Data Mining of Operations with Card Accounts of Bank Clients

Musa A. Subeh
Head of Accounting Section, Bethlehem Municipality, Palestine.

Hanna Yarovenko
PhD, Associate Professor of the Economic Cybernetics Department, Sumy State University, Ukraine.

Abstract

The article is devoted to the expediency of using the data mining and the construction of the neural network for the evaluation of transactions with card accounts for detecting attempts of frauds. The authors proposed a scheme for customer interaction with the bank when transaction is performing with the payment cards. The process is carried out using the verification module with data mining. The article was built a neural network with using software "Statistica". The authors selected a data set that contains amounts of transaction, time intervals, fraud identifiers. As a result, it was got a multilayer perceptron with nine inputs, five hidden neurons and two outputs that can be used to predict an attempt at fraud with card accounts of bank clients.

Keywords: bank, fraud, credit card, data mining, modeling, neural network, Statistica
JEL Classification: C45, D85, G21, L86.

Introduction

Modern economy provides widespread use of non-cash tools for the implementation of various types of payments. The benefits of making such payments to both banks and their customers are significant. But such operations are increasingly becoming objects of criminal and fraudulent actions by third parties that harm the citizens, banks and the economy as a whole. This negative phenomenon is characteristic for many countries of the world, which is connected with the development of information society, the rapid progress in the field of computer technology, increasing access to various data. The latest technologies for a number of reasons are increasingly attracting fraudsters for their use in committing unlawful actions, especially in the banking sector.

Thus, according to the Ukrainian Interbank Association of EMA payment system members, the losses of citizens due to the actions of fraudsters with payment cards in 2016 amounted to 339.13 million UAH, including as a result of telephone fraud (stitching) – 275.45 million UAH, as a result of embezzlement of confidential data (phishing) – 63.68 million UAH. The amount of losses in 2016 exceeds the losses earned in 2015 (181 million UAH) and in 2014 (90 million UAH). The average amount lost by a bank client from fraudulent activities in 2015 reached 800 UAH, and in 2016 it increased to 1,500 UAH. Statistics for 2017 are also unfavorable. So, for only two months – August and September 2017, fraudsters stole about 238 million UAH from card accounts of citizens. Statistical information shows that from year to year the number of fraud increases, their ways are modified. This situation is a rather negative phenomenon for the country as a whole as it undermines the trust of users of services to banks. Banks also display bottlenecks in the system of protection and lose potential customers. It is difficult to improve this situation completely, but there is an urgent need to improve the system of bank protection by introducing more advanced methods and technologies. In our opinion, one of the possible directions for modernizing the security system is the use of data mining of operations with card accounts.

Literature review

Modern mathematical tools are quite versatile and popular in the field of analytics of various economic objects. One of such areas is the data mining, which allows you to identify certain patterns and new knowledge in the data environment to make effective management decisions. This term was introduced by Gregory Piatetsky-Shapiro in 1989. The scope of its application is quite diverse. Thus, its application in the field of fraud detection accounts for 21.8%, which makes it a very popular means of finding errors, illegal operations, manipulations with information (Levkovich-Maslyuk, 2007).

The application of intellectual analysis is expanding and deepened by domestic and foreign scientists and practitioners for the study of economic phenomena and processes. Thus, recent developments and theoretical developments of practitioners and scientists in the field of intellectual analysis, which are presented in scientific publications, were systematized in the following directions:

1) general issues of modeling, fundamental concepts and algorithms of data mining: J. Poveda Poveda & J. Turmo Borras (2007), B. Klemens (2008), J. Stanton (2013), M. J. Zaki and W. Meira (2014);

2) methods and models of data analysis, basic algorithms of Data Mining: A. Barsehian, M. Kupriianov, V. Stepanenko, I. Kholod (2004, 2007), N. Paklin, V. Oreshkov (2009);

3) Microsoft multidisciplinary data analysis and OLAP tools: N. Yelmanova, O. Fedorov (2002), Ye. Melomed, V. Stepanenko, V. Shcherbynin, I. Horbach, O. Berger (2007), D. Mclennen, Ch. Tang, B. Kryvat (2008);

4) programming languages for the implementation of Data Mining:

– R: J. Gareth, D. Witten, T. Hastie, R. Tibshirani (2014), A. Shypunov, Ye. Baldin, P. Volkova, A. Korobeinikov et al. (2014), S. Mastytskii, V. Shytikov (2014);

– Python: W. J. Chun (2012), J. Grus (2015), W. McKinney (2016);

5) Big Data and Machine Learning: Ian H. Witten & Eibe Frank (2005), Jared Dean (2014), A. S. Müller, S. Guido (2016);

6) data mining of social networks such as Facebook, Twitter, Linkedin, Google, Github etc.: Matthew A. Russell (2013).

Thus, data mining is a fairly widespread trend for scientific research and the use of practices in the process of studying economic phenomena and processes.

Results of the research

Today there are quite a lot of methods of credit card fraud. The most common are:

1) physical methods: the use of fake cards, special devices that are installed on ATMs, fake terminals and ATMs that are not owned by the bank, and others;

2) software methods: the use of fake web pages, fake online stores, breaking up customer accounts, and others.

