I. J. Computer Network and Information Security, **2022, 4, 1-13**
Published Online August 2022 in MECS (http://www.mecs-press.org/)
DOI:10.5815/ijcnis.2022.04.01

Risk Forecasting of Data Confidentiality Breach Using Linear Regression Algorithm

Oleksandr Korystin
State Scientifically Research Institute of the MIA of Ukraine, Kyiv, Ukraine
E-mail: alex@korystin.pro

Svyrydiuk Nataliia
State Scientifically Research Institute of the MIA of Ukraine, Kyiv, Ukraine
E-mail: S_N_P_@ukr.net

Olena Mitina
Odesa Polytechnic National University, Odesa, Ukraine
E-mail: olenamitina@ukr.net

Received: 13 January 2022; Revised: 11 March 2022; Accepted: 28 May 2022; Published: 08 August 2022

Abstract: The paper focuses on the study of cyber security in Ukraine and creation of a predictive model for reducing the risk of identified cyber threats. Forecasting is performed using a linear regression model, taking into account the optimal dependence of specific threats in the field of cyber security of Ukraine on variables characterizing capabilities / vulnerabilities of cyber security. An unique empirical base was used for the analysis, which was formed on the basis of an expert survey of the cyber security system's subjects in Ukraine. In order to increase the representativeness of the research, based on the selection of reliable expert population, data cleaning is provided. Methodological research is based on a risk-oriented approach, which provided a risk assessment of the spread of cyber threats and, on this basis, the determination of capabilities / vulnerabilities of the cyber security system in Ukraine. The value of the research is formed not only by assessing the risks of the spread of cyber threats, but by a more in-depth analysis of the dependence of the cyber threats' level on the vulnerability of the cyber security system based on the search for optimal and statistically significant relationships. The experiment was conducted on the basis of determining the optimal model for forecasting the risk of the spread of one of the most significant threats in Ukraine – data confidentiality breach (54.67%), depending on the variables that characterize the capabilities / vulnerabilities of the cyber security system in Ukraine. The experiment showed that the optimal model emphasizes the predictors characterizing the vulnerability of the organizational cyber security system – "Departmental level of cybersecurity monitoring" and capabilities: "The level of use of risk management approaches at the operational level" and "The level of methodological support for cybersecurity of the critical infrastructure system".

Index Terms: Cybersecurity, cyber threats, data confidentiality breach, capabilities, vulnerabilities, risk-oriented approach, risk assessment, linear regression algorithm, predictive model, forecasting.

1. Introduction

Current trends and extremely high rates of informatization of social development require adequate security measures, as the protection of information, information and telecommunications systems is fundamental in society and protects the most important social values. Cybersecurity is an important area of national security in the information society era. Technological, economic and cultural development are directly related to cyberspace, which at the same time carries a variety of threats that significantly affect the level and development pace of society. Various foreign institutions constantly analyze cyber threats that are based on the quantitative assessment of incidents and facts known to experts or their specific probability.

In terms of hybrid war for Ukraine became real large-scale cyberattacks and their negative consequences. Considering military aggression against Ukraine and the need to directly protect the Ukrainian information space, critical infrastructure, socio-political and economic relations, the key is scientific justification, on theoretical basis using the results of experimental confirmation of key hypotheses of stability in cybersecurity. That is why it is important to use a risk-oriented approach, which is recommended by international acts as one of the key areas of ensuring cyber security and is based on assessment of both the likelihood and consequences of cyber threats. Along with this, a full-fledged analysis in this area cannot be completed by assessing only the risks of the spread of cyber threats, it is also

This work is open access and licensed under the Creative Commons CC BY 4.0 License.

important to assess the capabilities of the cyber security system in Ukraine and determine its vulnerabilities. Such problems, especially at the level of strategic analysis, are primarily solved within the scope of scientific applied research.

It is under these conditions that a risk-oriented approach in the field of cybersecurity plays a key role in forming a system of knowledge and awareness of ability of the cybersecurity system to withstand threats. Special components of the formation of sustainable cyber security system are also the development of predictive models based on the analysis and interpretation of dependence of the cyber threats' level on the vulnerability of the cyber security system based on the search for optimal and statistically significant connections, appropriate definition of the optimal model for forecasting the risk of the spread of the most significant threats in the cyber security system and optimization of practical measures, with defining the key components of the cyber security system.

2. Related Works

Much research has been done on cybersecurity. The directions of such research were chosen by various scientists. Some works focused on defining the principles of building a modern communication system and requirements for them [1], as well as recommendations for assessing the reliability of certain types of communication technology [2, 3]. Various proposed "ontological techniques" are considered, as well as comprehensive analysis of various models to ensure the safety of the cloud environment [4–7]. Much work has been done to identify and analyze software vulnerabilities, methods for reporting and classifying software security vulnerabilities [8–12]. Specific features of information protection focus on existing methods of localization of anomalies and current hazards in networks, statistical methods are considered as effective methods of detecting anomalies and experimental detection of the chosen method, methods of capturing and analyzing network traffic during passive monitoring of network segment [13–18]. The analysis of different areas of information security management in organizations of various activities is not left out: the management of information security system resources in the organization [19–21]; risk management and threats to information security [20, 22, 23]; management of documentation and information assistance system in the organization [19, 20, 24, 25]; information security audit management [23, 26, 27]; management of information security system efficiency analysis [19, 21, 28]. Some sources are associated with the use of different forecasting methods based on the construction of appropriate models [29–35].

