Credit Risk Assessment in the Banking Sector Based on Neural Network Analysis

Vera Ivanyuk[1,2(✉)] , Egor Slovesnov[1], and Vladimir Soloviev[1]

[1] Financial University Under the Government of the Russian Federation, Moscow, Russia
[2] Bauman Moscow State Technical University, Moscow, Russia

Abstract. The present research explores the possibility of using neural networks to predict credit risk in the banking sector through a case study of a database of one of the American banks. Scoring is a mathematical or statistical model used by a bank to determine, based on the credit history of "past" clients, how likely it is that a particular potential borrower will repay the loan on time. A scoring model is a weighted sum of certain characteristics. The result is an integrated parameter (score); the higher it is, the more reliable the client is, and the bank can order the clients according to their level of creditworthiness in increasing order.

The integrated parameter of each client is compared with a certain numerical threshold, or boundary line, which is essentially a break-even line and is obtained from the reckoning of the average number of clients paying on time needed to compensate for losses from a single debtor. Clients with an integrated parameter above this line are given credit, while clients with an integrated parameter below this line are not.

Theoretical aspects of the neural network application were considered. A basic table of real data on the bank's clients was studied. Based on the results of the study, conclusions were made that helped solve the problem of building a neural network.

Keywords: Credit risk · Neural network · Assessment

1 Introduction

To assess credit risk, the borrower's creditworthiness is analyzed. In banking practice, creditworthiness is interpreted as a desire combined with the ability to repay the issued obligation in a timely manner. According to this definition, the main goal of scoring is not only to find out whether the client is able to pay back the loan or not but also to examine the degree of client's reliability and commitment.

In the banking system, when a person applies for a loan, the bank may have the following information to analyze:

- the questionnaire filled out by the borrower
- information on this borrower from the credit bureau, an organization that stores the credit histories of the entire adult population of the country
- the borrower's account history, if he or she is the bank's client.

Credit analysts use the following concepts: clients' "attribute-characteristics" (in terms of mathematical logic—variables, factors) and "grade-values" that a variable takes. In the questionnaire that the client fills out, the characteristics are represented by the questions (age, marital status, profession), and the grade-values are the answers to these questions.

In its simplest form, a scoring model is a weighted sum of certain characteristics. The result is an integrated parameter (score); the higher it is, the more reliable the client is, and the bank can order the clients according to their level of creditworthiness in increasing order.

The integrated parameter of each client is compared with a certain numerical threshold, or boundary line, which is essentially a break-even line and is obtained from the reckoning of the average number of clients paying on time needed to compensate for losses from a single debtor. Clients with an integrated parameter above this line are given credit, while clients with an integrated parameter below this line are not.

Currently, it is customary to distinguish four areas of scoring:

1. Application scoring—models for evaluating the financial status of an entity to decide on the feasibility of a transaction
2. Behavioural scoring—models for evaluating the financial status of an entity in the process of implementing a transaction
3. Collection scoring—models for building relationships with entities in high-risk transactions
4. Fraud scoring—models for building relationships with entities to minimize non-financial (in particular, legal) transaction risks.

The first credit scoring models were developed by Fair Isaac Corporation more than half a century ago. The scores resulting from these models are named after the company—FICO. Now the FICO score is widely known and massively used in the United States and Canada when making decisions about issuing loans. The FICO score is calculated based on information from the three largest national credit bureaus: Experian, Equifax, and TransUnion. Depending on the credit bureau whose data is used for calculation, the credit score varies slightly.

The FICO score ranges from 300 to 850. A higher score, as in most other models, corresponds to lower risks. It should be noted that determining the threshold for screening applications that will not be satisfied requires additional efforts. There is no strictly defined procedure and the choice of this feature depends on the bank's strategy: what risks the bank is willing to accept, how much it seeks to expand its loan portfolio, etc.

2   Credit Risk Assessment in the Banking Sector

Currently, the main algorithms used in credit scoring models are:

- logistic regression
- neural networks

- decision trees (and their ensembles such as random forests and gradient boosting).

The use of various machine learning methods, such as neural networks, logistic regression, random forests, etc. in credit scoring has been considered in many papers.

