC and profit score are improved with the increase in the Random Forest scale. However, if the scale is more than 100, the performance of Random Forest only shows small fluctuations. From the microscopic fluctuations of the three curves, when the highest

**Table 2**
The selected features and pretreatments.

| Feature name | Description and Pretreatment | Type |
|---|---|---|
| Annual income | The annual income provided by the borrower during registration. Pretreatment: the natural logarithm function. | Numeric |
| Credit age | Data of the earliest credit line opened by the borrower, which is converted to months. Pretreatment: the natural logarithm function. | Numeric |
| Delinquencies | The delinquency in the borrower's credit file for the past 2 years. Pretreatment: right-censor ⩾2, meaning values more than 2 were set to 2. | Numeric |
| Employment length | Employment length in years. Pretreatment: value is between 0 and 10 with 0 meaning less than 1 year and 10 meaning 10 or more years. | Numeric |
| Home ownership | The home ownership status provided by the borrower. Values include Rent, Own and Mortgage. Pretreatment: binary discretization. | Nominal |
| Inquiries | The number of inquiries in the past 6 months. Pretreatment: right-censor ⩾3. | Numeric |
| Loan amount | The listed amount of the loan applied for by the borrower. | Numeric |
| Loan purpose | A category provided by the borrower for the loan request. Values include Debt Consolidation, Credit Card, Major Purchase, Home Improvement, Moving, Renewable Energy, Vacation, Car, Medical, House, Small Business, Wedding, Educational and Other. Pretreatment: binary discretization. | Nominal |
| Open account | The number of open credit lines in the borrower's credit file. | Numeric |
| Total accounts | The total number of credit lines currently in the borrower's credit file. | Numeric |
| Term | The number of payments on the loan. Values are in months and can be either 36 or 60. Pretreatment: binary discretization. | Nominal |
| DTI | Ratio calculated using the borrower's total monthly debt payments divided by the borrower's monthly income. | Numeric |
| Income to payment ratio | Ratio of the borrower's monthly payment to the monthly income. Pretreatment: the natural logarithm function | Numeric |
| Revolving utilization ratio | The amount of credit the borrower is using relative to all available revolving credit. | Numeric |
| Income to revolving ratio | Ratio of the borrower's monthly income to the total credit revolving balance. Pretreatment: the natural logarithm function. | Numeric |
| Loan rate | Interest rate on the loan. This feature is not used in the model; only for the profit score. | Numeric |
| Loan status | Current status of the loan. Pretreatment: "Default" and "Charged Off" correspond to 0 and "Fully Paid" corresponds to 1. | Numeric |

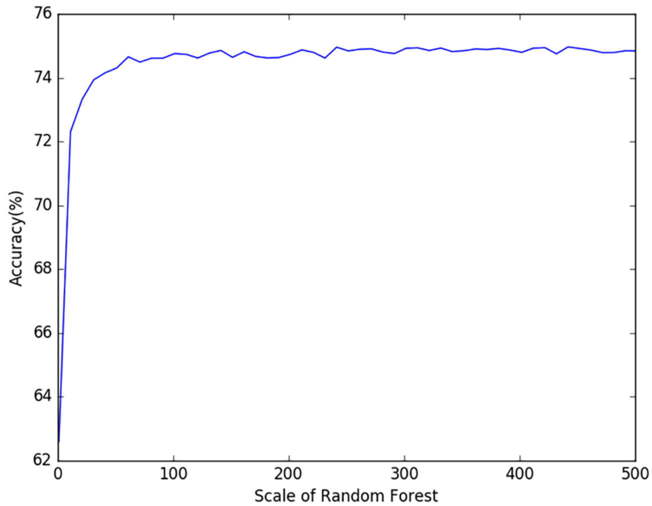



**Fig. 5.** Accuracy with respect to the Random Forest scale.

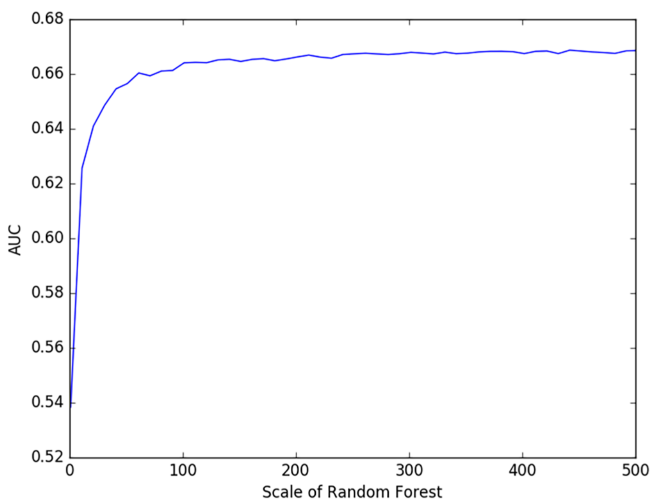



**Fig. 6.** AUC with respect to the Random Forest scale.

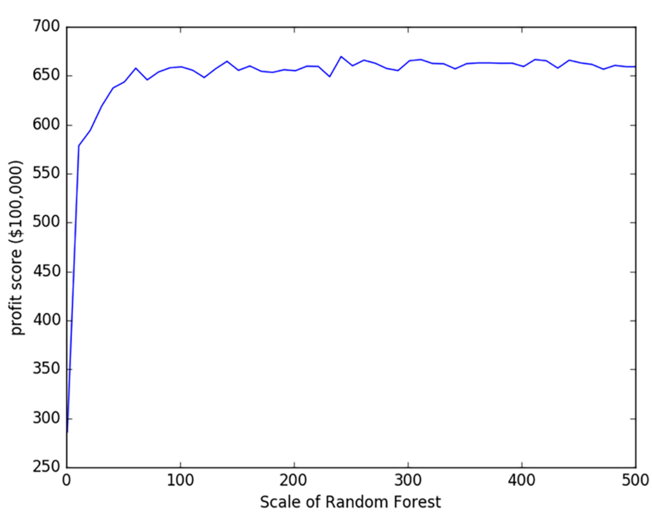



**Fig. 7.** Profit score with respect to the Random Forest scale.

value of the accuracy or AUC is achieved, the profit score did not reach the maximum, and the fluctuations among the curves are quite different. The experimental results show that the AUC and accuracy are maximum when the Random Forest scale is 440. In this case, the AUC is

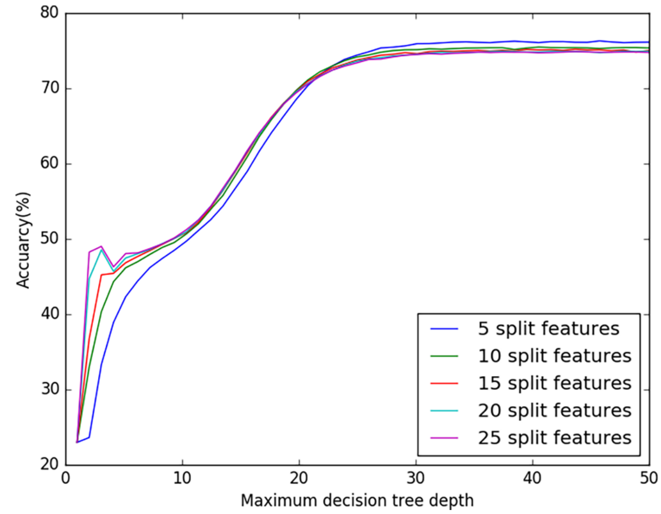



**Fig. 8.** Accuracy with respect to the number of split features and the maximum decision tree depth.

0.668, the accuracy rate is 74.92%, and the profit score is 666.48. It can be seen that the profit score does not reach its highest level, which is 669.55 in this case. However, when the Random Forest Scale is 240, the profit score reaches its maximum, while the accuracy and AUC of the classifier are only 74.95% and 0.667, respectively. Therefore, it is difficult to select a high-profit classifier using the accuracy and AUC, while the profit score can solve this problem properly. Thus, the profit score is used as the main criterion of the model selection in the following research, and 240 decision trees is the optimal scale of the Random Forest.

