Received: 7 January 2020 | Revised: 2 April 2020 | Accepted: 19 June 2020
DOI: 10.1002/isaf.1480

RESEARCH ARTICLE

WILEY

A neural-network-based decision-making model in the peer-to-peer lending market

Golnoosh Babaei | Shahrooz Bamdad

Department of Industrial Engineering, Islamic Azad University, Tehran, Iran

Correspondence
Shahrooz Bamdad, Department of Industrial Engineering, Islamic Azad University, South Tehran Branch. Tehran, Iran.
Email: sh_bamdad@azad.ac.ir

Summary

This study proposes an investment recommendation model for peer-to-peer (P2P) lending. P2P lenders usually are inexpert, so helping them to make the best decision for their investments is vital. In this study, while we aim to compare the performance of different artificial neural network (ANN) models, we evaluate loans from two perspectives: risk and return. The net present value (NPV) is considered as the return variable. To the best of our knowledge, NPV has been used in few studies in the P2P lending context. Considering the advantages of using NPV, we aim to improve decision-making models in this market by the use of NPV and the integration of supervised learning and optimization algorithms that can be considered as one of our contributions. In order to predict NPV, three ANN models are compared concerning mean square error, mean absolute error, and root-mean-square error to find the optimal ANN model. Furthermore, for the risk evaluation, the probability of default of loans is computed using logistic regression. Investors in the P2P lending market can share their assets between different loans, so the procedure of P2P investment is similar to portfolio optimization. In this context, we minimize the risk of a portfolio for a minimum acceptable level of return. To analyse the effectiveness of our proposed model, we compare our decision-making algorithm with the output of a traditional model. The experimental results on a real-world data set show that our model leads to a better investment concerning both risk and return.

KEYWORDS

net present value, peer-to-peer lending, portfolio optimization

1 | INTRODUCTION

Peer-to-peer (P2P) lending has become popular in recent years because it reduces financing costs by eliminating the need for a traditional financial intermediary (Guo, Zhou, Luo, Liu, & Xiong, 2016). In the P2P lending market, borrowers sign up on online platforms and submit applications for loans. Lenders (or investors) browse the available applications, which are called listings, and partially select some of them for investment (Xia, Liu, & Liu, 2017). If a listing has a sufficient amount of money, it becomes a loan. Normally, to reduce the systematic risk, a lender divides the money into different loans, but most of them are not expert and cannot transform the available information into smart investments (Mild, Waitz, & Wöckl, 2015). Therefore, loan evaluation is an effective tool to guide unprofessional lenders in making rational decisions (Guo *et al*., 2016). As making decisions on whether or not to fund a certain loan is challenging for lenders in the P2P lending market, various loan evaluation models, namely random forest (Malekipirbazari & Aksakalli, 2015), decision tree (Serrano-Cinca & Gutiérrez-Nieto, 2016), artificial neural network (Byanjankar, Heikkila, & Mezei, 2015), and kernel method (Guo *et al*., 2016), have been established.

The P2P lending market is new and developing, so research on the loan evaluation models for P2P lending is limited. In 2015, a random-forest-based classification method was proposed by

Malekipirbazari and Aksakalli (2015) that, while decreasing the performance, is superior to rating-based models in identifying the best borrowers. Additionally, cost sensitivity was considered in their study, but they failed to enhance the capability of the proposed model to distinguish default borrowers. Emekter, Tu, Jirasakuldech, and Lu (2015) used logistic regression and survival analysis to develop a credit risk evaluation model for P2P lending, and they found the variables most related to loan default.

Serrano-Cinca and Gutiérrez-Nieto (2016) proposed the profit scoring approach as an alternative to credit scoring for P2P lending. They focused on the expected profitability of investing in P2P loans using the well-known financial variable called the internal rate of return (IRR), and the results showed promising IRR. Many studies have used IRR to evaluate the profitability of loans in this market (Bastani, Asgari, & Namavari, 2019; Cho, Chang, & Song, 2019), but IRR can scarcely be used to rate mutually exclusive loans (Magni, 2013). Therefore, investigating models that completely consider the special requirements of loan evaluation in P2P lending is crucial. The net present value (NPV) is another financial variable that has been used in many investment decision-making problems (Magni, 2005; Puška, Beganovic, & Šadic, 2018). This measure is the difference between the benefits and the costs while considering the discounted values. Using NPV leads to a fair comparison among different projects because it considers the time value of money and transfers all forecasts to today's scale (Karellas, Boukis, & Kontopoulos, 2010). Also, discount rates are utilized to compute NPV while these rates of projects are different, so using NPV provides decision-makers an effective assessment of a loan based on its financial context (Hopkinson, 2016).