Analysis of scientific sources, approaches to researching cybersecurity threats, identifying and assessing vulnerabilities, as well as developing models for predicting the impact on cyber threats, formed the author's approach to the methodology of further analysis, assessing the risks of cyber threats, assessing cybersecurity capabilities of data in the cybersecurity system of Ukraine [36].

3. Proposed Methods

The stages in this research flow shown in Fig. 1

Fig.1. Research Flow

3.1. Data Collection

The data set used in this study was obtained by identifying cyber threats, identifying indicators of cybersecurity capabilities and based on this survey of cybersecurity experts in Ukraine [36], and therefore reflects the professional experience of respondents and professional awareness of the survey. Questionnaires were completed on ON-LINE with confidentiality and anonymity. Data obtained in the ON-LINE mode are increasingly used to obtain information from respondents. This approach makes it possible to optimally learn the opinions of a large number of issues and a large number of experts who are at a distance. To ensure risk assessment, each indicator was assessed on two characteristics: "Likelyhood" and "Consequences".

3.2. Data cleaning

Data cleaning is the stage of preliminary data processing, cleaning them from unreliable expert sample. The statistical rationale for the sample restriction procedure is based on the fact that due to the large volume of the questionnaire, experts could make mistakes in the answers, as the complexity of questions and the short time of their comprehension leads to instability of attention [37]. In addition, the quality of data obtained in the ON-LINE mode may be significantly reduced due to incompetence of experts, lack of motivation to provide reasonable answers, as well as due to exhaustion or inattention in answering a large number of questions [38]. In order to extract the most reliable information from the obtained data, only those experts who provided logically consistent answers were selected.

3.3. Risk Assessment

The general methodological approach is based on the recommendations of ISO 31000 [39]. The author's is implementation of the ISO 31000 risk assessment algorithm. This applies to the system of indicators, which are formed into three groups (threats, capabilities and vulnerabilities of the cybersecurity system in Ukraine), data structure, evaluation grading and formation of the general expert set, as well as approaches to implementing certain methods and tools of data processing and analysis and interpretation of results [40].

3.4. Linear Regression Algorithm Forecasting

The task of researching complex systems and processes is often is checking the presence and establishing the type of relationship between independent variables (predictors, factors), the values of which may vary by the researcher and have a certain predetermined error, and the dependent variable [41-43]. Regression analysis includes methods for constructing mathematical models of the studied systems, methods for determining the parameters of these models and verifying their adequacy. It suggests that regression is a linear combination of linearly independent basis functions from factors with unknown coefficients (parameters). It is important to take into account multicollinearity, which causes the instability of the computational procedure due to high computational error. As a result, the interpretation of the results becomes impossible, and the values of particular coefficients – statistically insignificant. In some cases, to eliminate multicollinearity, related variables are alternately excluded and the results are compared. One of the methods of selecting the most significant factors is the stepwise regression procedure.

4. Results and Analysis

4.1. Data Collection

First of all, it should be noted that solving the issue of data collection had its problems. Members of the expert working group, which was formed to identify indicators that characterize threats, capabilities and vulnerabilities in the field of cybersecurity in Ukraine, could not clearly approach this task, due to their professional commitment, obsession, and thus bias in the expression of expert opinion. This was especially evident in the application of the risk-oriented approach. In our opinion, the limitations of their views were formed under the influence of understanding the problem exclusively within the technical sciences, IT technologies and did not take into account social, economic, psychological, law enforcement and security aspects. Therefore, modern world approaches and trends to characterize the components of cybersecurity and their rationale for use were taken into account.

European cybersecurity practice emphasizes that the uncontrolled use of cyberspace allows various destructive forces to spread cyber threats and dangers. And the biggest limitation on the effectiveness of national cybersecurity legislation is the inability to counter cyberspace.

The European Convention on Cybercrime was developed to solve this problem and it was adopted by the Committee of Ministers of the Council of Europe on 23 November 2001 [44]. During the preparation of the Convention on Cybercrime, the goal was to form a common law enforcement system to ensure cybersecurity and create conditions for the exchange of information between all signatory countries. The provisions of the Convention on Cybercrime also stipulate common rules for all ISPs to store customers' personal information in the event that such information is required in the investigation of cybercrime.