In fact, in the process of committing a fraudulent operation, fraudsters can use different methods, but for the bank the results of such actions are reflected in its database. If you analyze operations that were subject to fraudulent actions, then there are the following generalizations that make it possible to identify the operation in most cases as fraudulent;

1) withdrawal of the entire amount from the account, i.e. its reset, or partial case - withdrawal of a large amount from the account exceeding the possible limit;

2) implementation of a large number of operations on one account in a short period of time (time, day), which ultimately leads to its withdrawing;

3) transfer of a large amount of funds not inherent in prior operations to a third party account opened in another bank;

4) operation from the territory of another country.

Any fraud situation, regardless of the method, will in any case have one of these attributes. If the automated banking system will be able to automatically monitor the transactions of its clients in the process of their implementation and identify operations that potentially meet these features, then the system will signal the need for customer control and re-identification. If the operation is really fraudulent, then the person who carries it out will not have permission to conduct it. The system must then inform the reasons for the identification.

How will the system be able to detect fraudulent operations on these grounds? This is possible with the use of data mining tools to detect such situations. Of course, the use of such tools requires the creation of an automated module. This will allow continuous monitoring of operations and preventative and operative detection of violations. The process of functioning of such a system in this case will have the form shown in Figure 1. The process is constructed in the BPMN notation using Bizagi Modeller.

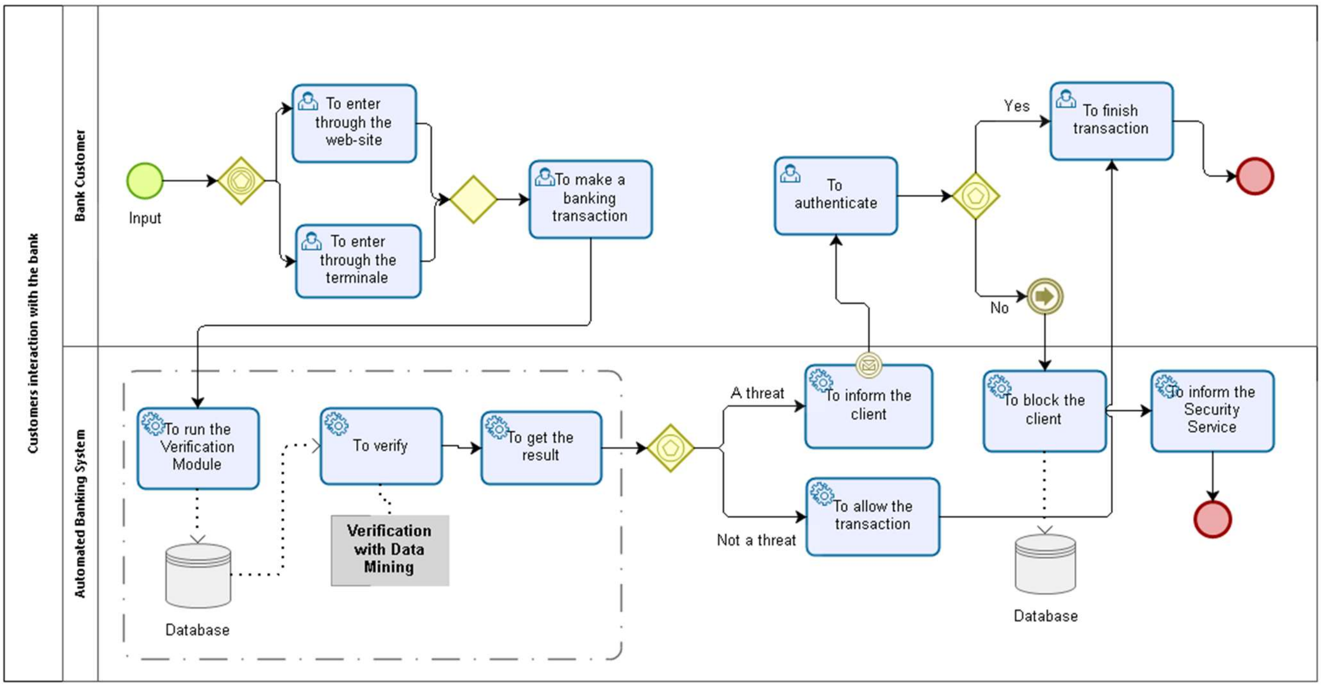

Customers interaction with the bank
Bank Customer
Automated Banking System
Input
To enter through the web-site
To enter through the terminale
To make a banking transaction
To authenticate
Yes
No
To finish transaction
To run the Verification Module
To verify
To get the result
Verification with Data Mining
Database
A threat
Not a threat
To inform the client
To allow the transaction
To block the client
To inform the Security Service
Database

Figure 1. Scheme of the process of interaction between the client and the bank when performing the operation with a payment card

Figure 1 highlights a verification unit that banks do not currently use, but with an increase in the number of fraud cases, there is such a need for its development and application. The scheme provides for a case where a fraud acts as a "customer", then in this case he will not be able to pass additional authentication. Also, in case of contact of a bank customer with a fraudster and informing him about codes, passwords, account numbers, cards, such a scheme will work also effectively. This is due to the fact that the customer will be warned not only about entering his account, but also about the operation that has signs of fraud.

The main function of this block is to conduct the data mining of transactions not only available, but also those that are in the process of processing or initialization. Data mining has in its arsenal a large list of methods. The most widespread of them are presented in Table 1.

Table 1. Advantages and disadvantages of the most widespread methods of data mining.

