APPLICATION OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES FOR CREDIT RISK ASSESSMENT

Muhamedieva D.T.
Egamberdiev N.A.
Xolmuminov O.T.

Abstract. *In the conditions where the information representing the creditworthiness of bank customers has a large volume and features of uncertainty, it is necessary to identify the hidden relationship between the data, the specific laws of predicting the course of the processes, classification, intellectualization of the studied process are important issues of analysis. Therefore, it is important to determine and evaluate the creditworthiness of bank customers through neural network algorithms. The principles and methods of intelligent data analysis, prediction, evaluation, and object-oriented programming were used in the research process. In the research work, a model for assessing the creditworthiness of bank customers is proposed; it is proposed to build a classification and evaluation model of intellectual analysis of creditworthiness of bank customers based on a multi-layer neural network algorithm; an algorithm for determining and evaluating the creditworthiness of bank customers was developed using multilayer neural networks.*

Keywords: *artificial intelligence, neural networks, risk, credit, bank, customer, classification, evaluation, activation function.*

ПРИМЕНЕНИЕ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ОЦЕНКИ КРЕДИТНОГО РИСКА

Аннотация. *В условиях, когда информация, отражающая кредитоспособность клиентов банка, имеет большой объем и черты неопределенности, необходимо выявление скрытой связи между данными, специфические закономерности прогнозирования протекания процессов, классификация, интеллектуализация изучаемых процесса являются важными вопросами анализа. Поэтому важно определять и оценивать кредитоспособность клиентов банка с помощью нейросетевых алгоритмов. В процессе исследования использовались принципы и методы интеллектуального анализа данных, прогнозирования, оценки и объектно-ориентированного программирования. В исследовательской работе предложена модель оценки кредитоспособности клиентов банка; предлагается построить классификационно-оценочную модель интеллектуального анализа кредитоспособности клиентов банка на основе многослойного нейросетевого алгоритма; разработан алгоритм определения и оценки кредитоспособности клиентов банка с использованием многослойных нейронных сетей.*

Ключевые слова: *искусственный интеллект, нейронные сети, риск, кредит, банк, клиент, классификация, оценка, функция активации.*

INTRODUCTION

The decisive role of the mortgage market in causing the financial crisis as a result of the pandemic has led to an increase in academic research on bank regulation and credit risk modeling. Banks spend significant resources developing internal credit risk models to more accurately determine expected credit losses and assign required economic capital. Rigorous analysis of credit risk is important not only for lenders and banks, but also for the proper

formulation and regulation of economic policy, as it provides a check on the financial system and a healthy economy in general. One of the main practices of banking institutions is to lend money to their customers. A common reason customers borrow is to finance their home purchase. Despite the fact that these future homeowners are looking for banks that offer the lowest interest rates, banks will lend to customers who can meet their financial obligations. In order for banks to weigh the default risk of their future borrowers, they collect information on borrowers and mortgages. The results of this collected data are called credit worthiness of applicants. These applicants will then receive a credit score rating. Banks then use the collected information to decide whether to lend or not [1-3].

In order to assess the bank's credit risk, the following issues must be resolved: the level of default, the probability of default and the damage caused. Currently, credit risk determination in developed countries is carried out on the basis of artificial intelligence models and algorithms. This serves to increase the level of accuracy and reduce the human factor [4-5].

The purpose of the research work is to develop an algorithm and a software tool for determining the creditworthiness of bank customers based on neural networks.

To achieve the goal, the following research tasks are set:

analysis of existing methods of assessing the creditworthiness of a bank client;

identify the problems of developing a model for assessing the creditworthiness of a bank client;

formulation of the general issue of applying classification algorithms in creditworthiness assessment systems of bank clients;

solving the problems of assessing the bank client's creditworthiness based on neural networks;

development of an algorithm for assessing the creditworthiness of a bank client;

development of the functional structure and software tool of the bank client's creditworthiness evaluation software tool;

conducting computational experiments to evaluate the effectiveness of developed algorithms and programs and analyzing the obtained results.

In our study, the advantages and disadvantages of algorithms for assessing the creditworthiness of a bank customer were analyzed.

K-nearest neighbor classifiers (KNN) are based on similarity based learning. Given an unknown sample, the KNN classifier searches the template space for the closest KNN to the unknown sample. Proximity is determined by distance. An unknown sample is assigned to the most common class among KNN. The main advantage of this approach is that no pre-classification model specification is required. The disadvantage of KNN is that it does not generate a simple classification probability formula and its prediction accuracy is greatly affected by the distance measure and the degree of cardinality k of the neighbor [6-7].