EU countries are demonstrating a common position on cybersecurity and human rights standards in cyberspace, which is dynamic and evolving through a rethinking of approaches. Given the level of development and key vectors of European democracy, the achieved level of human rights and freedoms, European cybersecurity policy always balances between state and public interests, which distinguishes the European model of cybersecurity, based on the social direction of domestic policy.

Thus, in the EU there is a clear distinction between the features of information security of person and society, information security of the state and international information security. At the same time, the interests of man and society became fundamental, which led to the intensive development of such areas as personal data security, access to information, as well as ensuring the implementation of democracy in the construction of the information society.

Hereby, a set of data characterizing threats in the field of cybersecurity is formed on the basis of 83 indicators. This list of indicators is based on different approaches:

- cyberattacks by distribution sectors (management, economy, infrastructure, defense, security, etc.);
- cyberincidentals in the same distribution sectors;
- cyber attacks on the subject (government agencies, law enforcement agencies, information resources, process control systems, etc.);
- according to the ENISA classification - *European Network and Information Security Agency* (top 15 threats);
- variety of malware (virus, worm, ransomware, trojan-malware, dialer, rootkit, polymorphic and metamorphic, XSS, keylogger etc.);
- variety of information-gathering (scanner, variety of sniffing, social-engineering etc.);

- variety of intrusion;
- fundamental cyber threats are presented NCSI (*National Cyber Security Index*) *e-Governance Academy* (Violation of data security: Data confidentiality breach – secrecy is exposed; Data integrity breach – unauthorized modification; Denial of e-services – services are not accessible).

We have also proposed 66 indicators of cybersecurity abilities in Ukraine, which can be grouped into the following groups:

- abilities at the national level;
- abilities of scientific and analytical nature;
- precautionary abilities;
- abilities according to the level of communication of NSCS subjects;
- abilities by the nature of competence;
- abilities in terms of efficiency and responsiveness.

We also used 21 indicators of vulnerabilities of the cybersecurity system, which are grouped into three homogeneous groups:

- technological vulnerabilities;
- technical and software vulnerabilities;
- legal and organizational vulnerabilities.

4.2. Data cleaning

For further analysis, it was necessary to limit the sample to the most qualitative and reliable data, i.e. to check the respondents for their logical error. For this purpose, during the development of the questionnaires, questions were entered in various sections, the logical answer to which was their assessment in the characterization of "Likelyhood" as "high" or "medium", but not "low" in wartime. For example, "Cyberattacks as an element of hybrid warfare" and so on.

Therefore, in the basic set of further analysis there were 508 questionnaires of only those experts who provided logically consistent answers, which is 63.66%. Despite the fact that after filtering the data there were 63.66% of the initial sample, the quality of the results increased significantly. It can be seen in the example of evaluating the indicator under question 1.8 "Cyberattacks in the field of defense" (Section 1 "Threats") and the distribution in the group of those who were selected for the filter no logical errors, compared with those who did not pass this filter (Table 1).

Table 1. Logical error filter analysis

Threats (Likelyhood)		**Sample**		**Total**
		Unreliable part	**Reliable part**	
1.8. Cyberattacks in the field of defense	low	34,4%	8,0%	11,0%
	average	37,5%	27,9%	29,0%
	high	28,1%	64,1%	60,1%
Total		100,0%	100,0%	100,0%

As you can see, the difference in distributions is glaring: 34.4% of unreliable experts indicated a low probability of a threat, while reliable ones chose this option in only 8.0% of cases. The "high probability" option was chosen by them in 64.1% of cases. This difference is not only statistically significant (criterion $\chi^2 = 25.102$, $p < 0.000$), but also the magnitude of the effect is very significant (V Cramer = 0.298, $p < 0.000$). Similar trends are observed in other important questions of the questionnaire.

Hereby, sampling constraints based on logical error checking are statistically significant, reliable, and representative.

4.3. Risk Assessment

A risk-oriented approach is one of the priorities of the cyber security system, which is determined by international legal acts. In particular, UN General Assembly Resolution 57/239 (December 20, 2002), "Elements of a global cyber security culture" [45] emphasizes the need for risk assessment to identify threats and vulnerabilities.

The EU Directive on Measures to Ensure a High General Level of Security of Network and Information Systems (NIS Directive) [46] complements the list of international legal acts, which determine the assessment of cyber security risks as one of the priorities. But it should be noted that the NIS Directive, although it defines the standards for ensuring cyber security in the jurisdictions of EU members, nevertheless provides the opportunity for countries to implement the specified requirements of the Directive by introducing their own legal mechanism into the system of national legislation. In this regard, it is important to highlight that in order to implement the the Directive's requirements, implementation of risk management in the cyber security system is one of the four key priorities along with increasing the capacity of the

cyber security system and pan-European cooperation at the national level, as well as the obligation of digital service operators to report cyber incidents.

Based on the mentioned international legal acts, we can confidently state that the implementation of risk management in the cyber security system is an obligation of each individual country and it is motivated by the need to increase the level of knowledge regarding the possible vulnerability of the cyber security system.