So in the paper West, D. [1] the use of 5 different neural network architectures for credit scoring tasks was considered and benchmarked against classical statistical methods including discriminant analysis, logistic regression and non-parametric methods. The results of the study showed that the use of artificial neural networks could significantly improve the quality of classification.

Boguslauskas, V., Mileris, R. concluded in their work [2] that neural networks and logistic regression appeared to be the most effective models for solving the problem of credit scoring. Their analysis showed that artificial neural networks were superior to other methods in terms of prediction accuracy.

The paper Pawel Plawiak, Moloud Abdar, Joanna Plawiak, Vladimir Makarenkov [3] also considered the credit scoring application of neural networks. The authors used a genetic algorithm of a deep learning neural network on a small data set, which resulted in outperforming traditional scoring algorithms.

The purpose of the present research is to study and summarize theoretical and practical issues of using neural networks in the banking sector for credit risk analysis.

The information base is a table of clients of an American bank with data for the period 2011–2015. The database contains information about more than 800,000 clients including 74 characteristics for each client (see Fig. 1).

	id	member_id	loan_amnt	funded_amnt	funded_amnt_inv	term	int_rate	installment	grade	sub_grade
0	1077501	1296599	5000.0	5000.0	4975.0	36 months	10.65	162.87	B	B2
1	1077430	1314167	2500.0	2500.0	2500.0	60 months	15.27	59.83	C	C4
2	1077175	1313524	2400.0	2400.0	2400.0	36 months	15.96	84.33	C	C5
3	1076863	1277178	10000.0	10000.0	10000.0	36 months	13.49	339.31	C	C1
4	1075358	1311748	3000.0	3000.0	3000.0	60 months	12.69	67.79	B	B5

5 rows × 74 columns

Fig. 1. Information base.

The objective of the study is to build a neural network using real data to assess credit condition (good, bad).

Formation of requirements to the model. Below we list some important clients' attribute-characteristics:

1. ID. Borrower's identification number
2. Loan_amnt. Amount of the loan requested by the borrower
3. Funded_amnt. Amount of the loan issued

4. Term. Period for which the loan was issued
5. Int_rate. Interest rate of the loan
6. Installment. Amount of regular payment
7. Grade and Sub_grade. Score of the borrower's reliability
8. Emp_length. Borrower's employment length
9. Home_ownership. The form of housing tenure of the borrower (own, rent, mortgage)
10. Annual_inc. Borrower's annual income
11. Loan_status. Current status of the loan (current, fully paid, late).
12. Issue_d. Loan issue date
13. Purpose. The purpose provided by the borrower for the loan request (car, business, educational).

Let us look at the amounts of loans requested by the clients and what loans were issued overall and by year (see Fig. 2).
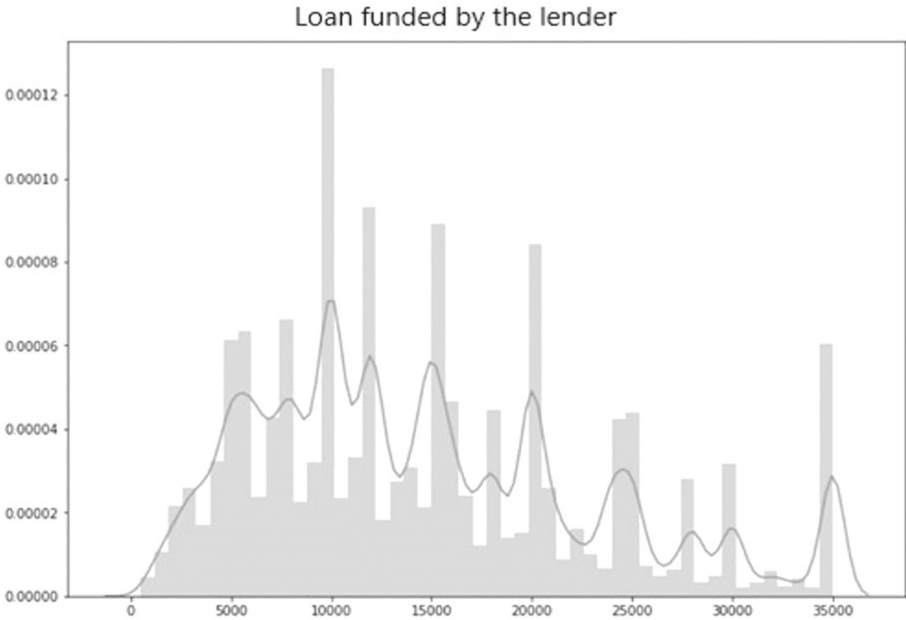