It was claimed by some researchers that lenders could not create complex models for their investment strategies (Klafft, 2008; Mild *et al*., 2015), but only a few studies examined the investment portfolio in P2P lending. In this context, Emekter *et al*. (2015) proposed a strategy based on lending only to the safest borrowers in the LendingClub (an American P2P lending company headquartered in San Francisco, California). However, Malekipirbazari and Aksakalli (2015) found that borrowers with the highest credit scores are not necessarily borrowers from a profitability perspective. In 2016, an instance-based loan evaluation model was proposed by Guo *et al*. (2016). They formulated the investment decision in P2P lending as a portfolio optimization problem and solved it by linear programming. Zhao, Liu, Wang, Ge, and Chen (2016) considered the investment decision in P2P lending as a multi-objective problem for the first time. Three objectives of their model were non-default probability, fully funded probability, and winning–bidding probability. They proposed two portfolio allocation strategies based on weighted objective optimization and multi-objective optimization in selecting portfolios for lenders. However, the majority of the platforms have replaced the Dutch auction with the buyout auction recently, so their proposed model seems to be obsolete (Xia *et al*., 2017). Babaei and Bamdad (2020) utilized the non-dominated sorting genetic (NSGA2) algorithm to propose a multi-objective instance-based decision support system for investment recommendation in P2P lending. Their proposed model was compared with single-objective and profit-scoring methods, and they concluded that, although the profit-based approach led to the most profitable portfolio, this portfolio, in comparison with their proposed portfolio, had a higher level of risk. Moreover, their proposed model in comparison with the single-objective model improved the investment based on two objectives (i.e. return maximization and risk minimization).

The focus of this study is on comparing different ANN models to choose the best ANN for proposing an investment recommendation model in the P2P lending market. In order to evaluate the profitability of loans, NPVs of available investment choices are obtained by the use of the ANNs. This supervised learning method has been widely used in many financial studies because it was claimed by many researchers that ANNs are one of the strongest models to learn the relationship between inputs and outputs for the financial predictions (Asadi, Hadavandi, Mehmanpazir, & Nakhostin, 2012; Enke & Thawornwong, 2005; Fernandez-Rodrıguez, González-Martel, & Sosvilla-Rivero, 2000), but few studies have compared different ANN models in the context of return evaluation in the P2P lending market. So, we first compare three various models of ANNs and then choose the best model for our proposed model to improve its performance. In general, the technical novelty of our research is to integrate different machine learning models with a portfolio optimization problem. Another supervised learning method that is utilized to estimate the probability of default (PD) of loans is logistic regression. Many researchers have used logistic regression in different credit-scoring models (Bensic, Sarlija, & Zekic-Susac, 2005; Joanes, 1993; Puro, Teich, Wallenius, & Wallenius, 2010). For instance, Zekic-Susac, Sarlija, and Bensic (2004) used logistic regression as a methodology for small business credit scoring. In our study, after training models based on finished loans in the past, the return and risk of open listings are evaluated using the trained models, and then we apply an optimization method to find the optimal portfolio. As a result, investors can find how much money they should allocate to each listing. As the contribution of our study, we aim to make a comparison between different ANN models for return prediction in the P2P lending market and select the best ANN model to improve P2P investment decision-making.

The remainder of this paper is organized as follows: Section 2 presents the methods used in our study. Section 3 presents the empirical study. The results and discussion are presented in Section 4. A sensitivity analysis is undertaken in Section 5. Finally, Section 6 concludes this study.

2 | METHODOLOGY

For the purpose of helping lenders to evaluate listings and improve their investments in P2P lending platforms, we develop an investment recommendation model. Figure 1 illustrates the proposed model, which first takes raw data and then balances it. When the number of instances in one class is different from the number of instances in another class, a class imbalance problem occurs. In analysing an imbalanced data set, classifiers tend to be biased towards the majority class,

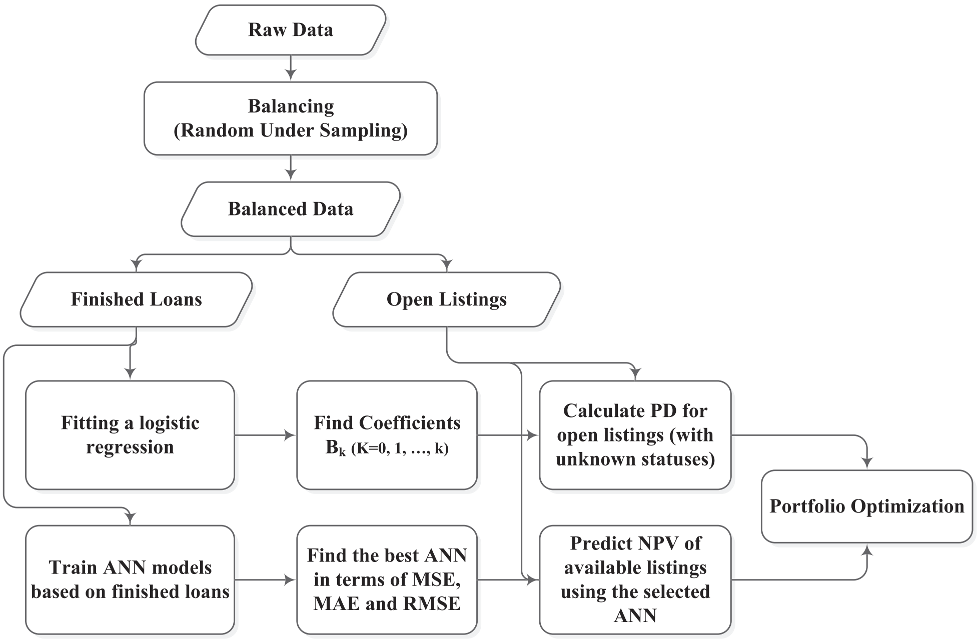

FIGURE 1 The proposed model process

whereas the detection of the minority class is much more important (Chawla, Japkowicz, & Kotcz, 2004). For example, in most P2P lending data sets, the majority class contains fully paid loans, whereas predicting loans with a high probability of failing is more important because if lenders choose a failed loan they face a loss. There are many algorithms to address imbalanced data classification problems, and resampling is one of the most important strategies for solving this issue (Guo *et al*., 2017) that generates a balanced data set before building the model. The three types of resampling techniques are oversampling, undersampling, and a hybrid of the two (Guo *et al*., 2017). In terms of computational time, the undersampling method is the best method for big data sets (Loyola-González, Martínez-Trinidad, Carrasco-Ochoa, & García-Borroto, 2016; Napierała & Stefanowski, 2015), as it reduces the number of samples from the majority class. We utilize this method for the preprocessing part of our model.