Given the challenge of hybrid war in Ukraine, and extreme importance of cyber threats in it, it is important to pay attention to concrete legal acts of the EU that also emphasize the need to implement risk management: Joint Report to the European Parliament, the European Council and the Council on the Implementation of the Joint Framework on countering hybrid threats (April 6, 2016 [47]; July 19, 2017 [48]; June 13, 2018 [49]) and Joint Staff Working Document Report on the implementation of the 2016 Joint Framework on countering hybrid threats and the 2018 Joint Communication on increasing resilience and bolstering capabilities to address hybrid threats [50]. In general, these are periodic reports to the European Community on countering hybrid threats, which focus on raising awareness in this area of society's resilience, increasing crisis prevention capacity, and expanding further international cooperation. It is in the context of raising awareness of hybrid cyber threats that emphasis is placed on identifying significant system vulnerabilities and conducting risk analysis.

That is why further analysis focuses on determining the level of risk of cyber threats. The implemented methodology involves risk assessment of the spread of threats on a scale from 0% to 100% and provides for the following threshold levels [51]: above 60% - red risk zone (the most significant threats); 50 - 60% - orange risk zone (significant threats); 40 - 50% - yellow risk zone - threats that need attention; 40% - green risk zone.

The overall picture of risk assessment of the identified threats is quite variable, which is also perceived as a level of adequacy in terms of representativeness of the results (Fig. 2).

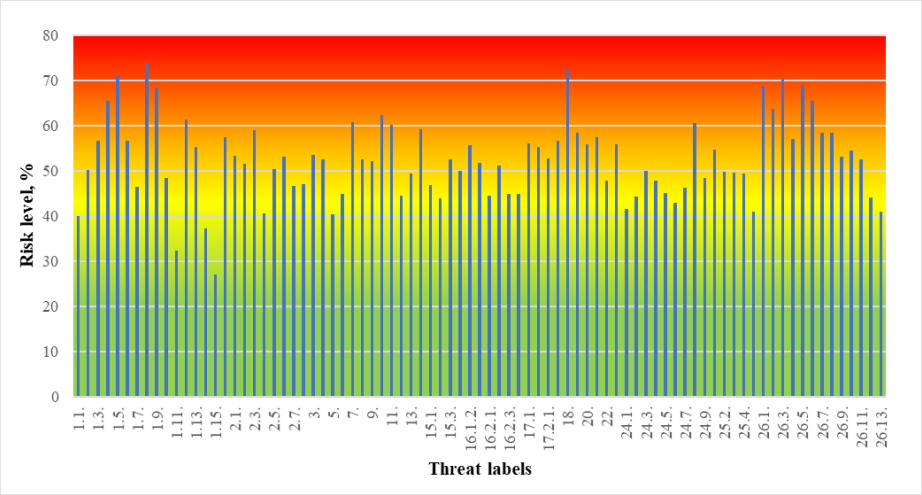

Fig.2. Overall ranking of cybersecurity threats

From the whole list of cyber threats, further analysis focuses on Data confidentiality breach – secrecy is exposed, which is among the threats that characterize violation of data security, has the highest level of risk – 54,67% (Fig. 3).

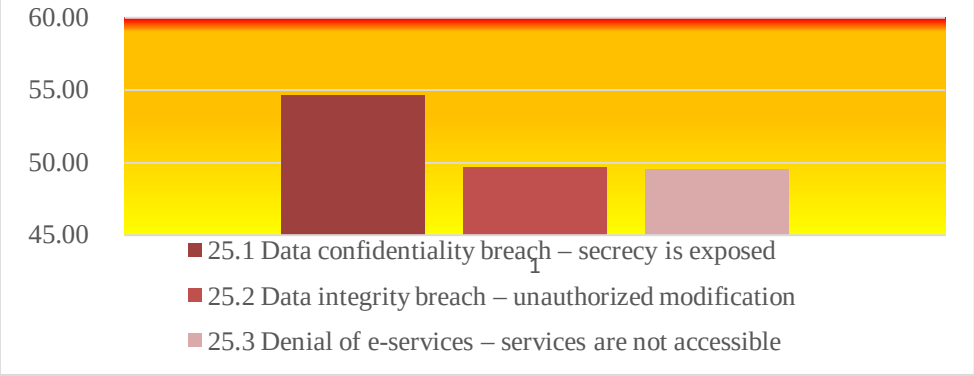

Fig.3. Cyber threat rating– violation of data security

It should also be noted that our analysis is based on a comparison of the mean value in assessing the level of risk. Expert assessments, despite the reliability of the sample, still have some differences (Fig.4.) But we perceive this as the variability of expert assessment, which once again emphasizes the representativeness of the empirical basis. Along with

this, the use of the average value of expert evaluation is a common approach in our interpretation and is used by us mainly to analyze the trend, compare the risk of other indicators, as well as to build a forecast model.