Fig. 2. Amount of loans issued

We also consider the average amount of loans issued by year (see Fig. 3):

According to the charts, most of the loans issued were in the range of $10,000 to $20,000. We can also note a steady increase in the average amount of the loans issued.
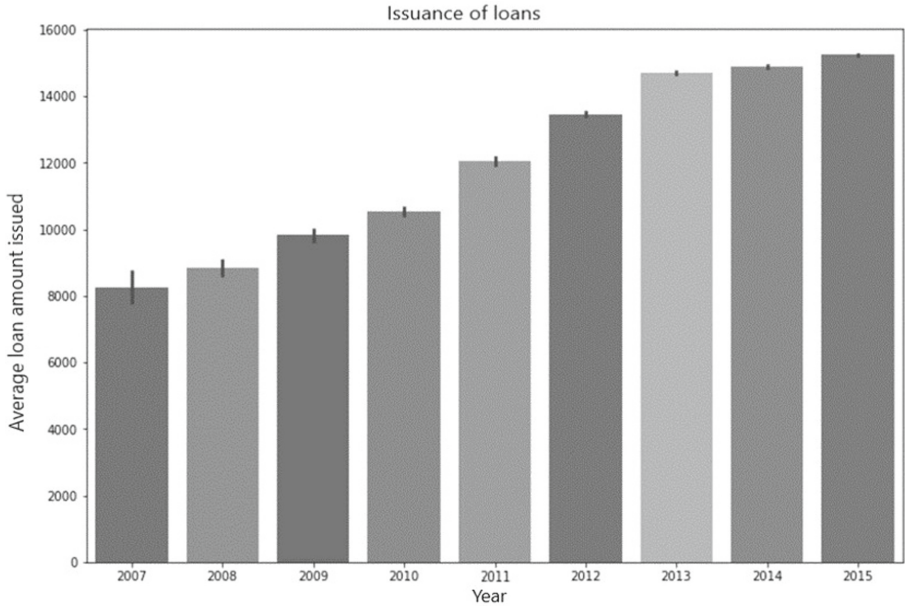
Fig. 3. Average amount of loans issued by year.

Let us consider the number of loans by their status:

1. Current – 601779
2. Fully paid – 207723
3. Charged off – 45248
4. Late (31–120 days) – 11591
5. Issued – 8460
6. In Grace Period – 6253
7. Late (16–30 days) – 2357
8. Does not meet the credit policy. Status: Fully Paid – 1988
9. Default – 1219
10. Does not meet the credit policy. Status: Charged Off – 761.

We shall consider Charged Off, Default and Late (in any stage) as bad loans. Now let us look at the ratio of good loans to bad loans, as well as their number by year (see Fig. 4 and 5).

Two important conclusions can be drawn from these charts. First, bad loans comprise only 7.6% of all loans issued. Second, it is important to remember that the database contains a lot of current loans, which may become bad and somewhat affect the quality of the network.

