MODELING REGARDING DETECTION OF CYBER THREATS FEATURES IN BANKS ACTIVITIES

Olha Shulha, State University of Intelligent Technologies and Telecommunications
Iryna Yanenkova, State organization "Institute for Economy and Forecasting of National academy of Ukraine"
Mikhaylo Kuzub, Kyiv National University of Trade and Economics
Iskandar Muda, Universitas Sumatera Utara, Medan, Indonesia
Viktor Nazarenko, Dragomanov National Pedagogical University

ABSTRACT

The following groups of banks` cyber threats have been identified: network and application layer attacks; social engineering; developed stable threats; cybercrime, master data violation. Banks are suffering from phishing attacks and bank roses, which is the most common mobile cyber threat, since most of smart phones holders also have a bank card. A neural network was created, consisting of a 1st hidden layer with two neurons, moreover, a mathematical interpretation of the source layer wa represented, as well as the 1st and 2nd hidden layers of the neuron. The constructed models were tested for quality and adequacy using the following coefficients: MISC misclassification rates, MSE and ASE error rates. Their results were analyzed and it was found that all the models created were describing the input data with the same accuracy, but nevertheless, the neural network model was the best. The probability prediction of cyber-threat signs occurrence during the transaction of mobile and Internet banking users was performed on the basis of the selected neural network model. As a result, it was found that 21.9% of transactions were likely to contain cyber-threats. The application of this model will enable the banking sector employees to detect cyber threats in transactions, thereby preventing mobile and internet banking users from possible losses caused by certain criminal activity.

Keywords: Cyber threat; Internet banking; Neural network model; Regression model; Decision tree.

JEL Classifications: M5, Q2

INTRODUCTION

The routine activities of banking systems are directly related to the usage of modern computer technologies and are fully dependent on the reliable and stable operation of electronic computers. Cyber threats in today's society are becoming more widespread. Nowadays, a successful attack by hackers can cut off electricity of an area or even a country, rob a bank or destroy a successful organization.

The banking sector is at the forefront of the list of targeted industries which is most prone to cybercrime attacks. That is why banks should always be vigilant and prompt in responding to information about potential threats, as well as the types of attacks and methods of their implementation.

Citation Information: Shulha, O., Yanenkova, I., Kuzub, M., Muda, I., & Nazarenko, V. (2022). Modeling regarding detection of cyber threats features in banks activities. *Journal of Management Information and Decision Sciences, 25*(S1), 1-8.

The topicality of this problem is determined by the high level of banking sector vulnerability to cyber threats, the significant costs of cybersecurity in the banking sector, as well as the rapid pace of new cyber threats development.

The main goal of the study is to construct a mathematical model for identifying signs of cyber threats in the bank sphere and its practical implementation using the methods of intellectual analysis with the help of analytical package SAS Enterprise Miner.

REVIEW OF PREVIOUS STUDIES

Banking activities are closely linked to the usage of modern computer technologies and are fully dependent on the reliable and resistant operation of electronic systems. Worldwide experience testifies to the vulnerability of any company due to the fact that cybercrime has no national borders, and therefore hackers have the ability to threaten information systems anywhere in the world (Qamar et al., 2017; Martins et al., 2020).

The most common cyber threats in the banking system:

network and application layer attacks (Camillo, 2017; Brasil & Campelo, 2018): interruption or suspension of servers and network resources activity connected to the Internet; easy attack of any startup, it is very difficult for banks to decide on their own; DDoS attack packages are easily accessible to anyone on the black market; DDoS attacks can be triggered by cybercriminals in order to distract banking staff from notable fraudulent transactions, such as unsanctioned money transactions; social engineering (Antonacci, 2018): phishing attacks often occur to banking customers; banking clients receive fake emails aimed at accessing their accounts or obtaining certain personal information; fake emails are carefully created in order to copy genuine emails that are usually sent by banks.