The balanced data are grouped into finished loans (known payback status) as the training data and open listings (unknown payback status) as the test data. The next process of our model is loan evaluation, in which we evaluate open listings from two perspectives: risk and return. For risk evaluation, a logistic regression model is fitted based on finished loans and the response variable is loan status, which is a binary variable where 1 stands for failed loans and 0 stands for fully paid ones. After fitting the regression model and finding coefficients $\hat{\beta}_{k\ k}$ (K = 0, 1, …, k), PDs of open listings are computed as follows:

$$PD = \frac{1}{1 + e^{-\left(\hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_k X_k\right)}} \tag{1}$$

Coefficients $\hat{\beta}_k$ show the relation among 19 independent variables and the response variable. The independent variables of our model are as follows (a description of the variables is located in the LendingClub data dictionary table[1]).

- Annual_Inc: the self-reported annual income provided by the borrower during registration;
- Credit Age: number of days of credit history considering the date when the borrower's earliest reported credit line was opened;
- Delinq_2yrs: the number of ≥30 days past-due incidences of delinquency in the borrower's credit file for the past 2 years
- Emp Length: employment length in years (possible values are between 0 and 10, where 0 means less than 1 year and 10 means ≥10 years);
- Home_Ownership: own, rent, mortgage;
- Public Records: number of derogatory public records;
- Months Since Last Delinquency: the number of months since the borrower's last delinquency
- Inq_Last_6mths: the number of inquiries in the past 6 months;
- Loan_Amnt: the listed amount of the loan applied for by the borrower.
- Purpose: 14 loan purposes; namely, wedding, credit card, car loan, major purchase, home improvement, debt consolidation, house, vacation, medical, moving, renewable energy, educational, small business, and other
- Open_Acc: the number of open credit lines in the borrower's credit life;
- FICO score: a measure of credit risk; based on credit reports and ranging from 300 to 850 (FICO is a registered trademarks of Fair Isaac Corporation).
- LC Grade: LendingClub categorizes borrowers into seven different loan grades from A down to G, with A-grade being the safest.

[1]https://resources.lendingclub.com/LCDataDictionary.xlsx

- LC Subgrade: LendingClub-assigned loan subgrade. There are 35 loan subgrades in total for borrowers from A1 down to G5, with A1 subgrade being the safest.
- Dti: a ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LendingClub loan, divided by the borrower's self-reported monthly income.
- Revol_Util: revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit.
- Int_Rate: the interest rate on the loan paid by the borrower.
- Loan Amount To Annual Income: loan amount to the reported annual income.
- Annual Instalment To Income: the annual payment owed by the borrower divided by the annual income provided by the borrower during registration.

For return evaluation, this study utilizes ANNs. ANNs are based on neurons that are connected via weights (Giannopoulos & Aggelopoulos, 2019; Sun & Vasarhelyi, 2018; Trinkle & Baldwin, 2016). This paper is one of the few comparative studies that utilize different ANN models and compare their performance for a special problem in the P2P lending market. We apply three ANN models, namely FeedForwardNet, CascadeForwardnet, and FitNet. We first train ANNs with finished loans and try to find the most suitable network for our investment recommendation model.

The structure of the feedforward ANN is presented in Figure 2. It contains inputs, outputs, and hidden layers and has no feedback elements (Ahmad AL-Allaf, 2012). The outputs are connected to the inputs by the weight vectors and activation functions (Gardner & Dorling, 1998; Guo *et al*., 2017). Therefore, it can model highly nonlinear functions.

A cascade-forward ANN generally is similar to FeedForwardNet, but in this network each layer is related to all previous layers (Badde, Gupta, & Patki, 2013). The structure of a cascade-forward ANN is shown in Figure 3.

A function-fitting neural network is also similar to a feedforward ANN model and applied to fit the input–output relationship. There are different methods to improve the performance of the network; for instance, finding the optimal network architecture. Thus, to find the optimal network structure, we compare these three ANN models and run them with different training functions (i.e. trainlm, trainbr, trainscg, and trainrp) and various numbers of hidden neurons. In addition, we use the three performance metrics mean square error (MSE), mean absolute error (MAE), and root-mean-square error (RMSE) to find the optimum network to predict the NPV of open listings. At the end of the loan evaluation step, all information about available listings is clear and lenders can create their portfolio. Our portfolio optimization model is as follows:

$$\text{Min} \sum_{i=1}^{I} \lambda_i^2 \text{PD}_i \tag{2}$$

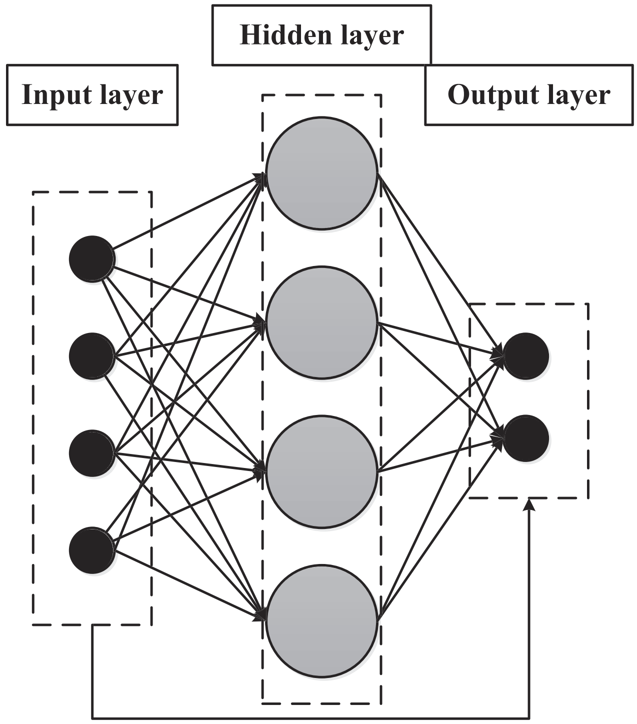

FIGURE 3 Cascade-forward ANN structure

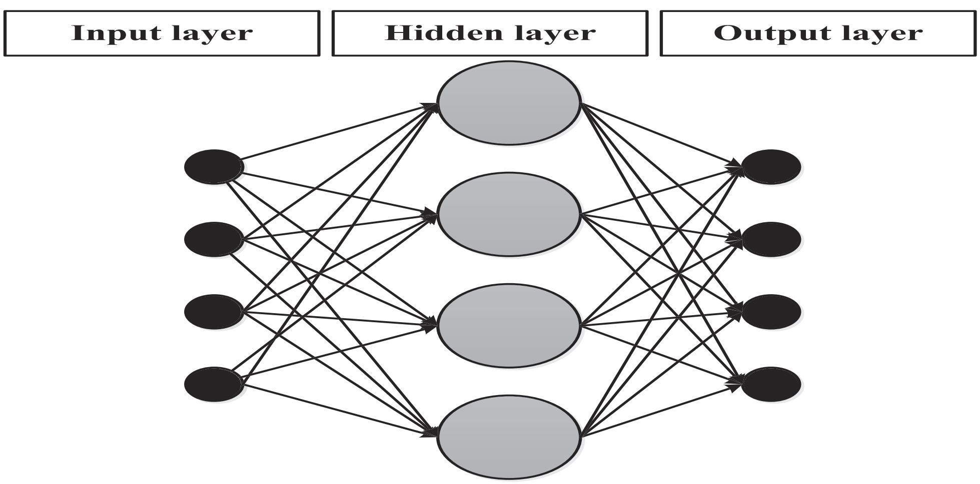

FIGURE 2 Feedforward ANN structure

Subject to:

$$\sum_{i=1}^{I} \lambda_i \mathrm{NPV}_i \geq R^* \tag{3}$$

$$\sum_{i=1}^{I} \lambda_i = 1 \tag{4}$$

$$m \leq \lambda_i M \leq \text{amount of loan}_i \tag{5}$$

$$\lambda_i \geq 0 \tag{6}$$

Here, λ_i is the decision variable, which represents the optimal proportion of the total investment amount allocated to the ith loan. As is shown in equation 2, the objective is to minimize the risk of investment. This model tries to find a portfolio that has an equal or higher level of return than R^*, equation 3, which is the minimum acceptable amount of NPV that can vary according to investors' preferences. Equation 4 indicates that the sum of the proportions should be equal to 1, and equation 5 shows the constraints on investment amounts. In the P2P lending platforms, a minimum investment amount m is set for issued loans. In LendingClub and our study, m = \$25. Moreover, it never can be possible to allocate an amount much more than the requested loan amount to that listing.

3 | EMPIRICAL STUDY

3.1 | Sample and Data

This study utilizes the LendingClub public data set to validate the proposed model. As we want to evaluate the return of open listings based on the available information about closed loans in the past, so we need to have completely finished loans. For this reason, 36-month loans in 2013 are selected for this study. This data set consists of 1,163 loans. As mentioned in Section 2, there are 19 independent variables in our data set: Annual_Inc, Credit Age, Delinq_2yrs, Emp Length, Home_Ownership, Inq_Last_6mths, Loan_Amnt, Purpose, Open_Acc, FICO score, LC Subgrade, Dti, Revol_Util, Int_Rate, Loan Amount To Annual Income, Annual Instalment To Income, Public Records, Months Since Last Delinquency, LC Grade.

3.2 | Balancing

As illustrated in Section 2, different methods for solving the imbalanced data problems have been proposed by researchers (Chawla *et al*., 2004; Han, Wang, & Mao, 2005; Longadge & Dongre, 2013). As our data set is big, we use the undersampling method. Loans are separated into two classes: failed loans, for which the label of this class is 1; and fully paid loans, for which the label is 0. The number of fully paid loans is more than failed loans, so the prediction model ignores the minority class, whereas the prediction of failed loans is much more important, because if failed loans are not forecasted correctly then lenders select them for their investment and suffer from the loss of profit. The distribution of classes before balancing is presented in Figure 4.