Let us consider the importance of the borrower's credit score. In order to understand exactly how the grade of the credit score affects the final risk, we need to consider the number of bad loans against the borrower's score. Let us plot the number of loans issued depending on the borrower's credit score (see Fig. 6).
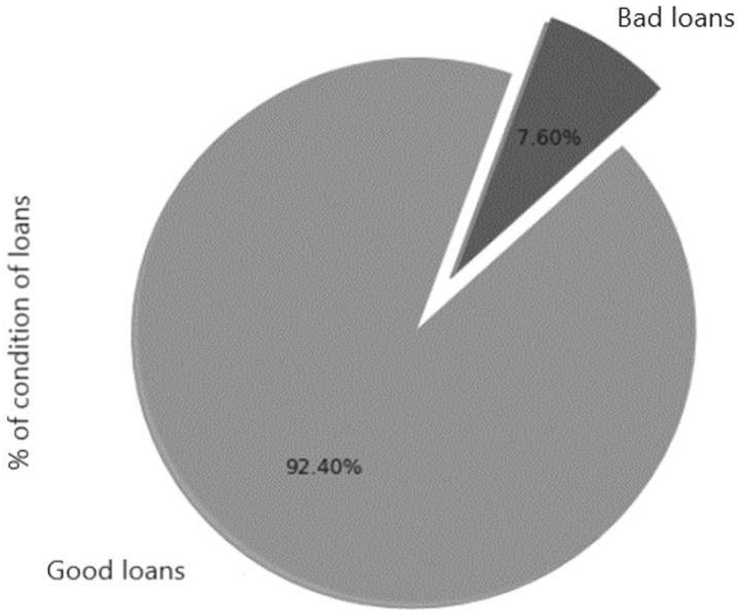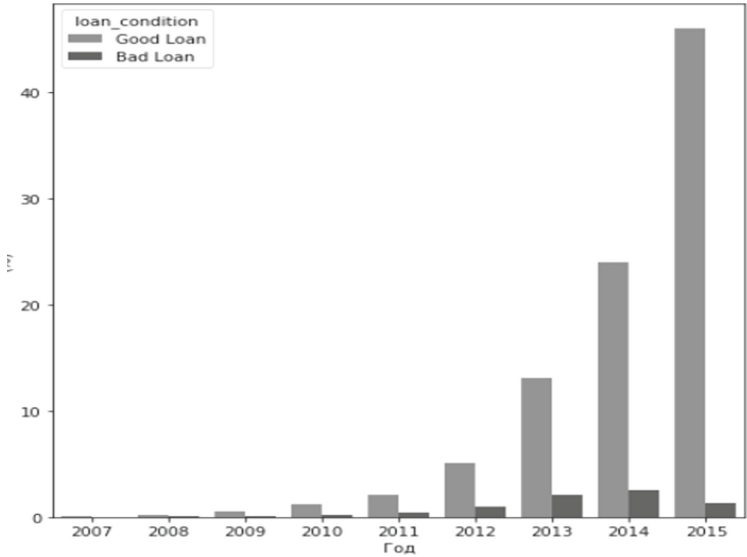
Fig. 4. The ratio of good loans to bad loans

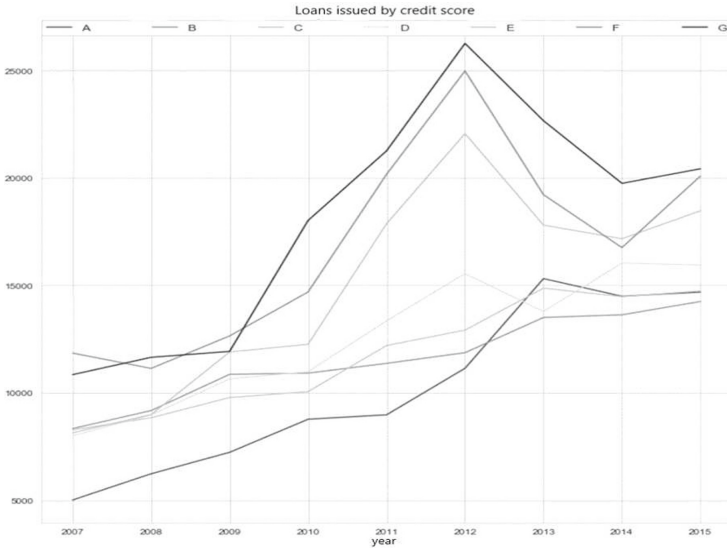**Fig. 5.** Percentage of good and bad loans by year

Fig. 6. Number of loans issued

Conclusions on the impact of credit score:

- The scores that had a lower grade received a larger amount of loans in comparison to the higher grade of credit score. This contributes to a higher level of risk for the bank as a whole.
- The interest rate increases as the grade deteriorates.
- Most bad loans were issued to borrowers with a grade of "B".