Then, the optimal number of split feathers and maximum decision tree depth are determined based on the results of the Random Forest scale tuning. The number of split features is set to 5, 10, 15, 20 and 25, and the maximum decision tree depth is set from 1 to 50. The experimental results are shown in Figs. 8–10.

The optimal number of split features is discussed first, where Figs. 8 and 10 show similar trends. The performance of the Random Forest with 5 split feathers is worse than the others when the decision tree maximum depth is less than approximately 25. However, with the increasing of the maximum decision tree depth, Random Forest with 5 split features has a higher accuracy and profit score than the others. In addition, Fig. 9 shows that the Random Forest with 5 split features

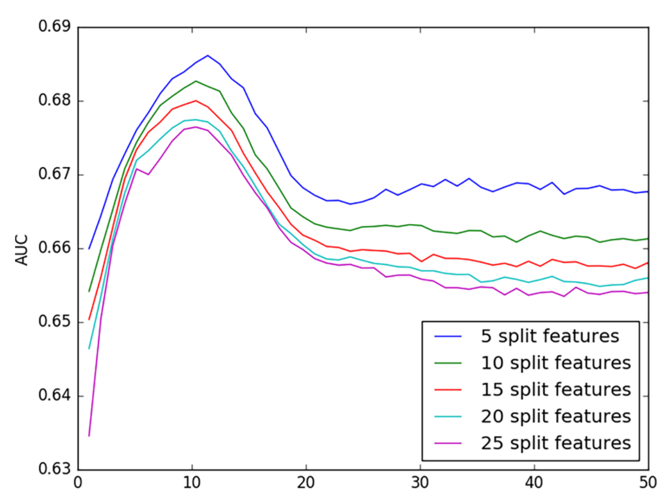



**Fig. 9.** AUC with respect to the number of split features and the maximum decision tree depth.

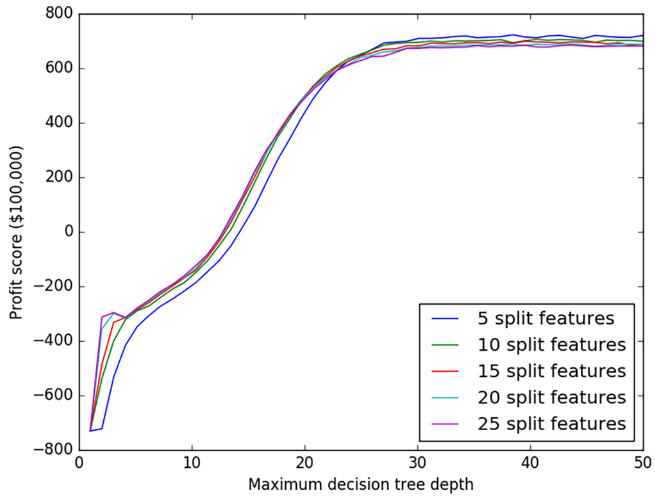



**Fig. 10.** Profit score with respect to the number of split features and the maximum decision tree depth.

always performs well for the AUC compared with the other split features. Therefore, the optimal number of split features is 5.

Next, the maximum decision tree depth is optimized. Comparing Figs. 9 with 10, when the maximum decision tree depth is near 10, Random Forest has the highest AUC, while the profit score of the classifier is not optimum. A Random Forest that can meet the optimal AUC and profit score at the same time is difficult to find. Comparing Figs. 8 with 10, the performance of the Random Forest improves as the maximum decision tree depth increases and eventually tends to converge. To obtain a high-profit Random Forest, the optimal value of the maximum decision tree depth is 37. Thus, via parameter tuning, the Random Forest with a scale of 240, 5 split features and maximum decision tree depth of 37 is selected.

### *4.5. Decision tree combination optimization based on GA*

In this section, the GA is used to optimize the combination of the decision tress in Random Forest after parameter tuning. The selection of GA parameters has an influence on the convergence speed and the performance of RFoGAPS (Kim and Oh, 2008; Li et al., 2016). The parameters of the GA are described in Table 3, which are chosen based on pilot experiments.

### *4.6. Experimental results*

To evaluate the RFoGAPS performance, six different approaches are conducted and compared as follows.

(1) *Actual profit* is the actual returns and losses that lenders can obtain. If the actual label of the loan is non-defaulter, the lender obtains the interest. If the actual label of the loan is defaulter, the lender loses the principal. Actual profit is the benchmark.
(2) *RF* is Random Forest with parameter tuning. The scale of RF is 240. In each tree, 5 split features are used, and splitting is continued until the tree reaches a depth of 37.
(3) *SVM* divides the data into two regions (one for each class) in the p-dimensional feature space via a hyperplane with the maximum margin width between instances of the two classes (Cortes and Vapnik, 1995). The SVM optimization problem can be described as in Eq. (7).

$$\min_{w,b} \frac{1}{2}||w||^2$$
$$\textbf{s. t.} \quad y_k(w^T x_k + b) \geq 1 \quad k = 1, 2, \ldots, m \tag{7}$$

where $x_k$ denotes the feature vector of the $k$-th instance, and $y_k$ represents the label of the $k$-th instance. The hyperparameters $w$ and $b$ determine the direction of the hyperplane and the distance from the hyperplane to the origin, respectively. In this study, the radial basis function (RBF) is specified as the kernel used in the SVM, which is chosen as a candidate method since it has been successfully used in Huang et al. (2004) and Moro et al. (2014).
(4) *DT* (decision tree) constructs a tree for classification by selecting the optimal partition feature. In this study, CART (Breiman et al., 1984) is selected, which applies the Gini index to select the optimal partition feature.
(5) *KNN* has been widely used in classification and it has good efficiency. In this study, the parameter $k$ of the KNN is set to 1 according to reference (Malekipirbazari and Aksakalli, 2015).
(6) *LR* is Logistic Regression and has been widely used in traditional loan evaluations for banks and P2P lending.