developed persistent threats (Kiwia et al., 2018): usage of system vulnerabilities in order to install «Backdoor»; with proper code, the attackers remain unnoticed and continue causing damage as long as possible; organized cybercrime (Tosh et al., 2017): risk of illegal possession of intellectual property, confiscation of bank accounts and loss of consumers as a result of business disruptions; easier to prevent than eliminate, cybercriminals specialize in the sale of personal information on the black market, using ransom and intimidation; violation of master data (Liu et al., 2019; Khraisat et al., 2019; Ferrag et al., 2020): highly professional hackers who use reliable infrastructure for targeted banking institutions, steal customer data and then sell it; using various methods, confidential information about banking institutions and their clients can be disclosed; business is disrupted, customer and business data are deteriorating, and recovery processes are costly.

METHODOLOGY

The following general scientific methods were used during the study of the chosen topic: analysis and synthesis, deduction, abstraction, concretization, argumentation, comparison, classification and method of generalization, with the help of which general conclusions were drawn.

Cybernetic methods are the most promising area of intellectual analysis, representing a multitude of approaches united by the idea of computer mathematics and the usage of artificial intelligence theory. The main methods of this group are the following: neural networks, decision trees, cluster and associative analyzes, construction of nonlinear regressions, etc. The usage of intellectual analysis helps to build a conceptual model whose main purpose is to identify cyber threats in banking institutions in order to prevent their occurrence in the future.

Citation Information: Shulha, O., Yanenkova, I., Kuzub, M., Muda, I., & Nazarenko, V. (2022). Modeling regarding detection of cyber threats features in banks activities. *Journal of Management Information and Decision Sciences, 25*(S1), 1-8.

The SAS regulatory and supporting documentation, the bank's empirical data on mobile and Internet banking user transactions, as well as scientific publications by experts in the field of economic and mathematical modeling made up the factual base of the study.

RESULTS AND DISCUSSIONS

We use the SAS Enterprise Miner analytics package in order to construct a cyber-threat detection model in banking institutions to prevent the occurrence of these threats in the future.

SAS Enterprise Miner streamlines and facilitates the data analysis process, while ensuring the creation of accurate predictive and descriptive models based on the analysis of the vast amount of information collected across the entire organization.

This toolkit helps to solve the following problems: detect fraud existence, identify and minimize risks, predict resource requirements, prevent incidents, increase response to marketing campaigns, reduce customer outflow, etc.

Next, we need to build neural network models using AutoNeural. This tool allows to build, refine, and test multilayered neural networks with direct communication automatically. That is, it conducts a limited search for a better network configuration. In general, each input element is fully associated with the first hidden layer, each hidden layer is fully associated with the next hidden layer, and the last hidden layer is fully associated with the output data.

Network optimization was done by minimizing the misclassification factor, the Figure 1 shows a graph of its changes depending on the number of relevant neurons.

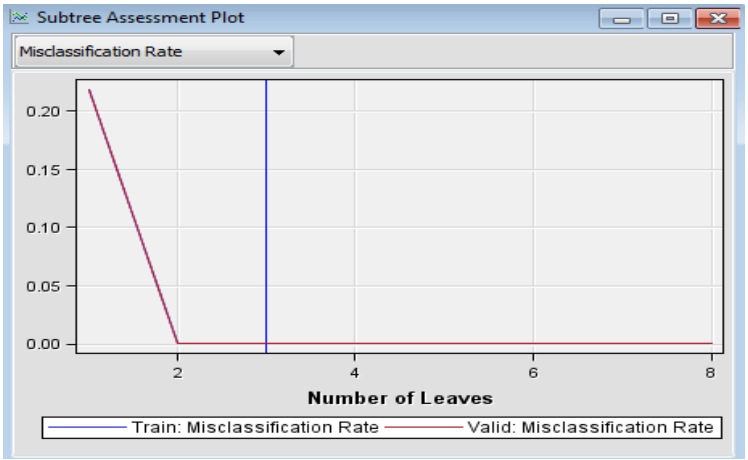

FIGURE 1

CONDUCTING NEURAL OPTIMIZATION USING THE SAS ENTERPRISE MINER MISCLASSIFICATION FACTOR

As can be seen from the Figure 1, the coefficient graph for the training and validation sets decreases rapidly at the interval (0;2). Then no change occurs for the training data or for the validation, so building a network with more layers is inexpedient.