3.3 | Prediction of the Return

We use NPV to predict the return of open listings. NPV is the difference between the present value of cash inflows and the present value of cash outflows. NPV is used in investment planning to analyse the profitability of a project. The equation of NPV is as follows:

$$\mathrm{NPV} = B - C \tag{7}$$

$$B = \sum_t \frac{b_t}{(1+r)^t} \tag{8}$$

$$C = \sum_t \frac{c_t}{(1+r)^t} \tag{9}$$

where B is the present value of all inflows b_t, C is the present value of all cash outflows c_t, and r is the discount rate, which is the interest rate of loans in our model. In addition, t shows the time horizon. The three ANN models mentioned in Section 2, namely FeedForwardNet, CascadeForwardNet, and FitNet, are run for the finished loans. In order to find the optimal number of hidden neurons in different models with various training functions (trainbr, trainlm, trainscg, and trainrp), a 10-fold cross-validation method is used to discover how many neurons lead to the least MSE. After finding the optimal hidden neurons, we compare three ANN models containing the optimal number of hidden neurons based on MSE, MAE, and RMSE. The results of this comparison are presented in Table 1. Weights in ANNs are initialized randomly, so all models are trained several times and the performance measures are then averaged. It can be concluded that the CascadeForwardNet model containing trainbr function and 40 neurons in the hidden layer has the best performance, and so it is chosen for the rest of our model. Selecting this ANN model for our algorithm can contribute to a good prediction of NPV, since it has the lowest level of error.

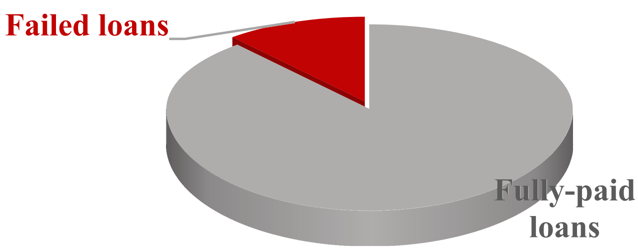

FIGURE 4 Distributions of classes before balancing

TABLE 1 Comparison of the three ANN models' performances

ANN model	Training function	Optimal hidden neurons	MSE	MAE	RMSE
FeedForwardNet	trainlm	15	0.0900	0.2360	0.2955
	trainbr	15	0.1076	0.2641	0.3277
	trainscg	25	0.0782	0.2194	0.2792
	trainrp	10	0.0768	0.2168	0.2768
CascadeForwardNet	trainlm	30	0.1084	0.2653	0.3285
	trainbr	40	**0.0617**	**0.1990**	**0.2484**
	trainscg	25	0.0804	0.2272	0.2833
	trainrp	10	0.0856	0.2314	0.2921
FitNet	trainlm	5	0.0690	0.2079	0.2626
	trainbr	20	0.0907	0.2375	0.3001
	trainscg	5	0.0676	0.2066	0.2599
	trainrp	15	0.0818	0.2250	0.2854

3.4 | Prediction of the Risk

In Section 3.3, the return of listings was forecasted by the use of ANNs, but as we know, investors pay attention to both return and risk of an investment, so in this section we aim to predict the risk of listings. As the loan statuses of finished loans are clear, a logistic regression model is fitted based on them and coefficients are found. After that, the PDs of open listings (with unknown statuses) are estimated by the use of equation 1.

3.5 | Investment Recommendation Algorithm

The investment recommendation model may be described as summarized in Algorithm 1.

Our algorithm has three parameters. (1) The minimum investment amount m that is set by the P2P lending platform. As we use the LendingClub data set in our model, m = \$25. (2) The total investment amount M is the amount of money that each investor has for investment in P2P loans. We assume M is \$15,000 for our numerical experiments. (3) R^* is the minimum amount of return that an investor accepts, and is \$1,000 in our model.

In the data preprocessing part (lines 1 and 2), we first balance the data set by the random undersampling method. After that, we prepare the training and test data sets. In fact, the training data are finished loans whose information is available and can be used to evaluate the test data, which include open listings. In the model training part (lines 3 and 4), the best ANN model is found for prediction of the return. For this purpose, we use a 10-fold cross-validation method to find the best number of hidden neurons for each ANN model with different training functions. After that, ANNs with optimal hidden neurons are run for several times to get the average amounts of MSE, MAE, and RMSE and understand which ANN model leads to the least error. In terms of risk, we build a logistic regression model with finished loans and extract the coefficients $\hat{\beta}_k$ of independent variables. In the next part (lines 5–8), test data are fed to the trained network found in line 3 to predict the NPV of open listings. After that, the PD of each listing is evaluated with the help of the coefficients generated by the logistic regression in line 4.

In the portfolio optimization part (lines 9–14), we try to find a portfolio with optimal risk with respect to a minimum acceptable level of return. Lines 9–14 are similar to equations 2–6. The output of our proposed algorithm is a set of proportions that can be used by investors to create an optimal portfolio.

4 | RESULTS AND DISCUSSION

Extensive experiments have been performed on the real-world data set described in Section 3.1. Table 2 shows the information of the selected portfolio from our proposed algorithm.