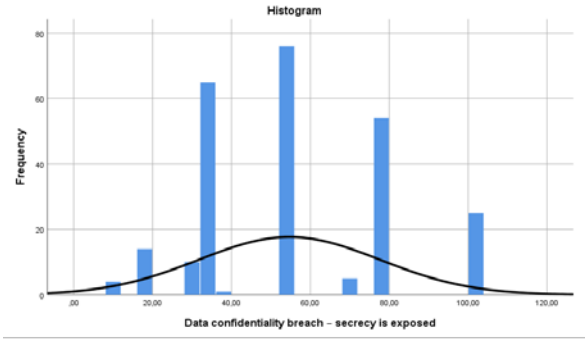

Fig.4. Variability of cyber threat assessment by experts– data confidentiality breach

Further analysis will focus on assessing cybersecurity abilities. A total of 66 ability indicators were identified and assessed (Fig. 5).

A calculation algorithm (experimentally substantiated) was used, in which the indicators of ability to risk assessment above 50% are characterized by a positive level, and below 50% - negative (as an insufficient level) and the content is more consistent with system vulnerabilities.

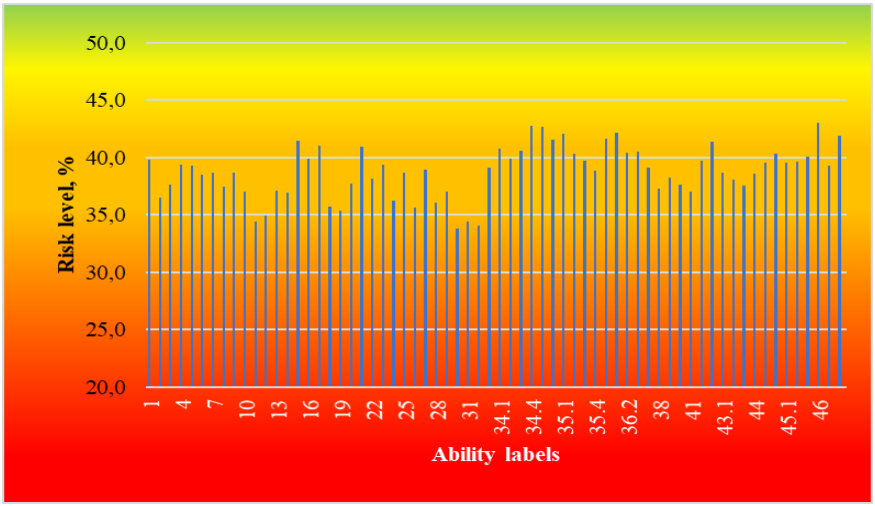

Fig.5. Overall rating of cybersecurity abilities

From the list of indicators characterizing the ability, all are in a high risk zone, which determines only the vulnerability of the cybersecurity system (less than 50% on a rating scale from 0 to 100%).

Vulnerabilities have been analyzed in 3 groups and the most relevant and potentially vulnerable have been identified:

- technological vulnerabilities: the next generation of mobile communications or 5G (62.71%); Internet of Things (IoT) (62.57%); quantum technologies (65.56%);
- technical and software vulnerabilities: the most significant among technical vulnerabilities are information and telecommunication systems (54.27%), and software vulnerabilities - disadvantages of software code (53.65%);
- legal and organizational vulnerabilities: legal: compliance with state and industry standards (61.55%) and adequacy of the level of responsibility for violations of cybersecurity legislation (60.43%); organizational: departmental monitoring level of cybersecurity (60.99%) and the level of public-private cooperation in the field of cybersecurity (62.95%).

4.4. Linear Regression Algorithm Forecasting

The analysis of internal factors (capabilities / vulnerabilities) does not end with their descriptive statistics. It is important to compare each threat identified and assessed by the cybersecurity expert community with the system of capabilities / vulnerabilities (a number of indicators).

EU directives on combating hybrid threats explicitly point to the need to find hidden links, including on specific threats with key capabilities / vulnerabilities in their impact [47-49].

This problem is solved by building an appropriate forecast model.

Hypothetically, internal factors are defined as our abilities or vulnerabilities, are characterized by a certain statistical relationship with particular cybersecurity threats and may directly or indirectly reduce the risk of the threat spreading.

Using the linear regression model, the optimal dependence of individual cybersecurity threats in Ukraine on variables characterizing the capabilities or vulnerabilities of the national cybersecurity system, including the data confidentiality breach, was determined by the method of phased inclusion, phasing out and finding the best subsets (Fig. 6) with the introduction of new technologies.

Because multiple linear regression is based on the large number of variations in possible sets of independent variables, we used IBM SPSS Statistics software for analysis.

As we have noted, general methodological approaches to the use of linear regression require consideration of the situation when independent variables are in some way characterized by affinity, i.e. multicollenarity, which affects the adequacy of the model and results of interpretation and conclusions. It is important to take into account multicollinearity, which causes the instability of the computational procedure due to high computational error. As a result, the interpretation of the results becomes impossible, and the values of particular coefficients – statistically insignificant.