As a result, a neural network model consisting of a 1st hidden layer with two neurons was generated.

The network architecture in SAS Enterprise Miner is depicted in Figure 2.

Citation Information: Shulha, O., Yanenkova, I., Kuzub, M., Muda, I., & Nazarenko, V. (2022). Modeling regarding detection of cyber threats features in banks activities. *Journal of Management Information and Decision Sciences, 25*(S1), 1-8.

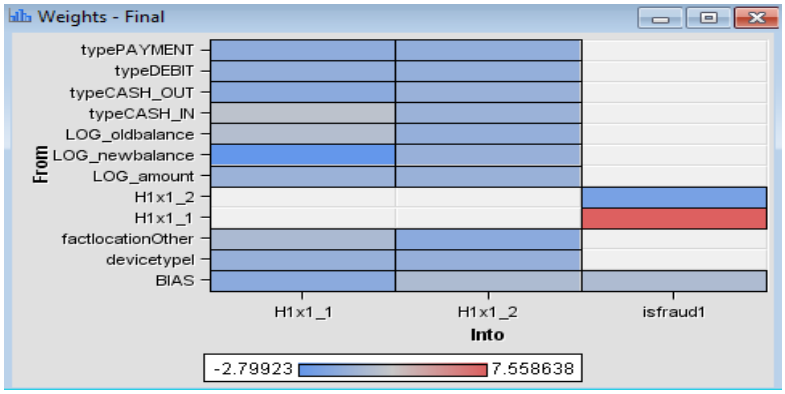

FIGURE 2

NEURAL NETWORK ARCHITECTURE IN SAS ENTERPRISE MINER

Neural networks have the ability to adapt to the external environment changes, that is, in unsteady conditions, when information changes over time. It is advisable to use this feature when creating a neural network in order to analyze constantly changing bank transactions.

Next, an analysis of the quality and adequacy of the models was carried out. Table 1 presents the classification characteristics using a regression model for training and validation samples.

TABLE 1
CHARACTERISTICS OF THE REGRESSION MODEL CLASSIFICATION FEATURES BASED ON TRAINING AND VALIDATION SAMPLES

The target variable	Result	The target variable, %	Result, %	Frequency of such cases	General classification, %
Training sample					
0	0	99,9923	99,9949	78093	78,0922
1	0	0,0077	0,0274	6	0,0060
0	1	0,0183	0,0051	4	0,0040
1	1	99,9817	99,9726	21898	21,8978
Validation sample					
0	0	99,9987	99,9974	78094	78,0948
1	0	0,0013	0,0046	1	0,0010
0	1	0,0091	0,0026	2	0,0020
1	1	99,9909	99,9954	21902	21,9022

Therefore, the Table 1 shows that the model in the training sample correctly classified 99.99% of transactions that are not cyber threats and 99.97% of transactions that are cyber threats. At the same time, the model classified 0.03% of cyber-threatened transactions as non-cyber-threatened transactions and 0,01% of non-cyber-threatened transactions were classified as cyber-threatening. Regarding absolute values, the model correctly classified negative result (not cyber threat) in 78093 transactions, and positive - 21898. Incorrectly classified the negative result in 4 transactions, and the positive - in 6. As a result, we can say that the proportion of correct classification was 99.99% (78.0922% + 21.8978%).