Name of the method and its essence	Advantages	Disadvantages
Association allows to find certain patterns between related events	Allows to find interesting patterns between data; works with data of any nature; the results are presented in the form of table, tree, text	Difficulty of understanding rules; the cumbersome rules, which sometimes causes the inconvenience of the analysis of regularities
Clustering breaks a set of objects into homogeneous groups (clusters)	Using different signs to break; has no limitations to the kind of observation	Due to compression of the information there may be some distortion; the number and composition of clusters have restrictions; it works only with quantitative data
Linear regression reveals the dependence of the researched indicator on one or more factors	Simplicity of construction and interpretation of results; fixed algorithm of calculations	Complexity of determining the type of functional communication and modeling of nonlinear processes; it is used for linear processes
Logit and Probit-models allow to determine the possibility of occurrence of an event by fitting data	Fixes the defects of linear regression in relation to the value of probability; easy to build and implement	Estimates are effective only if the number of observations exceeds 500
Decision-making trees graphically systematize the decision-making process for forecasting the value of the target variable, taking into account that each subsequent solution depends on the previous	Scalability that accelerates the calculation; unambiguousness of the learning process; self-adaptation with minimal human intervention; high accuracy of forecast; possibility of using categorical variables	Complexity of determining the number of optimal solutions; it takes a lot of time to be built; it can have many variants of branching; there is a need to use another methods to select factors

Table 1. (cont.). Advantages and disadvantages of the most widespread methods of data mining

Name of the method and its essence	Advantages	Disadvantages
Neural networks represent a system of artificial neurons that are interconnected and interact with each other, which allows, based on the learning process, to determine the result [6, p. 241]	Possibility of solving weakly formalized or non-formalized nonlinear problems; standard algorithm; ease of construction using software; learning opportunities both by a person and automatically	Difficult to interpret and understand; presence of inaccurate data with an accidental component; limited use
Bayesian analysis determines the most accurate probability of occurrence of a certain event in view of the emergence of new information	Possibilities of assessment by a person who decides on the probability of trust in the model; flexibility to take into account new information; application to situations that have not been analyzed before	Does not take into account the current state of the investigated object; complexity of calculation; it is not possible to choose a priori distribution
Genetic algorithms are designed to solve multidimensional random-search optimization problems	Possibility to apply for data of different types; allow to find universal solutions; find a plurality of solutions and choose the best ones; self-evolving	Unknown time to search; low search speed; a large number of free parameters; convergence can not be prooved

In our opinion, the most flexible instrument for analyzing data is neural networks, which are widely used in various spheres of the economy. The neural network imitates the behavior of the human brain, which allows it to be used to solve non-typical tasks. When comparing neural networks with traditional computing systems, then:

1) their use allows to solve problems with unknown laws of the development of the situation and the dependencies between input and output data that is not possible for traditional systems;

2) such systems have the ability to work with a large set of data, and what is more they choose the appropriate input signals independently;

3) they have the ability to adapt to changes in the environment, that is, in non-stationary conditions, when information changes over time. This property is just appropriate to use in the case of the creation of a neural network for the analysis of banking operations, where changes occur constantly;

4) such systems in case of damage of any links or the neuron itself do not lose their productivity;

5) they have the opportunity of speed action through the use of massive parallelism of information processing.

Based on the advantages and properties of neural networks, we can make the conclusion about their potential use for detecting fraud with banking operations. This is due to their properties of self-learning and the dynamic consideration of new information.

In this study, an algorithm for constructing a neural network for a set of data, taken on the example of the Sumy branch bank "A", which full name is not indicated in accordance with commercial secrets, has been implemented. The set of observations was formed from 5000 transactions of clients with bank cards. It was divided into three types of sample – 70% educational, 15% test, 15% control. 9 input variables were selected as input parameters, the description and values of which are presented in Table 2.

Table 2. Description of the input variables (Yarovenko, 2015).

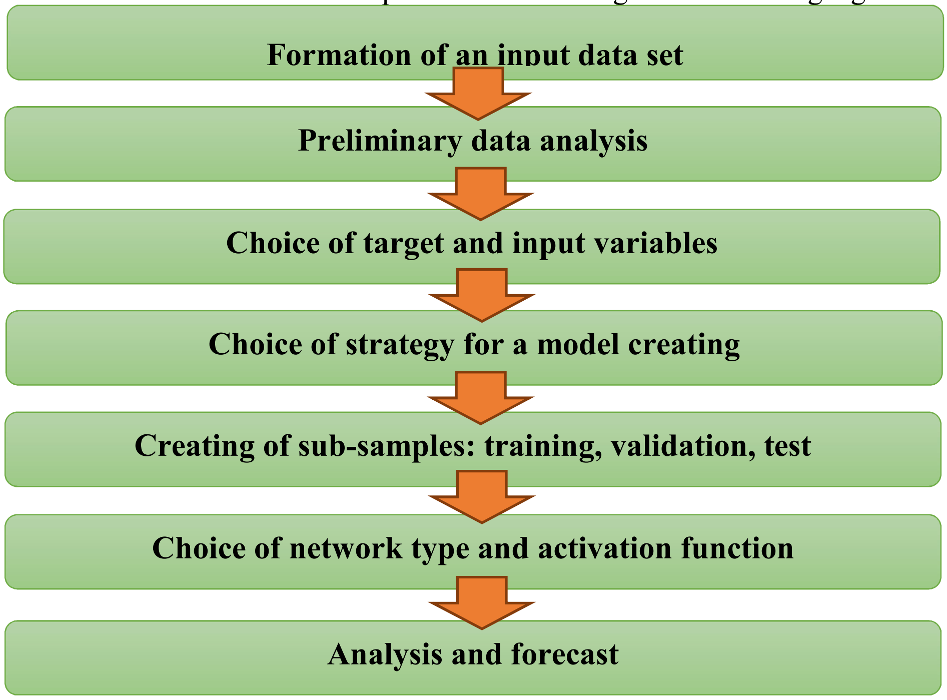
Variable name	Variable content	The role of the variable	Type
Y	Cases of fraud	target	binary
$X1$	Frequency of daily use of the card	input	interval
$X2$	The ratio of the logarithm of the transaction value to the logarithm of time between transactions	input	interval
$X3$	The ratio of the logarithm of the total transaction volume per day to the logarithm of the average time between transactions during the day	input	interval
$X4$	The standardized ratio of total transactions per day to the cumulative interval between transactions per day	input	interval
$X5$	Standardized volume of daily use of the card	input	interval
$X6$	Standardized transaction volume	input	interval
$X7$	Standardized time interval between card usage	input	interval
$X8$	Type of purchase according to volume	input	ordinal
$X9$	Type of purchase according to the probability of fraud	input	ordinal