Logistic regression can be considered a special case of linear regression models. But a binary response variable violates the usual assumptions of general regression models. A logistic regression model specifies that the appropriate function of the fitted probability of an event is a linear function of the observed values of the available explanatory parameters. The main advantage of this approach is that it can provide a simple classification of probability. Weaknesses are that logistic regression does not work well with problems of nonlinear and interactive effects from discrete variables [8-10].

Discriminant analysis, also known as Fisher's rule, is another method applied to the binary outcome of the response variable. Fisher's rule is a proxy for logistic regression, and the descriptive variables for each response variable class are based on a highly variable normal distribution function with a common coefficient-covariance matrix. The goal of Fisher's rule is to maximize the distance between different groups and minimize the distance within each group. Their pros and cons are similar to logistic regression [11-12].

Bayesian classifier is based on Bayesian theory and asserts that the effect of one attribute value in a given class is independent of the values of other attributes. This assumption is called class conditional independence. Bayesian classifiers are useful in providing a theoretical framework for other classifiers that do not explicitly use Bayes' theorem. The main weakness of the Bayesian classifier is that the prediction accuracy depends on the conditional independence of the class. This assumption makes the calculation easier. In practice, there may be a correlation between the variables [13-15].

Lending evaluation allows to eliminate the risk of influence of manager's subjective opinion on the decision on lending, simplifies the process of lending, helps to further increase the volume of lending. The evaluation model has the following main advantages:

1) shortening the time for making a decision on lending. Increasing the number and speed of consideration of applications by minimizing documents in lending to private clients;

2) Effective assessment and continuous control of the risk level of a particular borrower;

3) reducing the influence of subjective factors when making a decision on lending. Ensuring objectivity in evaluating applications of credit specialists in all branches and offices of the bank;

4) assessment and risk management of the portfolio of loans given to individuals of the bank as a whole, including its branches. Taking into account the profitability level and risk of the loan portfolio when determining the parameters of new loans. There is a wide statistical model for calculating points proposed by representatives of foreign banks. 'based on methods of predicting borrower behavior based on data.

The most powerful in terms of the accuracy of credit risk assessment are models that use a complex approach, i.e. accounting for all data and expert knowledge of bank management. Three main options are used in the construction of assessment models. The first is to create a profile of a specific target customer, and the second is to buy a ready-made model written off from another country (however, based on its experience, the bank should strengthen the requirements for borrowers). The third is a model adapted to the bank of individual characteristics. The third option is often used by strong banks entering the market with new products [16-17].

MATERIALS AND METHODS

To conduct research, it is necessary to use:

implementation of mathematical calculation of the problem, identification of data sources, identification of artificial neural network sources, model failure, creation of neural network sources. Let's formulate the problem mathematically

classification problem, where G = [personal information about the client] is a set of parameters describing a potential borrower, Q = {1, 0} is a single non-distinguished set (to whom credit is granted and to whom credit is denied) . about)).

As input data of the artificial neural network, we use a vector consisting of a set of questionnaires.

Customer data, they are reviewed for loan officers at banks and they make a decision on whether to grant a loan or reject an application. entrepreneur" and others), social status (options: "married", "divorced", "Single"), education ("unknown", "middle", "primary", "high" and others), average annual income, whether there are loans from other organizations ("yes" / "no") and whether payments are preloaded ("yes" / "no").

Training data expressed in $x_{p1}, x_{p2}, \ldots, x_{pm_p} \in X_p, p = \overline{1, r}$. Here $x_{pi} = \left(x_{pi}^1, x_{pi}^2, \ldots x_{pi}^N \right), i = \overline{1, m_p}$, N - considered in the range of dimensional characters, X_p $p = \overline{1, r}$ specifying a set of classes, u $x_{p1}, \ldots, x_{p_{m_p}}$ consists of objects.

Problem: to create a decision rule for identifying an unknown object, that is, to develop a rule to which class belongs to a set of trained options, and mathematical support for neural network algorithms.

Despite the diversity of neural networks, they all have common features. Thus, they all consist of many elements of the same type as the human brain - neurons that mimic the neurons of the brain.

The state of the neuron is determined by the following formula [1].

$$F = \sum_{i=1}^{n} x_i w_i . \tag{1}$$

Here,

n - is the number of input neurons;

x_i - value of input neurons;

w_i - i - synaptic weight.

Then the value of the neuron axon is determined by the following formula [1-2].

$$Y = f(S). \tag{2}$$

Here, f is a function called activation. Often sigmoid is used as an activation function with the following form:

$$f(x) = \frac{1}{1 + e^{-ax}} . \tag{3}$$

The main advantage of this function is that the derivative of the function is represented by the function itself:

$$f'(x) = \alpha f(x)(1 - f(x)). \tag{4}$$

During training, the task of minimizing the objective function of the neural network errors is set to limit the search space found using the least squares method:

$$E(w) = \frac{1}{2} \sum_{j=1}^{p} (y_j - d_j)^2 . \tag{5}$$

Here,

y_i - j - output neuron value;

d_i - j -j the ideal value of the output;

p - is the number of neurons in the output layer.