Specifically, we compare our model with an occasion that an investor knows nothing about the portfolio optimization methods and share his or her total money among open listings equally. The outputs of these models are shown in Table 3. It is clear that our proposed investment recommendation model outperforms the traditional model applied by inexpert lenders.

It is clear that the portfolio return of the traditional model is followed by an increase of 16.27% and reaches to 2273.95 in our proposed model. Moreover, the risk of investment in our proposed model is 33.33% smaller than the risk of the generated portfolio with equal proportions in the traditional model. Therefore, our model can improve the traditional model based on risk and return.

Finally, it should be noted that our algorithm contains an optimization problem that results in an optimal solution and is validated based on the selected data explained in Section 3.1. Therefore, investors can trust the proposed model based on the general patterns and rules generated from these data.

Algorithm 1

Investment decision model

Inputs:
m: the minimum investment amount required for funding a loan;
M: the total amount available for investment;
R^*: the minimum acceptable level of return;
Output:
A set of proportions λ_i ($i = 1, \ldots, l$) which make an optimal portfolio;
//Data preprocessing
1. Running random undersampling to balance the data set;
2. Generate train (finished loans with known statuses) and test (open listings with unknown statuses) data sets.

//Model training
3. Find the best ANN model to predict NPV:
   - ➢ Ten-fold cross-validation method to find the optimal number of hidden neurons for feedforwardnet cascadeforwardnet, and fitnet with different training functions (trainbr, trainlm, trainscg, and trainrp)
   - ➢ Run different ANN models with optimal hidden neurons for several times
   - ➢ Compare the average of MSE, MAE, and RMSE of ANNs
   - ➢ Select the ANN model with the lowest amounts of error
4. Fit the logistic regression model on train data $\rightarrow (\hat{\beta}_k \ (K = 0, \ldots, k))$;

//Loan assessment
5. for: $i = 1:l$
6. Apply the best ANN model for test data to predict the return of open listings:
   - ➢ Best ANN (listing(i)) $\rightarrow$ NPV(i);
7. Apply coefficients found by the fitted logistic regression for test data to predict the risk of open listings:
   - ➢ $\frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\cdots+\beta_{19} x_{19})}} \rightarrow \mathrm{PD}(i)$;
8. end

//Portfolio optimization
9. MIN $\sigma^2 = \sum_{i=1}^{l} \lambda_i^2 \mathrm{PD}_i$;
10. Subject to:
11. $\sum_{i=1}^{l} \lambda_i \mathrm{NPV}_i \geq R^*$;
12. $\sum_{i=1}^{l} \lambda_i = 1$
13. $m \leq \lambda_i M \leq$ amount of listing$_i$
14. $\lambda_i \geq 0$

Prog (m, M, R^* $\{\mathrm{PD}_i, \mathrm{NPV}_i\}_{i=1}^{l}$) $\rightarrow \lambda_i$ $\rightarrow$ Create a portfolio

TABLE 2 The return and risk of the selected portfolio generated by the proposed model

Portfolio return	**2273.95**
Portfolio risk	0.004

TABLE 3 Performance measures for all investment models

	Our proposed model	**Traditional model**
Portfolio return	2273.95	1955.81
Portfolio risk	0.004	0.006

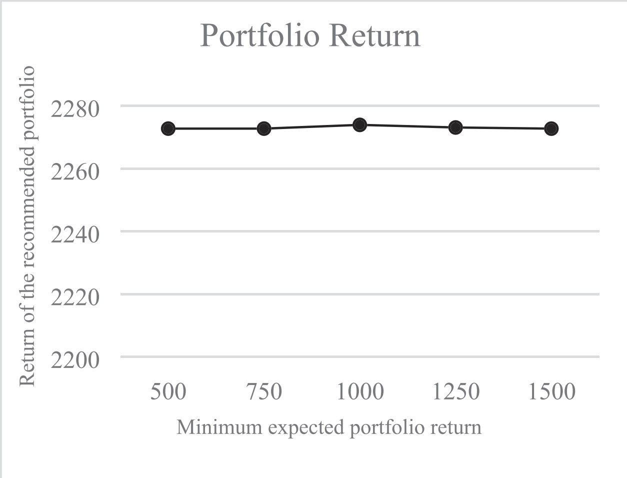

FIGURE 5 The portfolio return sensitivity to R^*

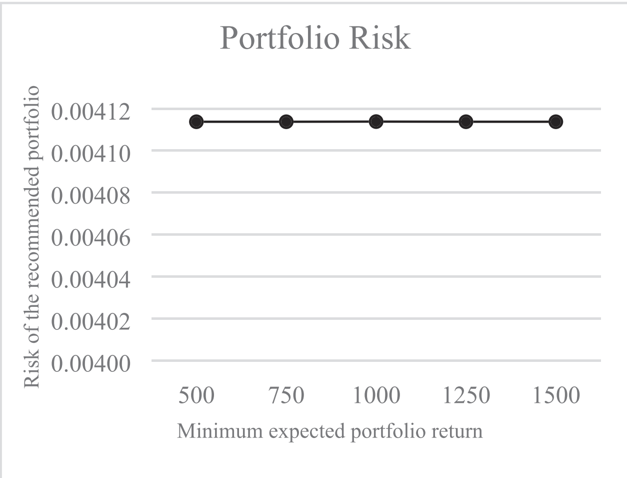

FIGURE 6 The portfolio risk sensitivity to R^*

5 | SENSITIVITY ANALYSIS

Sensitivity analysis is the study of how the uncertainty in the outputs of a mathematical model can be allocated to different sources of uncertainty in inputs. In this study, we set various R^* in our model to check the uncertainty of the proposed model. Figures 5 and 6 represent the outputs of the investment decision model for different amounts of R^*. It is clear that though R^* changes, the risk and return of the portfolio remain almost stable, which means that the model is not sensitive to R^*.