**Table 3**
Parameters of GA.

| Parameter | Chromosome length | Population size | Number of iterations |
|---|---|---|---|
| Value | 240 | 100 | 200 |
| Parameter | Mutation rate | Digits mutation rate | Crossover rate |
| Value | 0.1 | 0.1 | 1 |

Table 4 and Fig. 11 present the Actual profit and experimental results of the 10-fold cross-validation for six methods. Based on the average profit score of the 10-fold cross-validation, all methods are sorted by their profit scores: 787.20 (RFoGAPS) > 730.7 (Actual profit) > 717.31 (RF) > 583.64 (SVM) > 484.20 (KNN) > 381.44 (DT) > −79.05 (LR). Among them, RFoGAPS has the highest profit score at 787.20, and it is the only method that is higher than the Actual profit. The other methods are all lower than the Actual profit, especially LR. The reason for the low profit score of LR is that it is unable to effectively distinguish a non-defaulter and defaulter. This conclusion is consistent with the experimental result that the accuracy of the Logistic Regression is the lowest in reference (Malekipirbazari and Aksakalli, 2015). Compared with the Actual profit, the profit score of RFoGAPS increases by 7.73%. This shows that the RFoGAPS performs the best and can help lenders obtain 5.65 million additional returns than the Actual profit in years 2014–2016. In addition, from the perspective of the profit score in each fold, RF and RFoGAPS have higher profit scores than the Actual profit in folds 1–8. However, in folds 9 and 10, the profit score of RF shows a decline, which causes it to preform worse than Actual profit, which is more prominent in fold 9. The significant decline of profit score also appeared in DT and RFoGAPS. The analysis results of the dataset of folds 9 and 10 show that the borrowing times of the records are mainly in 2016. The main reason for this phenomenon is the differences in the feature distributions between the 2016 dataset and the 2014–2015 dataset. So the classifier trained by the 2014–2015 dataset cannot effectively classify the 2016 dataset. Therefore, the descriptive statistical analysis of the numerical features using the datasets in 2014–2015 and 2016 is analyzed, as shown in Table 5.

In the descriptive statistical analysis results, the maximum loan amount increased from 35,000 to 40,000 in 2016. The standard deviation of the loan amount, DTI, income to revolving ratio and loan rate in 2016 are higher than in 2014–2015. In addition, the nominal feature also has a significant difference, of which the most significant difference is in home ownership, as shown in Fig. 12. For home ownership, the proportion of borrowers that rent declined, while the proportion of borrowers who own increased in 2016. Furthermore, the proportion of non-defaulters in the 2016 record is higher than in 2014–2015.

There are two main reasons for the change in the feature distribution as follows. (1) The change in the loan amount shows that the Lending Club adjusted their loan policy at the beginning of 2016, which affects the feature distribution of the records. (2) Another important

**Table 4**
Compared experimental results (2014–2016).

| Fold | Profit Score ($100,000) | | | | | | | Improvement rate compared with actual profit (%) | | | | | | Scale of RFoGAPS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Actual profit | RF | RFoGAPS | SVM | DT | KNN | LR | RF | RFoGAPS | SVM | DT | KNN | LR | |
| 1 | 694.56 | 722.92 | **764.33** | 533.31 | 408.59 | 481.10 | −26.64 | 4.08 | **10.05** | −23.22 | −41.17 | −30.73 | −103.84 | 134 |
| 2 | 747.12 | 764.67 | **822.50** | 641.42 | 427.18 | 511.16 | −68.90 | 2.35 | **10.09** | −14.15 | −42.82 | −31.58 | −109.22 | 109 |
| 3 | 726.49 | 768.23 | **811.30** | 557.76 | 500.19 | 492.88 | −10.68 | 5.75 | **11.67** | −23.22 | −31.15 | −32.16 | −101.47 | 101 |
| 4 | 643.16 | 707.45 | **753.55** | 593.25 | 436.23 | 475.51 | 67.47 | 10.00 | **17.16** | −7.76 | −32.17 | −26.07 | −89.51 | 93 |
| 5 | 660.31 | 738.85 | **797.07** | 651.70 | 428.69 | 480.48 | 34.27 | 11.89 | **20.71** | −1.30 | −35.08 | −27.23 | −94.81 | 121 |
| 6 | 751.90 | 826.65 | **888.02** | 535.34 | 488.18 | 516.26 | −79.92 | 9.94 | **18.10** | −28.80 | −35.07 | −31.34 | −110.63 | 123 |
| 7 | 715.62 | 757.73 | **823.16** | 530.40 | 464.02 | 458.85 | −90.76 | 5.88 | **15.03** | −25.88 | −35.16 | −35.88 | −112.68 | 119 |
| 8 | 725.50 | 743.34 | **806.84** | 463.69 | 407.88 | 475.90 | −110.09 | 2.46 | **11.21** | −36.09 | −43.78 | −34.40 | −115.17 | 129 |
| 9 | 810.89 | 321.71 | 550.45 | **644.43** | −161.24 | 492.81 | −171.01 | −60.33 | −32.12 | **−20.53** | −119.88 | −39.23 | −121.09 | 101 |
| 10 | 831.50 | 821.56 | **854.74** | 685.09 | 414.72 | 457.10 | −334.29 | −1.20 | **2.79** | −17.61 | −50.12 | −45.03 | −140.20 | 115 |
| Avg. | 730.70 | 717.31 | **787.20** | 583.64 | 381.44 | 484.20 | −79.05 | −1.83 | **7.73** | −20.13 | −47.80 | −33.73 | −110.82 | 114.5 |

reason is that most of the loans that started in 2016 are still in progress, and only a few of them are completed. However, this portion of the records may be special, and their feature distribution may be different from loans that started between 2014 and 2015. The approach that uses the classifier trained by the 2014–2015 dataset to classify the 2016 dataset is unreasonable. Therefore, the proposed method is retested using the 2016 dataset, as shown in Table 6 and Fig. 13.

The experimental results show that the average profit scores of RF and RFoGAPS are 205.59 and 219.51, respectively, which have better performances than the Actual profit at 202.07. In contrast, SVM, DT, KNN and LR have a lower profit score than the Actual profit. Compared with the Actual profit, the average profit score of RFoGAPS shows a significant improvement of 8.63%, while RF only increases by 1.74%. It also can be seen that RFoGAPS remains valid for datasets with characteristic distribution changes. In addition, through comparing the RF scale, the model complexity of RF is approximately twice as high as that of RFoGAPS.

In summary, comparing the Actual profit and the other five methods, the proposed RFoGAPS can obtain the highest profits for the lenders, and the model has lower complexity and higher efficiency.

### 4.7. Real example

Furthermore, for a clearer comparison of the proposed algorithm with the Random Forest, 35 loan records from the 2016 dataset are randomly selected. The details of the 35 loan records are described in the Appendix A. These loan records are predicted by the RFoGAPS and Random Forest, and the predictive results are shown in Table 7.

The experimental results show that the predictive accuracy of Random Forest is 77.14% (27/35). It is higher than RFoGAPS at 71.43% (25/35). However, the actual profit obtained from RFoGAPS for the lender is $59,205.87, which is much higher than the actual profit obtained by Random Forest ($27,125.325). It can be seen that a high predictive accuracy does not necessarily lead to a higher profit. Compared with Random Forest, RFoGAPS can obtain higher profits for lenders.

### 4.8. Discussion and suggestions

Based on the experimental results, some suggestions are proposed for P2P lending platforms as follows.