Let us explore the reasons why a loan becomes bad. Logically, it can be assumed that the borrower's credit score and annual income will have the greatest impact on the level of credit risk. We will identify factors that increase the risk of loan default, such as low annual income, high interest rate, and low grade of the credit score. Let us build a correlation heatmap based on numerical variables (see Fig. 7):

Let us plot the amounts of bad loans with a breakdown by condition (see Fig. 8).

According to this plot, bad loans tended to decline by 2015.

3   Analysis of the Results Obtained

Let us describe the structure of the neural network that will be used for prediction [4–6]. The neural network will consist of input neurons, two output layers, and two hidden layers, with 66 neurons in each. For the research, a feedforward network will be used. The activation function will be the ReLu function [7,8].

It is the most convenient function which often performs better than others. Schematically, the neural network has the form (see Fig. 9):
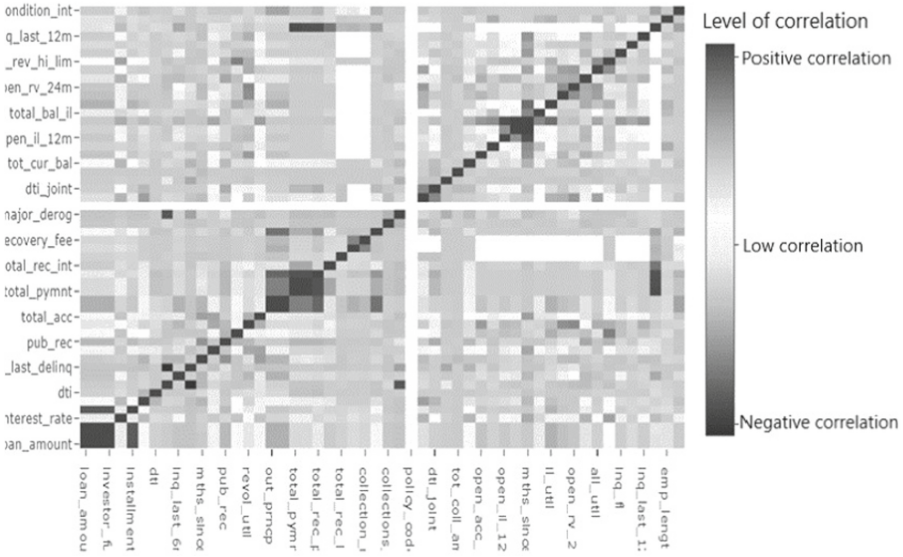
Fig. 7. Correlation heatmap.

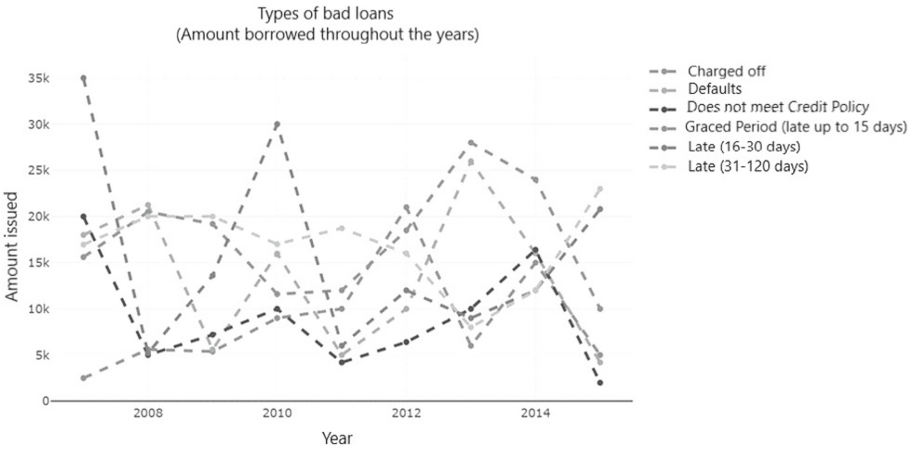**Fig. 8.** Amount of bad loans by condition.