6 | CONCLUSION

In this study, we proposed an investment recommendation model to help investors with finding the proportions of the total investment amounts that they should allocate to available loans for investment in the P2P lending marketplace. Our proposed model has three advantages. First, it can quantify the risk and return of each new loan. This feature of our model provides an easy way to compare loans with each other. Second, we compare three ANN models based on MSE, MAE, and RMSE to find the best ANN to predict the return of open listings, so an optimal ANN model that leads to the lowest level of error is used in our proposed decision-making algorithm. Finally, considering the advantages of using NPV and the availability of few studies that have used NPV as a dependent variable in the P2P lending context, our model improves the P2P investment decision-making models in this market. We utilized ANNs and logistic regression to evaluate loans. After analysing loans, a portfolio optimization problem was solved by the use of mathematical programming. As demonstrated by the experimental results on a real-world data set, the portfolio obtained by the use of our proposed model is better than the portfolio recommended by a simple traditional method. Therefore, investors can improve their investments concerning return and risk using the proposed algorithm in this study. The results of the sensitivity analysis revealed that the risk and return of the portfolio are not sensitive to different amounts of R^*.

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

DECLARATIONS OF INTEREST

None.

ORCID

Golnoosh Babaei https://orcid.org/0000-0001-6132-2271
Shahrooz Bamdad https://orcid.org/0000-0002-3217-1109

REFERENCES

Ahmad AL-Allaf, O. N. (2012). Cascade-forward vs. function fitting neural network for improving image quality and learning time in image compression system. *In Proceedings of the World Congress on Engineering*, *2*, 4–6.

Asadi, S., Hadavandi, E., Mehmanpazir, F., & Nakhostin, M. M. (2012). Hybridization of evolutionary Levenberg–Marquardt neural networks and data pre-processing for stock market prediction. *Knowledge-Based Systems*, *35*, 245–258. https://doi.org/10.1016/j.knosys.2012.05.003

Babaei, G., & Bamdad, S. (2020). A multi-objective instance-based decision support system for investment recommendation in peer-to-peer lending. *Expert Systems with Applications*, *150*, 113278. https://doi.org/10.1016/j.eswa.2020.113278

Badde, D. S., Gupta, A., & Patki, V. K. (2013). Cascade and feed forward back propagation artificial neural network models for prediction of compressive strength of ready mix concrete. *IOSR Journal of Mechanical and Civil Engineering*, *3*(1), 1–6.

Bastani, K., Asgari, E., & Namavari, H. (2019). Wide and deep learning for peer-to-peer lending. *Expert Systems with Applications*, *134*, 209–224. https://doi.org/10.1016/j.eswa.2019.05.042

Bensic, M., Sarlija, N., & Zekic-Susac, M. (2005). Modelling small-business credit scoring by using logistic regression, neural networks and decision trees. *Intelligent Systems in Accounting, Finance and Management*, *13*(3), 133–150https://doi.org/10.1002/isaf.261

Byanjankar, A., Heikkilä, M., & Mezei, J. (2015). Predicting credit risk in peer-to-peer lending: A neural network approach. *2015 IEEE Symposium Series on Computational Intelligence*, 719–725. https://doi.org/10.1109/SSCI.2015.109

Chawla, N. V., Japkowicz, N., & Kotcz, A. (2004). Editorial: special issue on learning from imbalanced data sets. *ACM SIGKDD Explorations Newsletter*, *6*, 1–6. https://doi.org/10.1145/1007730.1007733

Cho, P., Chang, W., & Song, J. W. (2019). Application of instance-based entropy fuzzy support vector machine in peer-to-peer lending investment decision. *IEEE Access*, *7*, 16925–16939. https://doi.org/10.1109/ACCESS.2019.2896474

Emekter, R., Tu, Y., Jirasakuldech, B., & Lu, M. (2015). Evaluating credit risk and loan performance in online peer-to-peer (P2P) lending. *Applied Economics*, *47*(1), 54–70. https://doi.org/10.1080/00036846.2014.962222

Enke, D., & Thawornwong, S. (2005). The use of data mining and neural networks for forecasting stock market returns. *Expert Systems with Applications*, *29*(4), 927–940. https://doi.org/10.1016/j.eswa.2005.06.024

Fernández-Rodríguez, F., González-Martel, C., & Sosvilla-Rivero, S. (2000). On the profitability of technical trading rules based on artificial neural networks: evidence from the Madrid stock market. *Economics Letters*, *69*(1), 89–94. https://doi.org/10.1016/S0165-1765(00)00270-6

Gardner, M., & Dorling, S. (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment*, *32*(14–15), 2627–2636. https://doi.org/10.1016/S1352-2310(97)00447-0