Firstly, a profit score that considers both the actual and potential returns and losses is better as an evaluation criterion for the loan evaluation model in P2P lending compared with traditional evaluation criteria, such as accuracy and AUC. As a new form of social lending, P2P lending obtains higher profits for lenders and is also accompanied by a larger risk. To improve the profit of lenders and P2P lending platforms, P2P lending platforms use a scorecard and machine learning methods to

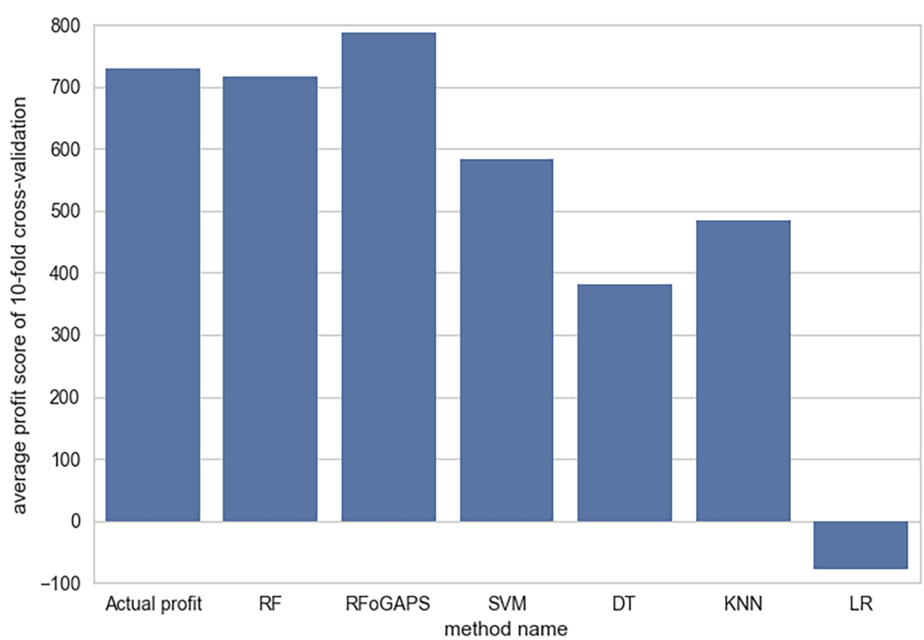
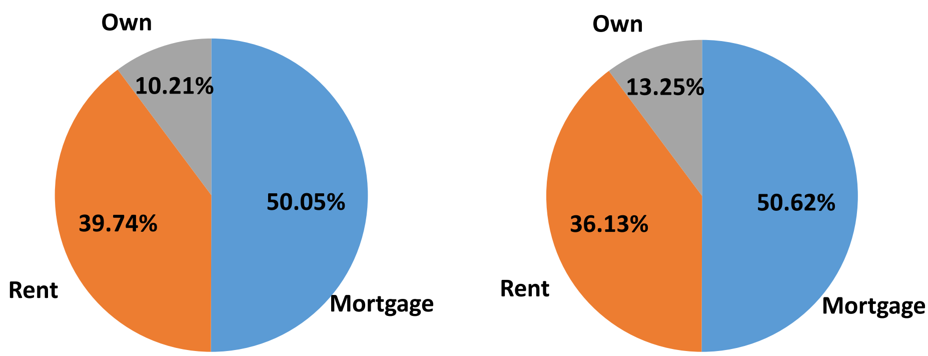



**Fig. 11.** Average profit score of 10-fold cross-validation of compared methods (2014–2016).

**Table 5**
Descriptive statistical analysis results.

| Feature name | Mean | | Std. | | Min | | Max | |
|---|---|---|---|---|---|---|---|---|
| | 2014–2015 | 2016 | 2014–2015 | 2016 | 2014–2015 | 2016 | 2014–2015 | 2016 |
| Annual Income | 11.09 | 11.14 | 0.53 | 0.54 | 8.01 | 6.40 | 16.00 | 15.69 |
| Credit Age | 5.36 | 5.26 | 0.39 | 0.43 | 4.01 | 3.78 | 6.80 | 6.66 |
| Delinquencies | 0.27 | 0.26 | 0.58 | 0.57 | 0.00 | 0.00 | 2.00 | 2.00 |
| Employment Length | 5.69 | 5.69 | 3.82 | 3.88 | 0.00 | 0.00 | 10.00 | 10.00 |
| Inquiries | 0.71 | 0.65 | 0.91 | 0.87 | 0.00 | 0.00 | 3.00 | 3.00 |
| Loan Amount | 14641.02 | 14744.14 | **8479.59** | **9231.77** | 1000.00 | 1000.00 | **35000.00** | **40000.00** |
| Open Accounts | 11.78 | 11.97 | 5.41 | 5.76 | 0.00 | 1.00 | 90.00 | 77.00 |
| Total Accounts | 26.32 | 26.07 | 12.18 | 12.52 | 2.00 | 2.00 | 169.00 | 176.00 |
| DTI | 18.22 | 18.60 | **8.35** | **11.38** | 0.00 | 0.00 | 380.53 | 1622.00 |
| Income to Payment Ratio | 2.68 | 2.76 | 0.60 | 0.69 | −0.13 | −2.73 | 9.00 | 7.18 |
| Income to revolving Ratio | **52.96** | **46.92** | **23.91** | **25.27** | 0.00 | 0.00 | 892.30 | 136.70 |
| Loan status | **0.76** | **0.82** | 0.43 | 0.39 | 0.00 | 0.00 | 1.00 | 1.00 |
| Loan rate | **13.50** | **13.85** | **4.44** | **5.31** | 5.32 | 5.32 | 28.99 | 30.99 |




(a) 2014-2015 (b) 2016

**Fig. 12.** Home ownership.

**Table 6**
Compared experimental results (2016).

| Fold | Profit Score ($100,000) | | | | | | | Improvement rate compared with actual profit (%) | | | | | | Scale of RFoGAPS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Actual profit | RF | RFoGAPS | SVM | DT | KNN | LR | RF | RFoGAPS | SVM | DT | KNN | LR | |
| 1 | 199.25 | 203.38 | **216.27** | 195.69 | 101.94 | 132.32 | 0.07 | 2.07 | **8.54** | −1.79 | −48.84 | −50.58 | −99.97 | 113 |
| 2 | 207.63 | 208.04 | **221.1** | 155.50 | 104.29 | 123.01 | 37.96 | 0.2 | **6.49** | −25.11 | −49.77 | −68.79 | −81.72 | 127 |
| 3 | 186.36 | 195.07 | **207.65** | 193.69 | 111.38 | 135.98 | 17.81 | 4.67 | **11.42** | 3.93 | −40.23 | −37.05 | −90.44 | 100 |
| 4 | 200.22 | 201.49 | **213.35** | 200.66 | 124.17 | 153.35 | 33.25 | 0.63 | **6.56** | 0.22 | −37.98 | −30.57 | −83.39 | 114 |
| 5 | 202.78 | 204.21 | **219.77** | 166.22 | 83.76 | 124.54 | 40.62 | 0.71 | **8.38** | −18.03 | −58.69 | −62.82 | −79.97 | 96 |
| 6 | 201.96 | 202.73 | **215.8** | 151.11 | 45.24 | 117.35 | 25.05 | 0.38 | **6.85** | −25.18 | −77.60 | −72.10 | −87.60 | 124 |
| 7 | 206.74 | 212.04 | **225.12** | 143.17 | 123.32 | 125.47 | 2.89 | 2.56 | **8.89** | −30.75 | −40.35 | −64.77 | −98.60 | 100 |
| 8 | 207.76 | 210.69 | **227.24** | 213.94 | 119.39 | 145.97 | −23.96 | 1.41 | **9.38** | 2.98 | −42.53 | −42.33 | −111.53 | 107 |
| 9 | 202.82 | 207.89 | **222.92** | 175.01 | 175.55 | 155.99 | −22.98 | 2.5 | **9.91** | −13.71 | −13.44 | −30.02 | −111.33 | 100 |
| 10 | 205.14 | 210.4 | **225.85** | 210.79 | 166.20 | 151.52 | −55.20 | 2.56 | **10.1** | 2.75 | −18.98 | −35.39 | −126.91 | 102 |
| Avg. | 202.07 | 205.59 | **219.51** | 180.58 | 115.52 | 136.55 | 5.55 | 1.74 | **8.63** | −10.64 | −42.83 | −47.98 | −97.25 | 108.3 |