The neural network is trained using the gradient descent method, which means that at each iteration, the weight is changed according to the following formula:

$$\Box w_{ij} = -\eta \cdot \frac{\partial E}{\partial w_{ij}} . \tag{6}$$

The parameter that determines the speed of learning η is calculated according to the following formula.

$$\partial_j^{(n)} = \left[\sum_k \partial_k^{(n+1)} \cdot w_{jk}^{(n+1)} \right] \cdot \frac{dy_j}{dS_j} . \tag{7}$$

Finding the neural network for the last layer is not difficult, since we know the target vector, that is, the vector of values that the neural network should generate for a given set of input values.

$$\partial_j^{(N)} = \left(y_j^{(N)} - d_j \right) \cdot \frac{dy_j}{dS_j} . \tag{8}$$

Finally, we write the formula (6) in an expanded form.

$$\Delta w_{ij}^{(n)} = -\eta \cdot \delta_j^{(n)} \cdot x_i^n . \tag{9}$$

Now let's look at the complete neural network training algorithm:

Step 1: determine the output values of the network neurons;

Step 2: calculate for the output layer of the neural network according to the formula (8) and calculate the change of weights of n output layers according to the formula (9);

Step 3: calculate formulas (7) and (9) respectively for the remaining layers of the neural network for n = N-1..1;

Step 4: Adjust the weights of all neural networks;

$$w_{ij}^{(n)}(t) = w_{ij}^{(n)}(t-1) + \Delta w_{ij}^{(n)}(t) .$$

Step 5: If the error is significant, go to step 1.

RESULTS

Assessing and determining the creditworthiness of a bank client.

• two dichotomous variables - usual payment (Yes = 1, No = 0) were used as response variables.

• Among the total 25 thousand observations, 5529 (22.12%) are non-paying card holders.

• the following 23 variables were used as descriptive variables:

• X1: Amount of loan granted (NT$): personal consumer loan and its family (additional) loan.

• X2: gender (1 = male, 2 = female).

• X3: Education (1 = Masters; 2 = University; 3 = High School; 4 = Others).

• X4: Marital status (1 = married, 2 = single, 3 = others). X5: Age (years).

• X6-X11: Past payment history. We observed monthly payment records from April to September 2005 as follows: X6 = repayment status in September 2005;

• X7 = return status in August 2005; . . .; X11 = April 2005 return status. The scale of repayment is as follows: 1 = pay on time; 1 = delay payment for one month; 2 = two months late payment;

• 8 = eight months payment delay; 9 = payment arrears of nine months or more. -X17: Tax amount (NT dollars). X12 = amount of budget declaration in September 2005; X13 = August 2005 bill amount;

• X17 = April 2005 bill amount.

• X18-X23: prepayment amount (NT$). X18 = amount of money paid in September 2005; X19 = amount paid in August 2005; . . .; X23 = amount paid in April 2005.

DISCUSSION

Creditworthiness of a client who wants to get a loan is a comprehensive assessment of the client's financial activity, which determines the ability of the borrower to repay the requested loan (principal amount and interest on it) on time and fulfill other debt obligations. to determine the ability to get lay.

At the first stage of lending, the bank must determine the following:

a) reliability and creditworthiness of the borrower, his

continuity and efficiency of the activity as a partner;

b) the validity of the loan order and the level of security of the loan repayment. If necessary, the bank develops its own requirements for the loan offer;

d) it is necessary to find an answer to the question of whether the bank's credit policy of the loan proposal leads to a decrease in the risk of granting a new loan or not.

The research object is a bank client.

The subject of the study is to assess the creditworthiness of the client.

The following methods and approaches are used in the research process:

- Assessing the creditworthiness of a new loan client;

- database;

In addition, the following results were obtained when the algorithm for selecting an effective method for classification was tested with the Blood transfusion dataset (Table 1):

Table 1

Analysis of results

Algorithm name	Xatolik	Vaqt(c)
Logistic	442	1,12
KStar	523	0,007
Decision tree (J48)	489	0,03
Tasodifiy oʻrmon	762	0,058
Neural network	399	0,007

The analysis of the results shows that Decision tree (J48), Logistic and Neural network performed well for this data set, but Neural network method gives better result if we compare with the time parameter.

CONCLUSIONS

The article describes the solution of artificial intelligence problems based on the neural network algorithm. The mathematical apparatus and algorithm of the neural network are presented. Based on the proposed algorithm, a program was written in the Java programming language, the selection of the number of neurons in the layers of the neural network and the adjustment of the parameters for the redistribution of errors in solving the problem of the iris

model were analyzed and the results obtained in several classification algorithms were compared.