The variable Y shows cases of fraud in this operation – "1", if there was a precedent of fraud; "0", if not. It serves as target, since it allows to identify the case of fraud.

Other variables were chosen according to the features mentioned above. Yes, the frequency of daily use of the card (x1) was chosen, which will allow us to select those clients who most often use the card during the day. As the frequency increases, the risk that the card is used by the fraud is increased. The variables x2 and x3 were calculated taking into account the volumes and time of transactions on a certain card. To bring the variables to normal distribution, their logarithm is performed. The variables x4-x7, which are formed taking into account the intervals of use of cards and volumes of daily use, have been standardized to eliminate the influence of various factors. In order to take into account the sign of the unusual use of card funds in case of its loss or theft, the variables x8 and x9 were introduced that characterize groups of potential purchases for the probability of fraud. These variables were evaluated by the bank independently according to its methodology.

For implementation of the model, the software "Statistica", developed by StatSoft, was chosen and designed for working with the data: implementation of various types of analysis, forecasting, data visualization. To study, we used the "Data Mining" module.

The construction of the neural network was performed according to the following algorithm (Fig. 2).



Formation of an input data set

Preliminary data analysis

Choice of target and input variables

Choice of strategy for a model creating

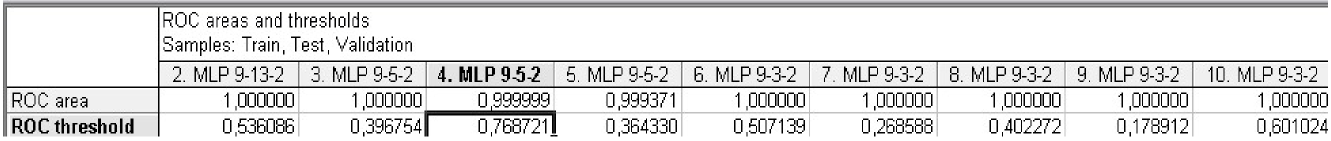
Creating of sub-samples: training, validation, test

Choice of network type and activation function

Analysis and forecast

Fig. 2. Algorithm for constructing the neural network

We have built several variants of the neural network using the automatic toolkit and with custom settings in the Statistica package. The most suitable for the parameters was a network, the value of the ROC-curve for which is 0.7687 (Fig.3).



ROC areas and thresholds
Samples: Train, Test, Validation

	2. MLP 9-13-2	3. MLP 9-5-2	**4. MLP 9-5-2**	5. MLP 9-5-2	6. MLP 9-3-2	7. MLP 9-3-2	8. MLP 9-3-2	9. MLP 9-3-2	10. MLP 9-3-2
ROC area	1,000000	1,000000	0,999999	0,999371	1,000000	1,000000	1,000000	1,000000	1,000000
ROC threshold	0,536086	0,396754	0,768721	0,364330	0,507139	0,268588	0,402272	0,178912	0,601024

Fig. 3. Selection of the neural network by the value of the ROC-curve

As a result, the neural network is constructed, the scheme of which is shown in Fig. 4.