construct a loan evaluation model in P2P lending. The purpose of these methods is to eliminate borrowers with high default risk to minimize the overall default rate. Although, these methods are effective at controlling the overall risk, they do not guarantee the maximum profits for lenders. Therefore, in the process of loan evaluation, P2P lending platforms need to not only pay attention to the identification of high-risk borrowers but also take the lender profits into account. The experimental results show that the profit score can evaluate the model from both the risk and the profit perspectives. In addition, it is more useful to minimize the risk and maximize the profits for lenders and P2P lending platforms, compared with traditional evaluation criteria, such as accuracy and AUC. Although the overall default rate from the proposed method is not minimal, it can maximize the profits for lenders, which is more valuable to lenders and P2P lending platforms. Therefore, we suggest that P2P lending platforms should not excessively pursue the minimization of default rates in loan evaluation but should put more emphasis on lender profits. Loans with high risk and high returns should not be rejected thoughtlessly.

Secondly, the analysis and discussion of feature importance are conducted as follows. The importance of the feature represents the

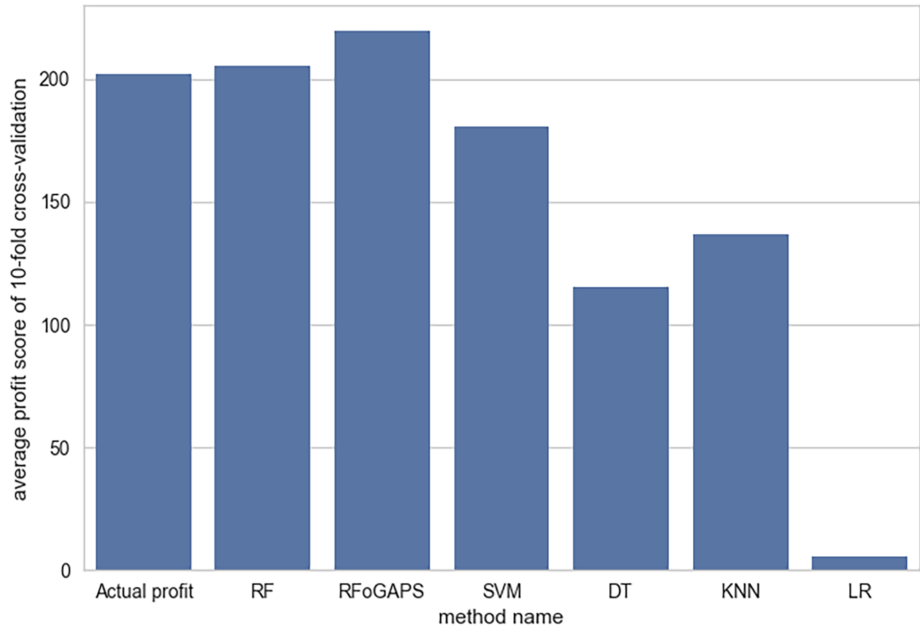



**Fig. 13.** Average profit score of 10-fold cross-validation of compared methods (2016).

**Table 7**
The predictive results of 35 loan records.

| ID | Actual label and predictive label | | | Actual profit of lender ($) | |
|---|---|---|---|---|---|
| | Actual label | Traditional Random Forest | RFoGAPS | Traditional Random Forest | RFoGAPS |
| 1 | 1 | 1 | 1 | 2457 | 2457 |
| 2 | 1 | 1 | 1 | 2050.5 | 2050.5 |
| 3 | 1 | 1 | 1 | 4398.3 | 4398.3 |
| 4 | 1 | 1 | 1 | 3687.3 | 3687.3 |
| 5 | 1 | 1 | 0 | 11568 | 0 |
| 6 | 1 | 1 | 1 | 911.865 | 911.865 |
| 7 | 1 | 1 | 1 | 4208.76 | 4208.76 |
| 8 | 1 | 1 | 0 | 8202 | 0 |
| 9 | 1 | 1 | 0 | 16786 | 0 |
| 10 | 1 | 1 | 1 | 2340 | 2340 |
| 11 | 1 | 1 | 0 | 3441 | 0 |
| 12 | 1 | 1 | 0 | 1806 | 0 |
| 13 | 1 | 1 | 1 | 3117.6 | 3117.6 |
| 14 | 0 | 0 | 0 | 0 | 0 |
| 15 | 1 | 1 | 1 | 10375 | 10375 |
| 16 | 1 | 0 | 0 | 0 | 0 |
| 17 | 1 | 1 | 1 | 9176 | 9176 |
| 18 | 0 | 0 | 0 | 0 | 0 |
| 19 | 0 | 0 | 0 | 0 | 0 |
| 20 | 0 | 0 | 0 | 0 | 0 |
| 21 | 0 | 0 | 0 | 0 | 0 |
| 22 | 0 | 0 | 0 | 0 | 0 |
| 23 | 1 | 0 | 1 | 0 | 10636.92 |
| 24 | 0 | 0 | 0 | 0 | 0 |
| 25 | 0 | 0 | 0 | 0 | 0 |
| 26 | 1 | 0 | 0 | 0 | 0 |
| 27 | 0 | 0 | 0 | 0 | 0 |
| 28 | 0 | 1 | 0 | −28000 | 0 |
| 29 | 0 | 0 | 1 | 0 | −15825 |
| 30 | 0 | 0 | 1 | 0 | −3425 |
| 31 | 0 | 0 | 1 | 0 | −10000 |
| 32 | 0 | 1 | 0 | −29400 | 0 |
| 33 | 1 | 0 | 1 | 0 | 15434.925 |
| 34 | 1 | 0 | 1 | 0 | 13664.7 |
| 35 | 1 | 0 | 1 | 0 | 5997 |