Input Layer ∈ $\mathbf{R}^2$ Hidden Layer ∈ $\mathbf{R}^{10}$ Hidden Layer ∈ $\mathbf{R}^{10}$ Output Layer ∈ $\mathbf{R}^2$
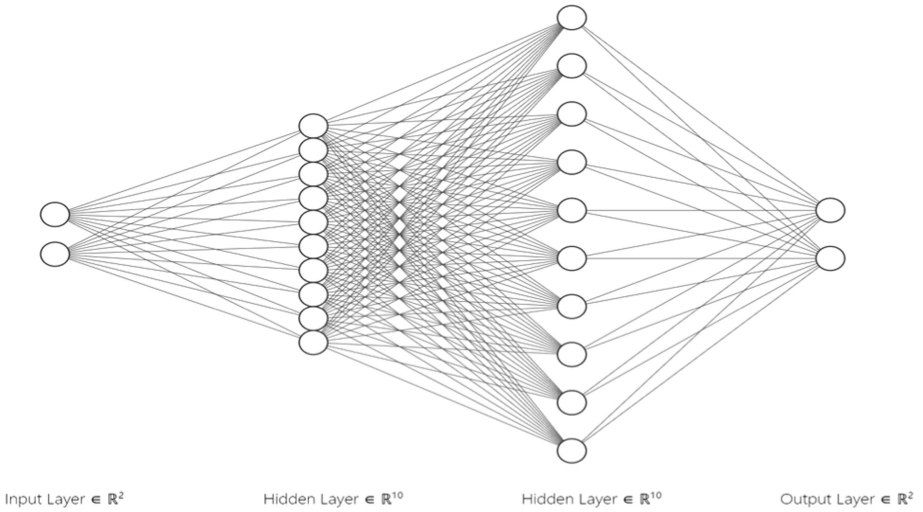
Fig. 9. Neural network scheme.

The scheme shows only 10 neurons in hidden layers for visual clarity. Also, the number of inputs in the scheme is two, but in fact, there will be more of them, equal to the number of attribute-characteristics.

Since there is only 9% of bad loans in the source data, the dataset can be considered unbalanced. Specifically for such cases, an algorithm named Synthetic Minority Over-Sampling Technique (SMOTE) was developed to improve the accuracy of predictions. We explain the principle of operation by giving an example. Assume that the total number of loans is D_0, then the number of good loans is S_0 and the number of bad loans is B_0. Consequently, $D_0 = S_0 + B_0$. Since the data is highly unbalanced, i.e. $S_0 \gg B_0$, we will increase the percentage of bad loans, as shown in Table 1.

Table 1. Example of SMOTE operation.

	"Good loans"		Bad loans		Total
	Quantity	Percentage $(\times 100\%)$	Quantity	Percentage (100%)	Quantity
Original data	S_0	S_0/D_0	B_0	B_0/D_0	D_0
SMOTE	$S_1 = S_0$	S_0/D_1	$B_1 = S_0$	S_0/D_1	D_1

It is important to note that instead of using existing data, SMOTE generates new rows by combining the characteristics of the target class with those of its neighbours.

The number of layers, the number of neurons per layer, and the learning rate were selected experimentally [9]. The learning rate is a setting parameter

in the optimization algorithm that determines the step size at each iteration when moving to the minimum of the loss function [10,11]. The loss function is a function that, in statistical decision theory, characterizes the loss associated with incorrect decision-making based on observed data.

Using the Tensorboard library, we will create the final scheme of the neural network (see Fig. 10).
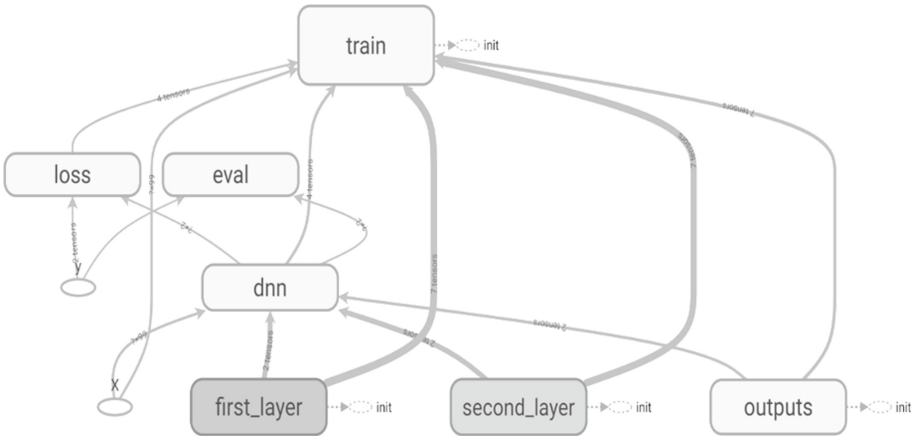

Fig. 10. Neural network scheme.