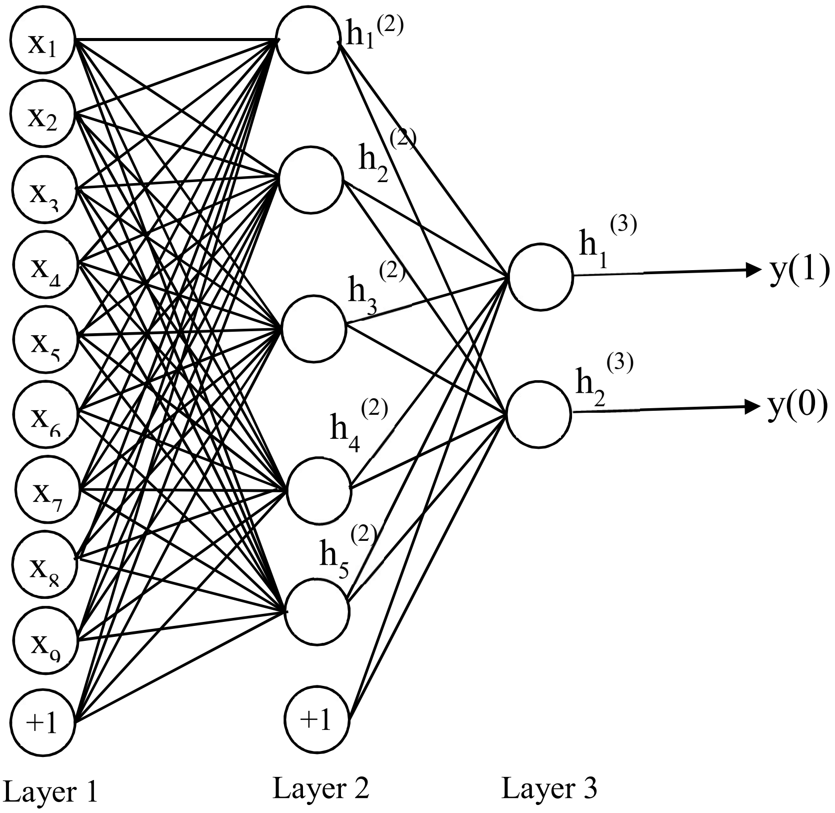


Fig 4. Scheme of the received neural network

The mathematical model of the neural network, taking into account the input and output variables presented in Table 2, can be represented in general form as follows (formulas 1-7):

$$h_1^{(2)} = f\left(w_{11}^{(1)}x_1 + w_{12}^{(1)}x_2 + \cdots + w_{19}^{(1)}x_9 + b_1^{(1)}\right), \quad (1)$$

$$h_2^{(2)} = f\left(w_{21}^{(1)}x_1 + w_{22}^{(1)}x_2 + \cdots + w_{29}^{(1)}x_9 + b_2^{(1)}\right), \quad (2)$$

$$h_3^{(2)} = f\left(w_{31}^{(1)}x_1 + w_{32}^{(1)}x_2 + \cdots + w_{39}^{(1)}x_9 + b_3^{(1)}\right), \quad (3)$$

$$h_4^{(2)} = f\left(w_{41}^{(1)}x_1 + w_{42}^{(1)}x_2 + \cdots + w_{49}^{(1)}x_9 + b_4^{(1)}\right), \quad (4)$$

$$h_5^{(2)} = f\left(w_{51}^{(1)}x_1 + w_{52}^{(1)}x_2 + \cdots + w_{59}^{(1)}x_9 + b_5^{(1)}\right), \quad (5)$$

$$y(1) = h_1^{(3)} = f\left(w_{11}^{(2)}h_1^{(2)} + w_{12}^{(2)}h_2^{(2)} + w_{13}^{(2)}h_3^{(2)} + w_{14}^{(2)}h_4^{(2)} + w_{15}^{(2)}h_5^{(2)} + b_1^{(2)}\right), \quad (6)$$

$$y(0) = h_2^{(3)} = f\left(w_{21}^{(2)}h_1^{(2)} + w_{22}^{(2)}h_2^{(2)} + w_{23}^{(2)}h_3^{(2)} + w_{24}^{(2)}h_4^{(2)} + w_{25}^{(2)}h_5^{(2)} + b_1^{(2)}\right), \quad (7)$$

where $f(\cdot)$ – unit activation function, in our case logistic function;

$h_1^{(2)}$ – output of the first unit in the second layer, the inputs of which is the output of the first unit in the second layer, that is $w_{11}^{(1)}x_1^{(1)}$, $w_{12}^{(1)}x_2^{(1)}$, …, $w_{19}^{(1)}x_9^{(1)}$ та $b_1^{(1)}$. These inputs are composed and transmitted to the activation function to calculate the output of the first unit. Other units $h_2^{(2)}$, $h_3^{(2)}$, $h_4^{(2)}$ and $h_5^{(2)}$ – similarly;

$h_1^{(3)}$ and $h_2^{(3)}$ – outputs of the second unit in the third layer, which takes the weighted outputs of the units of the second layer ($h_1^{(2)}$, $h_2^{(2)}$, $h_3^{(2)}$, $h_4^{(2)}$ and $h_5^{(2)}$).

As a function of activation for hidden layers and outputs, a sigmoid (logistic) function is used.

Logistic function for activating output units has the form (formula 8):