**Table 8**
Feature importance of RFOGAPS (%).

| Primary feature | Secondary feature | 2014–2016 (%) | | 2016 (%) | |
|---|---|---|---|---|---|
| | | Value | Total | Value | Total |
| DTI | / | 10.15 | 10.15 | 10.15 | 10.15 |
| Income to Payment Ratio | / | 10.13 | 10.13 | 10.13 | 10.13 |
| Revolving Utilization Rate | / | 9.50 | 9.50 | 9.50 | 9.50 |
| Revolving to Income Ratio | / | 8.98 | 8.98 | 8.98 | 8.98 |
| Credit Age | / | 8.63 | 8.63 | 8.17 | 8.17 |
| Annual Income | / | 8.40 | 8.40 | 7.79 | 7.79 |
| Total Accounts | / | 7.70 | 7.70 | 7.48 | 7.48 |
| Loan Amount | / | 7.57 | 7.57 | 7.40 | 7.40 |
| Open Accounts | / | 6.94 | 6.94 | 6.65 | 6.65 |
| Employment Length | / | 5.52 | 5.52 | 5.79 | 5.79 |
| Inquiries | / | 4.30 | 4.30 | 4.75 | 4.75 |
| Loan Purpose | debt consolidation | 0.94 | 3.73 | 1.08 | 4.46 |
| | credit card | 0.82 | | 0.92 | |
| | home improvement | 0.45 | | 0.57 | |
| | other | 0.45 | | 0.58 | |
| | major purchase | 0.24 | | 0.31 | |
| | small business | 0.21 | | 0.23 | |
| | medical | 0.17 | | 0.21 | |
| | car | 0.13 | | 0.16 | |
| | moving | 0.11 | | 0.14 | |
| | vacation | 0.10 | | 0.11 | |
| | house | 0.09 | | 0.14 | |
| | renewable energy | 0.02 | | 0.02 | |
| | wedding | 0.00 | | 0.00 | |
| | educational | 0.00 | | 0.00 | |
| Delinquencies | / | 3.35 | 3.35 | 5.18 | 5.18 |
| Term | 36 months | 1.62 | 3.14 | 0.70 | 1.40 |
| | 60 months | 1.52 | | 0.70 | |
| Home ownership | mortgage | 0.81 | 1.95 | 1.04 | 2.72 |
| | rent | 0.74 | | 1.15 | |
| | own | 0.40 | | 0.53 | |

degree to which the feature contributes to the classification effect. The greater the feature importance, the higher the contribution of the feature to the classification effect is. The calculation results of the feature importance for RFoGAPS are shown in Table 8, and some conclusions can be obtained as follows. On the one hand, the DTI, Income to Payment Ratio, Revolving Utilization Rate, Revolving to Income Ratio and Credit Age are the top 5 importance features in the experiments on both datasets from 2014 to 2016 and 2016. These five features are strongly related to the borrower's repayment ability. Therefore, we recommend that P2P lending platforms should pay more attention to the borrower's repayment ability, which has a closer relationship with loan repayment than the other features. If the P2P lending platforms can introduce more features related to the borrower's repayment ability for loan evaluation, the P2P lending platforms may obtain a better performance for loan evaluation. On the other hand, the feature importance of home ownership is 1.95% and 2.72% in the experiments based on the datasets of 2014–2016 and 2016, respectively. This phenomenon illustrates whether borrowers have a fixed asset, such as a house, has less impact on P2P lending. This is different from traditional lending where borrowers need to mortgage fixed assets to financial institutions, such as a bank. P2P lending platforms can also lend to high-quality borrowers even without fixed assets. Although borrowers lacking fixed assets have higher risk, the high profits that these borrowers can bring should not be ignored by P2P lending platforms.

Thirdly, we recommend that P2P lending platforms should establish the dynamic adjustment mechanism of the loan evaluation model, and the dynamic adjustment trigger rules should be defined. There are two reasons for this suggestion. On the one hand, due to problems such as asymmetric information and imperfect risk control measures in P2P lending, the lending policy of the P2P lending platforms is updated dynamically to ensure profits for lenders and P2P lending platforms. On the other hand, the distribution of features may be different for different periods. Therefore, the loan evaluation model in P2P lending should also be adjusted dynamically, and the incremental learning and transfer learning method can be used to solve this problem.

## 5. Conclusions

P2P lending is the intersection of e-commerce and sharing economy and has become an increasingly important lending method for individuals and start-ups. P2P lending is more vulnerable to risk and information asymmetry than traditional lending from banks and other financial organizations. Loan evaluation becomes more important to effectively identify high-risk borrowers and improve lender profits. To more effectively assist P2P lending platforms in loan evaluation, RFoGAPS is proposed and achieves better performances in loan evaluation for P2P lending. Therefore, the proposed method can facilitate the development of P2P lending, and its main contributions can be described as follows.

First, the profit score is used to evaluate the performance of the loan evaluation model. Compared with the traditional evaluation criteria, such as accuracy and AUC, the experimental results show that the profit score can better evaluate the performance of the loan evaluation model in terms of lender profits in P2P lending.

Second, the traditional Random Forest may not guarantee maximizing lender profits. To increase lender profits, a GA is used to optimize the combination of decision trees in the parameter-optimized Random Forest. The proposed method can bring higher profits to lenders than the other compared methods, and it has better computational efficiency than the traditional Random Forest.

Third, by analyzing the feature importance of RFoGAPS, we suggest that P2P lending platforms should lend to borrowers with strong repayment ability and also try to lend to high-quality borrowers, even without property.

In addition, the experimental results also show that the loan evaluation model needs to be dynamically adjusted. The fixed loan evaluation model is unable to adapt for actual needs. So dynamic adjustments to the model should be studied in the future.

## Appendix A. The details of 35 loan records

Table A.9.