Giannopoulos, V., & Aggelopoulos, E. (2019). Predicting SME loan delinquencies during recession using accounting data and SME characteristics: the case of Greece. *Intelligent Systems in Accounting, Finance and Management*, *26*(2), 71–82https://doi.org/10.1002/isaf.1456

Guo, Y., Zhou, W., Luo, C., Liu, C., & Xiong, H. (2016). Instance-based credit risk assessment for investment decisions in P2P lending. *European Journal of Operational Research*, *249*(2), 417–426. https://doi.org/10.1016/j.ejor.2015.05.050

Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., & Bing, G. (2017). Learning from class-imbalanced data: review of methods and applications. *Expert Systems with Applications*, *73*, 220–239. https://doi.org/10.1016/j.eswa.2016.12.035

Han, H., Wang, W.-Y., & Mao, B.-H. (2005). Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. https://doi.org/10.1007/11538059_91

Hopkinson, B. M. (2016). The case for project net present value (NPV) and NPV risk models. *PM World Journal*, 1–10.

Joanes, D. N. (1993). Reject inference applied to logistic regression for credit scoring. *IMA Journal of Management Mathematics*, *5*(1), 35–43. https://doi.org/10.1093/imaman/5.1.35

Karellas, S., Boukis, I., & Kontopoulos, G. (2010). Development of an investment decision tool for biogas production from agricultural waste. *Renewable and Sustainable Energy Reviews*, *14*(4), 1273–1282. https://doi.org/10.1016/j.rser.2009.12.002

Klafft, M. (2008). Peer to peer lending: auctioning microcredits over the internet. Proceedings of the International Conference on Information Systems, Technology and Management.

Longadge, R., & Dongre, S. (2013). Class imbalance problem in data mining review. *International Journal of Computer Science and Network*, *2*(1), 1305–1707. Retrieved from. http://arxiv.org/abs/1305.1707

Loyola-González, O., Martínez-Trinidad, J. F., Carrasco-Ochoa, J. A., & García-Borroto, M. (2016). Study of the impact of resampling methods for contrast pattern based classifiers in imbalanced databases. *Neurocomputing*, *175*, 935–947. https://doi.org/10.1016/j.neucom.2015.04.120

Magni, C. A. (2005). Investment decisions, net present value and bounded rationality. Retrieved from http://mpra.ub.uni-muenchen.de/7451/

Magni, C. A. (2013). The internal rate of return approach and the AIRR paradigm: a refutation and a corroboration. *The Engineering Economist*, *58*(2), 73–111. https://doi.org/10.1080/0013791X.2012.745916

Malekipirbazari, M., & Aksakalli, V. (2015). Risk assessment in social lending via random forests. *Expert Systems with Applications*, *42*(10), 4621–4631. https://doi.org/10.1016/j.eswa.2015.02.001

Mild, A., Waitz, M., & Wöckl, J. (2015). How low can you go?—Overcoming the inability of lenders to set proper interest rates on unsecured peer-to-peer lending markets. *Journal of Business Research*, *68*(6), 1291–1305. https://doi.org/10.1016/j.jbusres.2014.11.021

Napierała, K., & Stefanowski, J. (2015). Addressing imbalanced data with argument based rule learning. *Expert Systems with Applications*, *42*(24), 9468–9481. https://doi.org/10.1016/j.eswa.2015.07.076

Puro, L., Teich, J. E., Wallenius, H., & Wallenius, J. (2010). Borrower decision aid for people-to-people lending. *Decision Support Systems*, *49*(1), 52–60. https://doi.org/10.1016/j.dss.2009.12.009

Puška, A., Beganovic, A., & Šadic, S. (2018). Model for investment decision making by applying the multi-criteria analysis method. *Serbian Journal of Management*, *13*(1), 7–28. https://doi.org/10.5937/sjm13-12436

Serrano-Cinca, C., & Gutiérrez-Nieto, B. (2016). The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (P2P) lending. *Decision Support Systems*, *89*, 113–122. https://doi.org/10.1016/j.dss.2016.06.014

Sun, T., & Vasarhelyi, M. A. (2018). Predicting credit card delinquencies: an application of deep neural networks. *Intelligent Systems in Accounting, Finance and Management*, *25*(4), 1–16. https://doi.org/10.1002/isaf.1437

Trinkle, B. S., & Baldwin, A. A. (2016). Research opportunities for neural networks: the case for credit. *Intelligent Systems in Accounting, Finance and Management*, *23*(3), 240–254. https://doi.org/10.1002/isaf

Xia, Y., Liu, C., & Liu, N. (2017). Cost-sensitive boosted tree for loan evaluation in peer-to-peer lending. *Electronic Commerce Research and Applications*, *24*, 30–49. https://doi.org/10.1016/j.elerap.2017.06.004

Zekic-Susac, M., Sarlija, N., & Bensic, M. (2004). Small business credit scoring: a comparison of logistic regression, neural network, and decision tree models. In 26th International Conference on Information Technology Interfaces, 265–270.

Zhao, H., Liu, Q., Wang, G., Ge, Y., & Chen, E. (2016). Portfolio selections in P2P lending. KDD'16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2075–2084. https://doi.org/10.1145/2939672.2939861

How to cite this article: Babaei G, Bamdad S. A neural-network-based decision-making model in the peer-to-peer lending market. *Intell Sys Acc Fin Mgmt*. 2020;27:142–150. https://doi.org/10.1002/isaf.1480