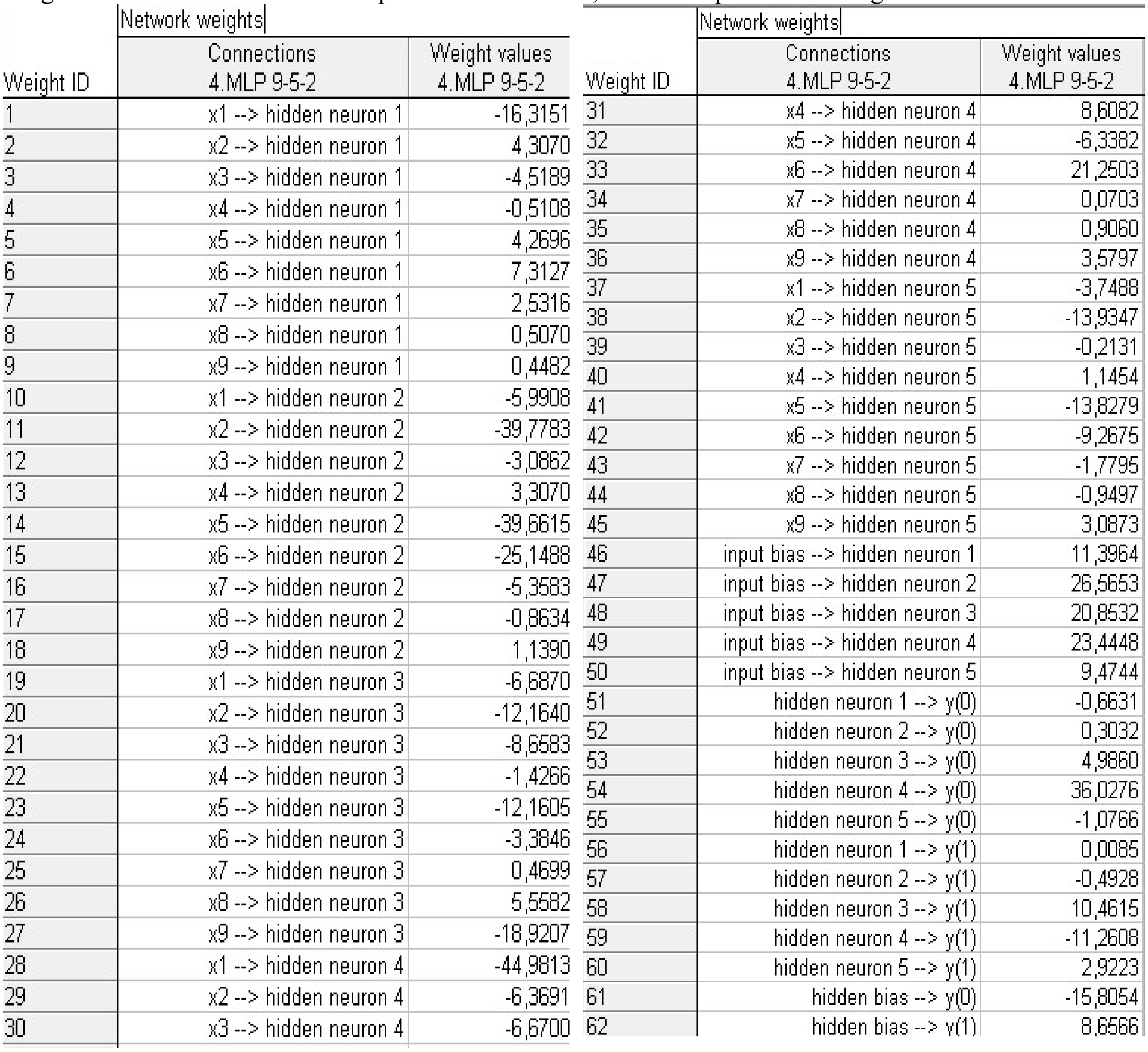
$$OUT = \frac{1}{1+\exp(-a\times net)}, \tag{8}$$

where OUT – outputs of the units of the neural network in the second and third layers, i.e. h1(2), h2(2), h3(2), h4(2), h5(2), h1(3) and h2(3);

net – amount of input signals multiplied by the corresponding weights for the second and third layers, for example, $\left(w_{11}^{(1)}x_1 + w_{12}^{(1)}x_2 + \cdots + w_{19}^{(1)}x_9 + b_1^{(1)}\right)$ for h1(2) (see formulas 1-7);

a – degree of slope of the logistic function.

As a result of using the algorithm for constructing the neural network (Figure 2), we obtained estimates of the weight coefficients of the most adequate neural network, which are presented in Figure 5.



Network weights

Weight ID	Connections 4.MLP 9-5-2	Weight values 4.MLP 9-5-2
1	x1 --> hidden neuron 1	-16,3151
2	x2 --> hidden neuron 1	4,3070
3	x3 --> hidden neuron 1	-4,5189
4	x4 --> hidden neuron 1	-0,5108
5	x5 --> hidden neuron 1	4,2696
6	x6 --> hidden neuron 1	7,3127
7	x7 --> hidden neuron 1	2,5316
8	x8 --> hidden neuron 1	0,5070
9	x9 --> hidden neuron 1	0,4482
10	x1 --> hidden neuron 2	-5,9908
11	x2 --> hidden neuron 2	-39,7783
12	x3 --> hidden neuron 2	-3,0862
13	x4 --> hidden neuron 2	3,3070
14	x5 --> hidden neuron 2	-39,6615
15	x6 --> hidden neuron 2	-25,1488
16	x7 --> hidden neuron 2	-5,3583
17	x8 --> hidden neuron 2	-0,8634
18	x9 --> hidden neuron 2	1,1390
19	x1 --> hidden neuron 3	-6,6870
20	x2 --> hidden neuron 3	-12,1640
21	x3 --> hidden neuron 3	-8,6583
22	x4 --> hidden neuron 3	-1,4266
23	x5 --> hidden neuron 3	-12,1605
24	x6 --> hidden neuron 3	-3,3846
25	x7 --> hidden neuron 3	0,4699
26	x8 --> hidden neuron 3	5,5582
27	x9 --> hidden neuron 3	-18,9207
28	x1 --> hidden neuron 4	-44,9813
29	x2 --> hidden neuron 4	-6,3691
30	x3 --> hidden neuron 4	-6,6700
31	x4 --> hidden neuron 4	8,6082
32	x5 --> hidden neuron 4	-6,3382
33	x6 --> hidden neuron 4	21,2503
34	x7 --> hidden neuron 4	0,0703
35	x8 --> hidden neuron 4	0,9060
36	x9 --> hidden neuron 4	3,5797
37	x1 --> hidden neuron 5	-3,7488
38	x2 --> hidden neuron 5	-13,9347
39	x3 --> hidden neuron 5	-0,2131
40	x4 --> hidden neuron 5	1,1454
41	x5 --> hidden neuron 5	-13,8279
42	x6 --> hidden neuron 5	-9,2675
43	x7 --> hidden neuron 5	-1,7795
44	x8 --> hidden neuron 5	-0,9497
45	x9 --> hidden neuron 5	3,0873
46	input bias --> hidden neuron 1	11,3964
47	input bias --> hidden neuron 2	26,5653
48	input bias --> hidden neuron 3	20,8532
49	input bias --> hidden neuron 4	23,4448
50	input bias --> hidden neuron 5	9,4744
51	hidden neuron 1 --> y(0)	-0,6631
52	hidden neuron 2 --> y(0)	0,3032
53	hidden neuron 3 --> y(0)	4,9860
54	hidden neuron 4 --> y(0)	36,0276
55	hidden neuron 5 --> y(0)	-1,0766
56	hidden neuron 1 --> y(1)	0,0085
57	hidden neuron 2 --> y(1)	-0,4928
58	hidden neuron 3 --> y(1)	10,4615
59	hidden neuron 4 --> y(1)	-11,2608
60	hidden neuron 5 --> y(1)	2,9223
61	hidden bias --> y(0)	-15,8054
62	hidden bias --> y(1)	8,6566