The use of neural network models and algorithms in the intellectual analysis of large volumes of data allows for the formation and adoption of more effective decisions.

The scientific significance of the results obtained in the research work is made up of the proposed models and algorithms.

Economic efficiency was achieved based on the generalization and systematization of knowledge about promising artificial intelligence technologies in the banking sector and their application to solving credit risk assessment issues.

REFERENCES

1. V.V. Kruglov, M.I. Dli, and R.Yu. Fuzzy logic and artificial neural networks. - M.: Fizmatlit, 2001.
2. Zagidullin BI, Nagaev IA, Zagidullin N.Sh., Zagidullin Sh.Z. _ A neural network model for the diagnosis of myocardial infarction. // Russian Journal of Cardiology. 2012; (6): 51-54.
3. Han Y., Lam W., Ling C.X. Customized classification learning based on query projections, Information Sciences 177 (2007) 3557–3573.
4. Jie Lu, Guangquan Zhang Da Ruan, Fendjie Wu.Multi-objective group decision Making. Imperial College Press, London, 2007, 390.
5. Mukhamedieva D.T., Egamberdiev N.A., Zokirov J.Sh. Mathematical support for solving the classification problem using neural network algorithms // Turkish Journal of Computer and Mathematics Education. Vol.12 No.10 (2021)
6. Oyang Y.J., Hwang S.C., Ou Y.Y., Chen C.Y., Chen Z.W. Data classification with the radial basis function network based on a novel kernel density estimation algorithm, IEEE Transactions on Neural Networks 16 (1) (2005) 225–236.
7. Peng L., Yang B., et al. (2009). "Data gravitation based classification." Inf. Sci. 179(6): 809-819.
8. Tozan H., Vayvay O. Analyzing Demand Variability Through SC Using Fuzzy Regression and Grey GM(1,1) Forecasting Models, Information Sciences 2007, World Scientific, 2007, pp. 1088-1094.
9. Vityaev E.E., Lapardin K.A., Khamicheva I.V., Proskura A.L. Transcription factor binding site recognition by regularity matrices based on the natural classification method. Intellegent Data Analysis. Special issue: "New Methods in Bioinformatics. Presented at the fifth International Conference on Bioinformatics of Genom Regulation and Structure" eds. Evgenii Vityaev and Nikolai Kolchanov. v.12(5), IOS Press, 2008 pp. 495-512.
10. Алиев Р.А., Алиев Р.Р. Теория интеллектуальных систем и ее применение. - Баку, Изд-во Чашыоглы, 2001. -720 с.
11. Egamberdiyev N.A., Muhamediyeva D.T., Jurayev Z.Sh. Qualitative analysis of mathematical models based on Z-number // Proceedings of the Joint International Conference STEMM: Science – Technology – Education – Mathematics – Medicine. May 16-17, 2019, Tashkent, pp.42-43.
12. Egamberdiev N., Mukhamedieva D. and Khasanov U. Presentation of preferences in multi-criterional tasks of decision-making // IOP Conf. Series:Journal of Physics: Conference Series 1441 (2020) 012137. DOI: https://doi.org/10.1088/1742-6596/1441/1/012137

13. Muhamediyeva D.T. and Egamberdiyev N.A. Algorithm and the Program of Construction of the Fuzzy Logical Model //2019 International Conference on Information Science and Communications Technologies (ICISCT), Tashkent, Uzbekistan, 2019, pp. 1-4.

14. Muhamediyeva D.T., Egamberdiyev N., Bozorov A. Forecasting risk of non-reduction of harvest //Proceedings of the 2nd International Scientific and Practical Conference "Scientific community: Interdisciplinary research". - Hamburg, Germany. 26-28.01.2021. Pp.694-699.

15. Muhamediyeva D.T.,Egamberdiyev N., Xushboqov I.U. Formulation of the problem particle swarm method for solving the global optimization // Proceedings of the 7th International Scientific and Practical Conference "Scientific horizon in the context of social crises". - Tokyo, Japan. 6-8.02.2021. Pp.1076-1082.

16. Mirzayan K., Dilnoz M., Barno S. (2021) The Problem of Classifying and Managing Risk Situations in Poorly Formed Processes. // In: Aliev R.A., Yusupbekov N.R., Kacprzyk J., Pedrycz W., Sadikoglu F.M. (eds) 11th World Conference "Intelligent System for Industrial Automation" (WCIS-2020). WCIS 2020. Advances in Intelligent Systems and Computing, vol 1323. Springer, Cham. Pp 280-286.