This is exactly the situation when using the entire list of independent variables that characterize the capabilities of the cybersecurity system (Rs1 - Rs66) and vulnerabilities (Va1 - Va21). To avoid multicollinearity, further analysis was performed using independent variables for certain homogeneous groups.

```
REGRESSION
/DESCRIPTIVES MEAN STDDEV CORR SIG N
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT TA66
/METHOD=STEPWISE Rs1 Rs2 Rs3 Rs4 Rs5 Rs6 Rs7 Rs8 Rs9 Rs10 Rs11 Rs12 Rs13 Rs14 Rs15 Rs16 Rs17
  Rs18 Rs19 Rs20 Rs21 Rs22 Rs23 Rs24 Rs25 Rs26 Rs27 Rs28 Rs29 Rs30 Rs31 Rs32 Rs33 Rs34 Rs35 Rs36 Rs37
  Rs38 Rs39 Rs40 Rs41 Rs42 Rs43 Rs44 Rs45 Rs46 Rs47 Rs48 Rs49 Rs50 Rs51 Rs52 Rs53 Rs54 Rs55 Rs56 Rs57
  Rs58 Rs59 Rs60 Rs61 Rs62 Rs63 Rs64 Rs65 Rs66 Va1 Va2 Va3 Va4 Va5 Va6 Va7 Va8 Va9 Va10 Va11 Va12 Va13
  Va14 Va15 Va16 Va17 Va18 Va19 Va20 Va21.
```

Fig.6. Syntax of the linear regression model "TA66 - data confidentiality breach" (IBM SPSS Statistics)

The first regression model is based on the dependence of the variable "Data confidentiality breach" on predictors that characterize vulnerabilities in the field of cybersecurity in Ukraine (Fig. 7, 8).

The adequacy of the obtained results of the linear regression model is characterized by the statistical significance of the obtained results (Significance ≤ 0.05).

From the general list of indicators (Va1 - Va21) that characterize the vulnerability in cybersecurity in Ukraine, based on linear regression analysis identified one key predictor for reducing the risk of cybersecurity threat "Data confidentiality breach" – Organizational: Departmental level of cybersecurity monitoring. Given that the regression coefficient is 0.571, and using a linear regression diagram, we can conclude that the risk of spreading the threat of "Data confidentiality breach" reduces the level of vulnerability of the cybersecurity system in the organizational aspect: Departmental level of cybersecurity monitoring. Given the above, it is possible to predict such a decline in stages by 20% and 40% (Table 2).

Thus, a 40% reduction in vulnerability according to the predictor "Organizational: Departmental level of cybersecurity monitoring" will lead to a reduction in the risk of spreading the threat "Data confidentiality breach" to the yellow level - 43.83%.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	32,511	3,595		9,043	0,000
	Organizational: Departmental level of cybersecurity monitoring	0,571	0,082	0,482	6,998	0,000

a. Dependent Variable: Data confidentiality breach – secrecy is exposed

Fig.7. Linear regression model "Data confidentiality breach" taking into account existing vulnerabilities (IBM SPSS Statistics)

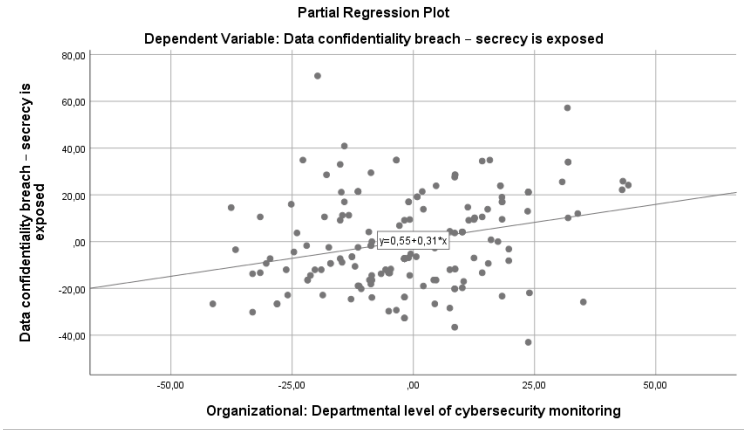

Fig.8. Linear regression diagram "Data confidentiality breach" taking into account the vulnerability "Organizational: Departmental level of cybersecurity monitoring" (IBM SPSS Statistics)

Table 2. Forecast of the threat level under the condition of changing the level of simulated predictors

THREAT	RISK LEVEL			
	Basic level	Simulated level	Changing predictors	
			20 %	40 %
Data confidentiality breach	54,67 %	**66,67 %**	55,25 %	43,83 %

For further analysis, we use step by step two groups of indicators of cybersecurity capabilities: "abilities of scientific and analytical nature" and "precautionary abilities".