Fig. 5. Weight coefficients of the neural network obtained in the Statistica package

We use the obtained values for constructing the mathematical model of the neural network (formulas 9-15).

$$h_1^{(2)} = f(-16{,}3151x_1 + 4{,}3070x_2 - 4{,}5189x_3 - 0{,}5108x_4 + 4{,}2696x_5 + 7{,}3127x_6 + 2{,}5316x_7 + 0{,}5070x_8 + 0{,}4482x_9 + 11{,}3964), \tag{9}$$

$$h_2^{(2)} = f(-5{,}9908x_1 - 39{,}7783x_2 - 3{,}0862x_3 + 3{,}3070x_4 - 39{,}6615x_5 - 25{,}1488x_6 - 5{,}3583x_7 - 0{,}8634x_8 + 1{,}1390x_9 + 26{,}5653), \tag{10}$$

$$h_3^{(2)} = f(-6{,}6870x_1 - 12{,}1640x_2 - 8{,}6583x_3 - 1{,}4266x_4 - 12{,}1605x_5 - 3{,}3846x_6 + 0{,}4699x_7 + 5{,}5582x_8 - 18{,}9207x_9 + 20{,}8532), \tag{11}$$

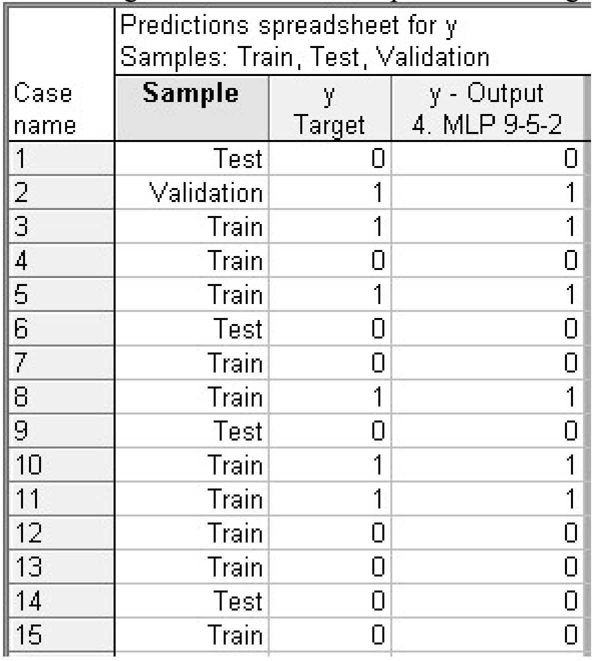
$$h_4^{(2)} = f(-44{,}9813x_1 - 6{,}3691x_2 - 6{,}6700x_3 + 8{,}6082x_4 - 6{,}3382x_5 + 21{,}2503x_6 + 0{,}0703x_7 + 0{,}9060x_8 + 3{,}5797x_9 + 23{,}4448), \tag{12}$$

$$h_5^{(2)} = f(-3{,}7488x_1 - 13{,}9347x_2 - 0{,}2131x_3 + 1{,}1454x_4 - 13{,}8279x_5 - 9{,}2675x_6 - 1{,}7795x_7 - 0{,}9497x_8 + 3{,}0873x_9 + 9{,}4744), \tag{13}$$

$$y(1) = h_1^{(3)} = f\left(0{,}0085h_1^{(2)} - 0{,}4928h_2^{(2)} + 10{,}4615h_3^{(2)} - 11{,}2608h_4^{(2)} + 2{,}9223h_5^{(2)} + 8{,}6566\right), \tag{14}$$

$$y(0) = h_2^{(3)} = f\left(-0{,}6631h_1^{(2)} + 0{,}3032h_2^{(2)} + 4{,}9860h_3^{(2)} + 36{,}0276h_4^{(2)} - 1{,}0766h_3^{(2)} - 15{,}8054\right). \tag{15}$$