Regarding the validation data, 99.99% of transactions that were not cyber-threatened and 99.99% of transactions that were cyber-threatened were classified correctly. At the same time, the model classified 0.005% of cyber-threat transactions as non-cyber-threatened

Citation Information: Shulha, O., Yanenkova, I., Kuzub, M., Muda, I., & Nazarenko, V. (2022). Modeling regarding detection of cyber threats features in banks activities. *Journal of Management Information and Decision Sciences, 25*(S1), 1-8.

transactions and 0.002% of non-cyber-threat transactions were classified as cyber-threats. With regard to absolute values, the model classified correctly negative result (not cyber-threat) in 78094 transactions, and positive rusult - in 21902. The negative result was qualified incorrectly in 2 transactions and a positive result - in 1 transaction. As a result, it should be mentioned that the proportion of correct classification was 99.997% (78.0948% + 21.9022%).

Table 2 represents the main coefficients characterizing the regression model quality.

TABLE 2
QUALITY COEFFICIENTS OF THE REGRESSION MODEL

Coefficient	Sample	
	Training	Validation
Misclassification Rate, MISC	0,0001	0,00003
Mean Square Error, MSE	0,001119	0,001091
Average Squared Error, ASE	0,001119	0,001091

Thus, the low values of the calculated coefficients testify to the quality and adequacy of the constructed model.

Based on the decision tree on the training and validation samples, Table 3 summarizes the classification characteristics.

TABLE 3
CHARACTERISTICS OF CLASSIFICATION FEATURES OF DECISION TREE BASED ON TRAINING AND VALIDATION SAMPLES

The target variable	Result	The target variable, %	Result, %	Frequency of such cases	General classification, %
Training sample					
0	0	99,9974	99,9949	78093	78,0922
1	0	0,0026	0,0091	2	0,0020
0	1	0,0183	0,0051	4	0,0040
1	1	99,9817	99,9909	21902	21,9018
Validation sample					
0	0	99,9987	99,9974	78094	78,0948
1	0	0,0013	0,0046	1	0,0010
0	1	0,0091	0,0026	2	0,0020
1	1	99,9909	99,9954	21902	21,9022

Thus, 99.99% of transactions that are not cyber-threatened and 99.99% of transactions that are cyber-threatened were classified correctly using the training sample of the model. Also, 0.009% of cyber-threat transactions were classified as non-cyber-threatened and 0.005% of non-cyber-threat transactions were classified as cyber-threat. Regarding the absolute values, model classified correctly the negative result in 78093 transactions, and the positive – in 21902. The negative result was classified incorrectly in 4 transactions and a positive result - in 2 transactions. In general, it can be said that the proportion of correct classification is 99.994%.

With regard to validation data, the model classified correctly 99.99% of non-cyber-threat transactions and 99.99% of cyber-threat transactions. At the same time, the model classified 0.005% of transactions that were cyber-threatened as non-cyber-threatened and 0.002% of transactions were classified mistakenly as cyber-threatened. In terms of absolute values, the model classified correctly negative result (not cyber threat) in 78094 transactions, and positive - 21902. The wrong result is classified as negative in 2 transactions and positive

Citation Information: Shulha, O., Yanenkova, I., Kuzub, M., Muda, I., & Nazarenko, V. (2022). Modeling regarding detection of cyber threats features in banks activities. *Journal of Management Information and Decision Sciences, 25*(S1), 1-8.

in 1. In general, we can say that the proportion of correct classification was 99.997% (78.0948% + 21.9022%).

Table 4 lists the coefficients that describe the decision tree quality.

TABLE 4
QUALITY COEFFICIENTS OF THE DECISION TREE

Coefficient	Sample	
	Training	Validation
Misclassification Rate, MISC	0,00009	0,00003
Average Squared Error, ASE	0,001112	0,001097

Thus, we can make a conclusion that the constructed decision tree is of high quality and adequate, which is confirmed by the analysis of modeling results.

The following Table 5 demonstrates the characteristics of neural network-based classification in training and validation samples.