Initially, the regression model was built based on the dependence of "Data confidentiality breach" on predictors (Rs8, Rs24, Rs26, Rs27, Rs28, Rs29, Rs31, Rs32, Rs58, Rs59, Rs60), which characterize the ability of cybersecurity in the field of scientific and analytical support. Fig. 9, 10).

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
2	(Constant)	104,924	6,955		15,087	0,000
	The level of use of risk management approaches at the operational level	-0,412	0,102	-0,333	-4,018	0,000
	The level of methodological support for cybersecurity of the critical infrastructure system	-0,376	0,104	-0,299	-3,610	0,000

a. Dependent Variable: Data confidentiality breach – secrecy is exposed

Fig.9. Linear regression model "Data confidentiality breach" taking into account abilities of scientific and analytical nature (IBM SPSS Statistics)

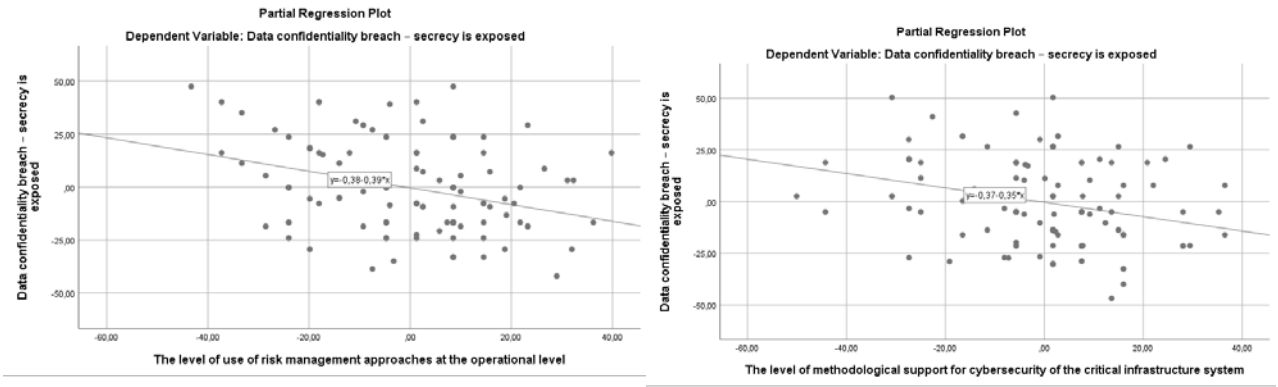

Fig.10. Linear regression diagram "Data confidentiality breach" taking into account abilities of scientific and analytical nature (IBM SPSS Statistics)

From the general list of indicators that characterize the ability of the cybersecurity system in the field of scientific and analytical support, based on linear regression analysis, two key predictors to reduce the risk of cybersecurity in the field of cybersecurity "Data confidentiality breach": "The level of use of risk management approaches at the operational level" and "The level of methodological support for cybersecurity of the critical infrastructure system". Given that the regression coefficients have the opposite values of -0.412 and -0.376, using a linear regression diagram, we can conclude that the risk of spreading the threat of "Data confidentiality breach" reduces the level of cybersecurity capabilities taking into account the abilities of scientific and analytical nature. Given the above, it is possible to predict such a decline in stages by 20% and 40% (Table 3).

Table 3. Forecast of the threat level under the condition of changing the level of simulated predictors

THREAT	**RISK LEVEL**					
	Basic level	**Simulated level**	**Changing predictors**			
			20 %		**40%**	
			The level of use of risk management approaches at the operational level	The level of methodological support for cybersecurity of the critical infrastructure system	The level of use of risk management approaches at the operational level	The level of methodological support for cybersecurity of the critical infrastructure system
Data confidentiality breach	54,67 %	**74,64 %**	58,9 %		43,15 %	

Thus, increasing the ability according to the predictors "The level of use of risk management approaches at the operational level" and "The level of methodological support for cybersecurity of the critical infrastructure system" by 40% will lead to a decrease in the risk of spreading the threat "Data confidentiality breach" to the yellow level - 43.15%.

This analysis requires other subgroups that characterize the capabilities of the cybersecurity system, but we will complete the analysis and build a regression model based on the dependence of "Data confidentiality breach" on predictors (Rs15, Rs16, Rs17, Rs18, Rs19, Rs20, Rs21, Rs22, Rs23, Rs30, Rs48, Rs49, Rs50, Rs51, Rs52, Rs53), which determine the precautionary nature of the capabilities of the cybersecurity system (Fig. 11, 12).