Assessment of model prediction accuracy. To assess the accuracy of the model, we compare it with the prediction obtained using logistic regression on the same data. The performance of the developed model is compared with the results of alternative models below (see Table 2)

Table 2. Performance comparison of different prediction methods

Prediction method		Prediction accuracy (%)
Parametric statistical methods	Logistic regression	84,5
	Linear discriminant analysis [12]	88,1
Nonparametric statistical methods	Decision tree (CRT) [13]	71,2
	Decision tree (CHAID) [13]	70,1
Support vector (SNN)	SVM in IBM SPSS modeler [13]	72,1
	Neural network with 1 hidden layer [13]	71,2
Deep neural networks (DNN)	Neural network with 2 hidden layers (our study)	92,4

The table shows that the accuracy of the neural network exceeds the accuracy of other prediction methods.

4   Conclusion

The study was conducted using a neural network with 2 hidden layers, with 66 neurons in each layer 22 categorical and numeric attribute-characteristics were used to create the model, and loans were divided into two classes: good and bad. Since categorical data cannot be used to build a neural network, it was converted to numeric data using the One-Hot Encoding algorithm 887379 observations were used to train the model. Due to the domination of good loans among all observations, SMOTE was used to increase the number of bad loans and thus balance the data. After training the neural network, the prediction accuracy comprised 0.92, which exceeds the results of other prediction methods.

References

1. West, D.: Neural network credit scoring models. Comput. Oper. Res. **27**(11–12), 1131–1152 (2000)
2. Boguslauskas, V., Mileris, R.: Estimation of credit risk by artificial neural networks models. Eng. Econ. **64**(4) (2009)
3. Plawiak, P., Abdar, M., Plawiak, J., Makarenkov, V., Acharya, U.R.: DGHNL: a new deep genetic hierarchical network of learners for prediction of credit scoring. Inf. Sci. **516**, 401–418 (2020)
4. Eliana, A., di Tollo, G., Roli, A.: A neural network approach for credit risk evaluation. Q. Rev. Econ. Financ. **48**(4), 733–755 (2008)
5. Ivanyuk, V., Tsvirkun, A.: Intelligent system for financial time series prediction and identification of periods of speculative growth on the financial market. IFAC Proc. Vol. **46**(9), 1128–1133 (2013)
6. Koroteev, M.V., Terelyanskii, P.V., Ivanyuk, V.A.: Approximation of series of expert preferences by dynamical fuzzy numbers. J. Math. Sci. **216**, 5692–695 (2016)
7. Chuang, C.-L., Huang, S.-T.: A hybrid neural network approach for credit scoring. Expert Syst. **28**(2), 185–196 (2011)
8. Lee, T.-S., et al.: Credit scoring using the hybrid neural discriminant technique. Expert Syst. Appl. **23**(3), 245–254 (2002)
9. Khemakhem, S., Said, F.B., Boujelbene, Y.: Credit risk assessment for unbalanced datasets based on data mining, artificial neural network and support vector machines. J. Model. Manage. **13**, 932–951 (2018)
10. Oreski, S., Oreski, D., Oreski, G.: Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment. Expert Syst. Appl. **39**(16), 12605–12617 (2012)
11. Wang, S., Yin, S., Jiang, M.: Hybrid neural network based on GA-BP for personal credit scoring. In: 2008 Fourth International Conference on Natural Computation, vol. 3. IEEE (2008)
12. Feis, A., et al.: P2P loan selection. Stanford Univesity Algorithmic Trading and Big Financial Data MS&E, p. 448 (2016)
13. Jin, Y., Zhu, Y.: A data-driven approach to predict default risk of loan for online peer-to-peer (P2P) lending. In: 2015 Fifth International Conference on Communication Systems and Network Technologies. IEEE (2015)