TABLE 5
CHARACTERISTICS OF THE CLASSIFICATION FEATURES OF THE NEURAL NETWORK MODEL IN THE TRAINING SAMPLE

The target variable	Result	The target variable, %	Result, %	Frequency of such cases	General classification, %
Training sample					
0	0	99,9949	99,9987	78096	78,0952
1	0	0,0051	0,0183	4	0,0040
0	1	0,0046	0,0013	1	0,0010
1	1	99,9954	99,9817	21900	21,8998
Validation sample					
0	0	99,9987	99,9987	78095	78,0958
1	0	0,0013	0,0046	1	0,0010
0	1	0,0046	0,0013	1	0,0010
1	1	99,9954	99,9954	21902	21,9022

Therefore, Table 5 shows that the training sample model classified correctly 99.99% of non-cyber-threat transactions and 99.98% of cyber-threat transactions. Meanwhile, the model has classified 0.018% of transactions that were cyber-threatened as non-cyber-threatened and 0.001% of non-cyber-threatened transactions were classified as cyber-threatening. In terms of absolute values, the model has classified correctly a negative result (not a cyber threat) in 78096 transactions and a positive - in 21900. A negative result was classified incorrectly in 1 transaction and a positive one - in 4 transactions. Generally, it can be said that the proportion of correct classification was 99,995% (78,095% + 21.8998%).

In terms of validation data, the model has classified correctly 99.99% of transactions that were not cyber-threatened and 99.99% of transactions that were cyber-threatened. Meanwhile, the model has classified 0.005% of cyber-threat transactions as non-cyber-threatened transactions and 0.001% of non-cyber-threat transactions were classified as cyber-threat. With regard to absolute values, the model has classified correctly a negative result (not a cyber threat) in 78095 transactions, and a positive – in 21902 transactions. A negative result as well as positive one were classified incorrectly in 1 transaction each of them. In general, it can be said that the proportion of correct classification was 99.998% (78.0958% + 21.9022%).

The main coefficients characterizing the quality of the neural network model are presented in Table 6.

Citation Information: Shulha, O., Yanenkova, I., Kuzub, M., Muda, I., & Nazarenko, V. (2022). Modeling regarding detection of cyber threats features in banks activities. *Journal of Management Information and Decision Sciences, 25*(S1), 1-8.

TABLE 6
NEURAL NETWORK QUALITY COEFFICIENTS

Coefficient	Sample	
	Training	Validation
Misclassification Rate, MISC	0,00005	0,00002
Mean Square Error, MSE	0,001105	0,001094
Average Squared Error, ASE	0,001105	0,001094

Thus, it can be stated that the constructed neural network model is qualitative and adequate, which is confirmed by the numerical characteristics of the modeling results. Comparative coefficients characterizing the quality and accuracy of models: regression, decision tree and neural network are represented in Table 7.

TABLE 7
COMPARATIVE MODELS` CHARACTERISTICS

No	Model	Misclassification Rate, MISC		Mean Square Error, MSE	
		Validation	Training	Validation	Training
1	AutoNeural	0,00002	0,00005	0,001094	0,001105
2	Decision Tree	0,00003	0,00009	0,001097	0,001112
3	Regression	0,00003	0,0001	0,001091	0,001119

The models in Table 7 are ranked from best to worst in terms of misclassification rate and mean square error meaning. The values of the misclassification rate and the mean squared error were the lowest in the AutoNeural model, in the second place - Decision Tree and in the third place - Regression model.

Thus, using the Model Comparison tool, it was determined that during a bank transaction of mobile and Internet banking, the neural network models better the assessment of the outcome of a cyber threat occurrence, than models of regression and decision tree.

In the validation dataset, 21.9032% of transactions were recognized as cyber threats and 78.0968% were not.

Almost equal percentage indicates stationarity between training, validation as well as scoring data.

The constructed cyber threat detection model can be implemented in banking institutions in order to prevent potential cyber threats.

The implementation of the constructed model guarantees that the bank will receive certain social and economic effects.

The social effect is manifested in the following aspects: increasing the level of customer confidence in banks through increased security and reliability level; the possibility of customers to choose from those banks that offer cyber-security services.

The economic effect involves: reducing the cost to banks for eliminating the effects of cyber threats by preventing them in advance; increase in profits due to increased confidence in banks and, as a result, attracting new customers.