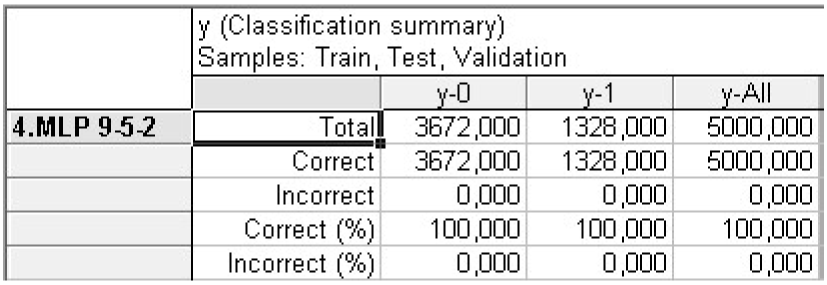
Using the received neural network, we will build a forecast of the probability of a fraudulent operation taking into account the given conditions. A fragment of the results is presented in Fig. 6.



Case name	Predictions spreadsheet for y Samples: Train, Test, Validation Sample	y Target	y - Output 4. MLP 9-5-2
1	Test	0	0
2	Validation	1	1
3	Train	1	1
4	Train	0	0
5	Train	1	1
6	Test	0	0
7	Train	0	0
8	Train	1	1
9	Test	0	0
10	Train	1	1
11	Train	1	1
12	Train	0	0
13	Train	0	0
14	Test	0	0
15	Train	0	0

Figure 6. A fragment of predictive values obtained with the help of the neural network

The error matrix of the built network in the three sub-categories shows good results (Fig. 7).



y (Classification summary)
Samples: Train, Test, Validation

		y-0	y-1	y-All
4.MLP 9-5-2	Total	3672,000	1328,000	5000,000
	Correct	3672,000	1328,000	5000,000
	Incorrect	0,000	0,000	0,000
	Correct (%)	100,000	100,000	100,000
	Incorrect (%)	0,000	0,000	0,000

Fig. 7. A fragment of predictive values obtained using the neural network

The constructed neural network can be used for further analysis of card operations to identify potential frauds with them. The model gives good results, but needs further development, taking into account new observations for its further training and improvement.

Conclusions

According to the results of research, the necessity of using data mining in the banking sphere was grounded in order to detect and prevent fraudulent operations with customer cards. For maximum effect it is expedient to create a module for monitoring operations in an automated banking system in which it is necessary to implement the algorithm of the neural network. Its application will allow automatic checking of operations at the stage of their initiation by a client or a potential intruder. In the process of checking, operations will be selected in accordance with the signs of fraud. The algorithm of the neural network will allow them to be evaluated.

In future, it is planned to develop the architecture of the proposed module and improve the neural network through its training on new data.

References

1. Bazy`lenko A. (2017). U 2016-mu shaxrayi vkraly` z raxunkiv ukrayinciv majzhe 340 mln grn. Retrieved from: http://watcher.com.ua/2017/02/07/u-2016-mu-shahrayi-vkraly-z-rahunkiv-ukrayintsiv-mayzhe-340-mln-hrn/.
2. "Kartkovi" shaxrayi zavdaly` bankam 180 mil`joniv gry`ven` zby`tkiv. (2016). Ukrayins`ka pravda. Retrieved from: https://www.epravda.com.ua/news/2016/02/26/583169/
3. Roman K. (2017). Vterly`sya v doviru: yak shaxrayi znimayut` groshi z bankivs`ky`x kartok ukrayinciv. Ukrayina. Retrived from: https://daily.rbc.ua/ukr/show/moshenniki-snimayut-dengi-bankovskih-kart-1500294135.html
4. Ryabuxa O. (2017). Nova sxema shaxrajstva. Pid udarom kliyenty` Pry`vatBanku i ne til`ky`. Minfin. Retrieved from: https://minfin.com.ua/ua/2017/11/16/30965613/
5. Levkovich-Maslyuk L. (2007). Velikie raskopki i velikie vyizovyi. *Kompyuternyiy poisk znaniy stanovitsya vse bolee tsennyim*, *11*(679), 48-51.
6. Yarovenko G.M. (2015). Modelyuvannya ymovirnosti viniknennya shahrayskih operatsIy z kreditnimi kartkami. ZbIrnik naukovih prats "*Problemi i perspektivi rozvitku bankIvskoii sistemi*", 41, 237-248.
7. Ivanov S.V. (2014). Preimuschestva geneticheskih algoritmov i ih primenenie v meditsine. *Aktualnyie problemyi gumanitarnyih i estestvennyih nauk*, 10, 44-47.
8. Preimuschestva neyronnyih setey (2016) Portal iskusstvennogo intellekta. Retrieved from: http://www.aiportal.ru/articles/neural-networks/advantages.html.
9. STATISTICA Automated Neural Networks (SANN) - Neural Networks: An Overview. Retrieved from: Portal StatSoft. http://documentation.statsoft.com/STATISTICAHelp.aspx?path=SANN/ Overview/ SANNNeural NetworksAnOverview
10. Berthold M., Hand D.J. (2003). Intelligent Data Analysis: An Introduction / M. Berthold, D.J. Hand. – Springer-Verlag Berlin Heidelberg, 515.