**Table A.9**
The details of 35 loan records.

| ID | Annual Income | Credit Age | Delinquencies | Employment Length | Home ownership | Inquiries | Loan Amount | Loan Purpose | Open Accounts | Total Accounts | Term | DTI | Income to Payment Ratio | Revolving Utilization Rate | Revolving to Income Ratio | Loan rate |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 10.31 | 4.41 | 0 | 1 | own | 1 | 8400 | credit card | 9 | 14 | 36 months | 22.8 | 2.23 | 19.4 | −1.25 | 9.75 |
| 2 | 11.26 | 5.68 | 0 | 10 | rent | 2 | 5000 | debt consolidation | 15 | 48 | 36 months | 19.17 | 3.64 | 71.8 | −0.46 | 13.67 |
| 3 | 10.62 | 5.88 | 0 | 0 | own | 0 | 9000 | debt consolidation | 6 | 11 | 36 months | 20.64 | 2.38 | 84.4 | −2.33 | 16.29 |
| 4 | 11.23 | 5.42 | 0 | 10 | mortgage | 3 | 8500 | debt consolidation | 17 | 25 | 36 months | 29.78 | 3.06 | 46.5 | −1.6 | 14.46 |
| 5 | 11.14 | 4.99 | 0 | 10 | mortgage | 0 | 16000 | home improvement | 5 | 11 | 60 months | 23.79 | 2.73 | 94.6 | −0.18 | 14.46 |
| 6 | 10.31 | 4.76 | 0 | 3 | rent | 1 | 2650 | debt consolidation | 6 | 12 | 36 months | 28.24 | 3.35 | 4.1 | 1.94 | 11.47 |
| 7 | 11.29 | 5.11 | 0 | 10 | mortgage | 1 | 10800 | debt consolidation | 7 | 18 | 36 months | 9.24 | 2.91 | 50.5 | −0.49 | 12.99 |
| 8 | 11.08 | 5.39 | 0 | 0 | rent | 1 | 12000 | debt consolidation | 33 | 62 | 60 months | 22.41 | 2.97 | 35.8 | −0.72 | 13.67 |
| 9 | 11.35 | 5.66 | 0 | 9 | mortgage | 1 | 28000 | debt consolidation | 11 | 20 | 60 months | 26.78 | 2.43 | 83 | −2.53 | 11.99 |
| 10 | 11.16 | 5.36 | 1 | 0 | rent | 1 | 8000 | debt consolidation | 7 | 19 | 36 months | 17.66 | 3.12 | 39.7 | −0.16 | 9.75 |
| 11 | 10.92 | 6.01 | 0 | 10 | mortgage | 1 | 10000 | other | 6 | 16 | 36 months | 11.02 | 2.63 | 73.9 | 0.03 | 11.47 |
| 12 | 11.16 | 5.21 | 0 | 6 | rent | 0 | 5600 | debt consolidation | 13 | 26 | 36 months | 19.25 | 3.46 | 52.9 | −0.01 | 10.75 |
| 13 | 10.6 | 5.38 | 0 | 10 | mortgage | 0 | 8000 | credit card | 5 | 16 | 36 months | 8.13 | 2.52 | 82.2 | −0.83 | 12.99 |
| 14 | 10.82 | 4.76 | 0 | 3 | rent | 3 | 14850 | other | 24 | 28 | 60 months | 13.68 | 2.23 | 39.1 | 0.31 | 26.57 |
| 15 | 11.86 | 5.32 | 0 | 3 | mortgage | 0 | 10000 | home improvement | 14 | 26 | 60 months | 27.9 | 3.78 | 48.9 | −0.39 | 20.75 |
| 16 | 11.26 | 5.11 | 1 | 0 | rent | 0 | 20000 | debt consolidation | 10 | 23 | 60 months | 18.29 | 2.54 | 84.2 | 0.07 | 18.25 |
| 17 | 10.86 | 5.66 | 0 | 1 | mortgage | 2 | 16000 | credit card | 4 | 14 | 60 months | 18 | 2.51 | 73.2 | −0.71 | 11.47 |
| 18 | 10.64 | 5.04 | 0 | 0 | rent | 0 | 7200 | debt consolidation | 9 | 24 | 36 months | 27.96 | 2.66 | 92.7 | −1.45 | 12.99 |
| 19 | 11.27 | 5.45 | 0 | 3 | rent | 0 | 18375 | debt consolidation | 23 | 54 | 60 months | 46.47 | 2.42 | 52.2 | −1.29 | 28.99 |
| 20 | 10.31 | 4.96 | 0 | 1 | mortgage | 0 | 20000 | other | 11 | 24 | 60 months | 41.44 | 1.53 | 70 | −1.67 | 21.18 |
| 21 | 10.71 | 4.98 | 0 | 1 | rent | 1 | 22425 | credit card | 20 | 33 | 60 months | 36.39 | 1.88 | 75 | −1.7 | 18.25 |
| 22 | 10.82 | 5.37 | 0 | 0 | rent | 2 | 23000 | debt consolidation | 7 | 29 | 36 months | 26.6 | 1.66 | 94.7 | −1.81 | 14.46 |
| 23 | 10.46 | 4.81 | 0 | 1 | rent | 0 | 12600 | debt consolidation | 7 | 9 | 36 months | 30.16 | 1.72 | 73.1 | −1.79 | 28.14 |
| 24 | 9.74 | 4.98 | 0 | 0 | rent | 3 | 5425 | debt consolidation | 9 | 27 | 36 months | 33.97 | 1.92 | 17.6 | −0.58 | 21.97 |
| 25 | 11.2 | 5.39 | 0 | 10 | rent | 1 | 35000 | debt consolidation | 9 | 16 | 60 months | 29.13 | 1.75 | 85.9 | −1.63 | 26.57 |
| 26 | 10.93 | 4.77 | 0 | 1 | rent | 0 | 18500 | debt consolidation | 15 | 20 | 60 months | 37.93 | 2.25 | 92.1 | −1.05 | 19.99 |
| 27 | 10.92 | 5.34 | 0 | 2 | rent | 2 | 24200 | credit card | 12 | 25 | 36 months | 33.67 | 1.64 | 75.7 | −1.52 | 18.99 |
| 28 | 11.61 | 5.3 | 0 | 4 | mortgage | 0 | 28000 | debt consolidation | 14 | 28 | 60 months | 13.95 | 2.51 | 70.6 | −0.83 | 19.99 |
| 29 | 10.49 | 5.37 | 0 | 0 | rent | 1 | 15825 | debt consolidation | 7 | 28 | 36 months | 47.43 | 1.64 | 93.7 | −1.44 | 19.53 |
| 30 | 9.35 | 5.03 | 0 | 0 | rent | 2 | 3425 | debt consolidation | 4 | 9 | 36 months | 30.28 | 1.93 | 84.9 | −0.41 | 26.57 |

(continued on next page)

**Table A.9** (*continued*)

| ID | Annual Income | Credit Age | Delinquencies | Employment Length | Home ownership | Inquiries | Loan Amount | Loan Purpose | Open Accounts | Total Accounts | Term | DTI | Income to Payment Ratio | Revolving Utilization Rate | Revolving to Income Ratio | Loan rate |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 31 | 10.78 | 4.98 | 0 | 1 | rent | 3 | 10000 | debt consolidation | 10 | 10 | 60 months | 22.7 | 2.7 | 54 | −0.28 | 20.75 |
| 32 | 11.16 | 5.42 | 2 | 9 | rent | 0 | 29400 | debt consolidation | 17 | 38 | 60 months | 26.11 | 1.91 | 30.4 | −1.21 | 24.99 |
| 33 | 10.78 | 4.36 | 0 | 6 | rent | 2 | 14575 | debt consolidation | 11 | 14 | 60 months | 24.3 | 2.31 | 42.6 | −0.07 | 21.18 |
| 34 | 10.71 | 5.15 | 0 | 4 | rent | 0 | 18900 | debt consolidation | 14 | 17 | 60 months | 28.61 | 2.13 | 41.7 | −1.3 | 14.46 |
| 35 | 10.36 | 4.81 | 2 | 5 | rent | 1 | 10000 | debt consolidation | 16 | 38 | 36 months | 29.47 | 1.96 | 53.6 | −0.28 | 19.99 |