CONCLUSIONS

From phishing attacks, as well as from the most common mobile cyber threat - the banking trojan, banks suffer the most, since most smart card holders also have a bank card. The main measures taken by banks in order to prevent this threat are the following:

Citation Information: Shulha, O., Yanenkova, I., Kuzub, M., Muda, I., & Nazarenko, V. (2022). Modeling regarding detection of cyber threats features in banks activities. *Journal of Management Information and Decision Sciences, 25*(S1), 1-8.

increasing the staff providing security services, attracting additional resources for carrying out special investigations. However, foreign banks are already applying more innovative and up-to-date approaches to this issue by introducing tools for business analytics.

To construct a model for detection of bank`s cyber threats as the source data was used information contained in the database of mobile and Internet banking of the bank "X". Also, eight input variables and a target, expressed in different cases of cyber-threat detection were selected.

In order to build a logistic regression, the stepwise exclusion method of minor factors was chosen, as a result, 3 significant factors were selected: the fixed location of the device from which the transaction was conducted; the client's balance before and after the transaction.

As a result of the decision tree creation, a three-tier classification tree was generated, which, in turn, has indicated the following major factors - the location of the device from which the transaction was performed and the type of device from which the transaction was performed. Also, a neural network consisting of a 1st hidden layer with two neurons was built.

The constructed models have been tested for adequacy and quality. Their results were analyzed and it was found that all created models have almost the same accuracy of describing the input data, but the results of all other indicators have shown that model based on the neural network is still the best.

The usage of intellectual analysis models will help the banking sector to detect cyber threats in transactions, thereby preventing mobile and Internet banking users from possible harm caused by such a criminal activity. However, these models require constant updating and improvement as new threats to mobile and Internet banking users emerge.

REFERENCES

Antonacci, P. (2018). The cyberthreat facing the financial services industry. *Cyber Security: A Peer-Reviewed Journal, 2*(2), 106-113.

Brasil, B. S., & Campelo, K. C. F. (2018). Banking phishing: The case of Brazil. *Cyber Security: A Peer-Reviewed Journal, 2*(1), 6-16.

Camillo, M. (2017). Cybersecurity: Risks and management of risks for global banks and financial institutions. *Journal of Risk Management in Financial Institutions, 10*(2), 196-200.

Ferrag, M. A., Maglaras, L., Moschoyiannis, S., & Janicke, H. (2020). Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. *Journal of Information Security and Applications*, *50*, 102419.

Khraisat, A., Gondal, I., & Vamplew, P. (2019). Survey of intrusion detection systems: techniques, datasets and challenges. *Cybersecurity, 2*, 20.

Kiwia, D., Dehghantanha, A., Choo, K. K. R., & Slaughter, J. (2018). A cyber kill chain based taxonomy of banking Trojans for evolutionary computational intelligence. *Journal of computational science, 27*, 394-409.

Liu, Q., Xing, L., & Zhou, C. (2019). Probabilistic modeling and analysis of sequential cyber-attacks. *Engineering Reports, 1*(4), e12065.

Martins, N., Cruz, J. M., Cruz, T., & Abreu, P. H. (2020). Adversarial machine learning applied to intrusion and malware scenarios: a systematic review. *IEEE Access*, *8*, 35403-35419.

Qamar, S., Anwar, Z., Rahman, M. A., Al-Shaer, E., & Chu, B. T. (2017). Data-driven analytics for cyber-threat intelligence and information sharing. *Computers & Security, 67*, 35-58.

Tosh, D. K., Shetty, S., Sengupta, S., Kesan, J. P., & Kamhoua, C. A. (2017, May). Risk management using cyber-threat information sharing and cyber-insurance. *Proceedings of International Conference on Game Theory for Networks* (pp. 154-164). Springer, Cham.

Citation Information: Shulha, O., Yanenkova, I., Kuzub, M., Muda, I., & Nazarenko, V. (2022). Modeling regarding detection of cyber threats features in banks activities. *Journal of Management Information and Decision Sciences, 25*(S1), 1-8.