## References


Adnan, M.N., Islam, M.Z., 2016. Optimizing the number of trees in a decision forest to discover a subforest with high ensemble accuracy using a genetic algorithm. Knowl.-Based Syst. 110, 86–97.
Bekhet, H.A., Eletter, S.F.K., 2014. Credit risk assessment model for jordanian commercial banks: neural scoring approach. Rev. Develop. Finance 4, 20–28.
Breiman, L., 2001. Random forest. Mach. Learn. 45, 5–32.
Breiman, L., Friedman, J.H., Olshen, R., Stone, C.J., 1984. Classification and regression trees. Encyclopedia Ecol. 40, 582–588.
Brown, I., Mues, C., 2012. An experimental comparison of classification algorithms for imbalanced credit scoring data sets. Expert Syst. Appl. 39, 3446–3453.
Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297.
Díez-Pastor, J.F., Rodríguez, J.J., García-Osorio, C.I., Kuncheva, L.I., 2015. Diversity techniques improve the performance of the best imbalance learning ensembles. Inf. Sci. 325, 98–117.
Eisenbeis, Robert, A., 1977. Pitfalls in the application of discriminant analysis in business, finance, and economics. J. Finance 32, 875–900.
Feldman, D., Gross, S.T., 2005. Mortgage default: classification trees analysis. J. Real Estate Finance Econ. 30, 369–396.
Gao, Y., Yu, S.-H., Shiue, Y.-C., 2018. The performance of the p2p finance industry in china. Electron. Commer. Res. Appl. 30, 138–148.
Guo, Y., Zhou, W., Luo, C., Liu, C., Xiong, H., 2016. Instance-based credit risk assessment for investment decisions in p2p lending. Eur. J. Oper. Res. 249, 417–426.
Holland, J.H., 1992. Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence. Ann. Arbor. 6, 126–137.
Huang, B.F.F., Boutros, P.C., 2016. The parameter sensitivity of random forests. BMC Bioinf. 17, 331.
Huang, Z., Chen, H., Hsu, C.-J., Chen, W.-H., Wu, S., 2004. Credit rating analysis with support vector machines and neural networks: a market comparative study. Decis. Support Syst. 37, 543–558.
Jin, Y., Zhu, Y., 2015. A data-driven approach to predict default risk of loan for online peer-to-peer (p2p) lending. In: Fifth International Conference on Communication Systems and Network Technologiespp. 609–613.
Kim, Y., Oh, I., 2008. Classifier ensemble selection using hybrid genetic algorithms. Pattern Recogn. Lett. 29, 796–802.
Lei, X., 2016. Discussion of the risks and risk control of p2p in china. Modern Econ. 07, 399–403.
Lessmann, S., Baesens, B., Seow, H.-V., Thomas, L.C., 2015. Benchmarking state-of-the-art classification algorithms for credit scoring: an update of research. Eur. J. Oper. Res. 247, 124–136.
Li, N., Cheng, X., Guo, H., Wu, Z., 2016. Recognizing human interactions by genetic algorithm-based random forest spatio-temporal correlation. Pattern Anal. Appl. 19, 267–282.
Lu, Z., Wu, X., Zhu, X., Bongard, J., 2010. Ensemble pruning via individual contribution ordering. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Miningpp. 871–880.
Ma, H., Wang, X., 2016. Influencing factor analysis of credit risk in p2p lending based on interpretative structural modeling. J. Discrete Math. Sci. Cryptogr. 19, 777–786.
Ma, B., Zhou, Z., Hu, F., 2017. Pricing mechanisms in the online peer-to-peer lending market. Electron. Commer. Res. Appl. 26, 119–130.
Malekipirbazari, M., Aksakalli, V., 2015. Risk assessment in social lending via random forests. Expert Syst. Appl. 42, 4621–4631.
Malhotra, R., Malhotra, D., 2002. Differentiating between good credits and bad credits using neuro-fuzzy systems. Eur. J. Oper. Res. 136, 190–211.
Martínez-Muñoz, G., Suárez, A., 2004. Aggregation ordering in bagging. In: Proc. of the IASTED International Conference on Artificial Intelligence and Applications Citeseer, pp. 258–263.
Metawa, N., Hassan, M.K., Elhoseny, M., 2017. Genetic algorithm based model for optimizing bank lending decisions. Expert Syst. Appl. 80, 75–82.
Moro, S., Cortez, P., Rita, P., 2014. A data-driven approach to predict the success of bank telemarketing. Decis. Support Syst. 62, 22–31.
Mundim, L.R., Andretta, M., de Queiroz, T.A., 2017. A biased random key genetic algorithm for open dimension nesting problems using no-fit raster. Expert Syst. Appl. 81, 358–371.
Ortega, A.C.B., Bell, F., 2008. Online social lending: borrower-generated content. In: Learning From the Past & Charting the Future of the Discipline Americas Conference on Information Systems, pp. 380.
Rodriguez-Galiano, V., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sanchez, J., 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS J. Photogrammetry Remote Sens. 67, 93–104.
Savvopoulos, A., 2010. Consumer credit models: pricing, profit and portfolios. J. R. Stat. Soc. 173 468–468.
Schebesch, K.B., Stecking, R., 2005. Support vector machines for credit scoring: extension to non standard cases. In: Innovations in Classification, Data Science, and Information Systems Springer, pp. 498–505.
Seiffert, C., Khoshgoftaar, T.M., Van Hulse, J., Napolitano, A., 2010. Rusboost: a hybrid approach to alleviating class imbalance. IEEE Trans. Syst. Man Cybern. Part A Syst. Humans 40, 185–197.
Serrano-Cinca, C., Gutiérrez-Nieto, B., López-Palacios, L., 2015. Determinants of default in p2p lending. PLoS One 10 e0139427.
So, M.C., Thomas, L.C., Seow, H.-V., Mues, C., 2014. Using a transactor/revolver scorecard to make credit and pricing decisions. Decis. Support Syst. 59, 143–151.
Thomas, L.C., 2000. A survey of credit and behavioural scoring: forecasting financial risk

of lending to consumers. Int. J. Forecast. 16, 149–172.
Verbraken, T., Bravo, C., Weber, R., Baesens, B., 2014. Development and application of consumer credit scoring models using profit-based classification measures. Eur. J. Oper. Res. 238, 505–513.
Wang, Y., Wang, D., Wang, Y., Jin, Y., 2017. Predicting survivability of colorectal cancer by an ensemble classification method improved on random forest. J. Manage. Sci. 30, 95–106.
Wang, Z., Jiang, C., Ding, Y., Lyu, X., Liu, Y., 2018. A novel behavioral scoring model for estimating probability of default over time in peer-to-peer lending. Electron. Commer. Res. Appl. 27, 74–82.
West, D., 2000. Neural network credit scoring models. Comput. Oper. Res. 27, 1131–1152.
Wiginton, J.C., 1980. A note on the comparison of logit and discriminant models of consumer credit behavior. J. Financ. Quantitative Anal. 15, 757–770.
Xia, Y., Liu, C., Liu, N., 2017. Cost-sensitive boosted tree for loan evaluation in peer-to-peer lending. Electron. Commer. Res. Appl. 24, 30–49.
Zhang, Y., Jia, H., Diao, Y., Hai, M., Li, H., 2016. Research on credit scoring by fusing social media information in online peer-to-peer lending. Procedia Comput. Sci. 91, 168–174.
Zhou, Z.-H., Wu, J., Tang, W., 2002. Ensembling neural networks: many could be better than all. Artif. Intell. 137, 239–263.