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
2	(Constant)	102,816	6,512		15,789	0,000
	The level of culture to ensure the confidentiality of business and private communications through electronic means	-0,433	0,100	-0,358	-4,327	0,000
	Level of cryptographic protection of information	-0,317	0,092	-0,286	-3,461	0,001
a. Dependent Variable: Data confidentiality breach – secrecy is exposed						

Fig.11. Linear regression model "Data confidentiality breach" taking into account precautionary abilities (IBM SPSS Statistics)

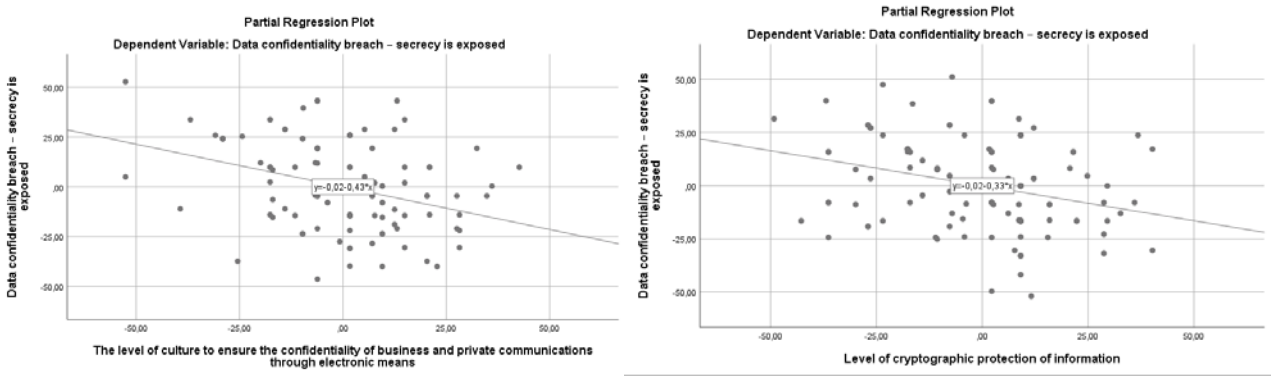

Fig.12. Linear regression diagram "Data confidentiality breach" taking into account precautionary abilities (IBM SPSS Statistics)

From the general list of indicators that characterize the security capabilities of the cybersecurity system, based on linear regression analysis, two key predictors have been identified to reduce the risk of cybersecurity threats: "Data confidentiality breach": "The level of culture to ensure the confidentiality of business and private communications through electronic means" and "Level of cryptographic protection of information". Taking into account that regression coefficients have the opposite values -0.433 and -0.317, using a linear regression diagram, we can conclude that the risk

of spreading the threat of "Data confidentiality breach" under the condition of increasing the level of cybersecurity taking into account precautionary abilities. Taking into account the above, it is possible to predict such a decline in stages by 20% and 40% (Table 4).

Table 4. Forecast of the threat level under the condition of changing the level of simulated predictors

THREAT	RISK LEVEL					
			Changing predictors			
			20 %		40%	
	Basic level	Simulated level	The level of culture to ensure the confidentiality of business and private communications through electronic means	Level of cryptographic protection of information	The level of culture to ensure the confidentiality of business and private communications through electronic means	Level of cryptographic protection of information
Data confidentiality breach	54,67 %	74,2 %	59,21 %		44,22 %	

Thus, increasing the ability according to the predictors "The level of culture to ensure the confidentiality of business and private communications through electronic means" and "Level of cryptographic protection of information" by 40% will lead to a decrease in the risk of spreading the threat "Data confidentiality breach" to yellow level - 44.22%.

5. Conclusion

Therefore, in order to predict changes in the level of threats in the field of cybersecurity, it is important not to complete the analysis of internal factors (capabilities / vulnerabilities) with their descriptive statistics. Comparison of each threat identified and assessed in the field of cybersecurity with the system of capabilities / vulnerabilities is solved by building an appropriate forecast model.

Applying the model of linear regression, by the method of gradual inclusion, gradual exclusion and search for the best subsets, the optimal dependence of a particular threat in the field of cyber security of Ukraine on variables characterizing the capabilities or vulnerabilities of the national cybersecurity system:

1) from the general list of indicators characterizing the vulnerability in the field of cybersecurity in Ukraine, a key predictor for reducing the risk of spreading the threat in the field of cybersecurity "Data confidentiality breach" – Organizational: Departmental level of cybersecurity monitoring.

2) from the list of indicators characterizing the ability of the cybersecurity system in the field of scientific and analytical support, two key predictors for reducing the risk of spreading the threat in the field of cybersecurity "Data confidentiality breach": "The level of use of risk management approaches at the operational level" and "The level of methodological support for cybersecurity of the critical infrastructure system".

3) it is indicated from the list of indicators that characterize the security capabilities of the cybersecurity system, based on linear regression analysis, two key predictors to reduce the risk of cybersecurity threat "Data confidentiality breach": "The level of culture to ensure the confidentiality of business and private communications through electronic means" and "Level of cryptographic protection of information".

Substituting into the forecast model and gradually, changing by 20% and 40%, the values of the estimated level, identified key predictors, the reduction of the cyber threat "Data confidentiality breach" is predicted.

References

[1] A. Tikhomirov, N. Kinash, S. Gnatyuk, A. Trufanov, O. Berestneva et al. (2014). Network Society: Aggregate Topological Models, Communications in Computer and Information Science. Verlag: Springer International Publ. Vol. 487. Pp. 415-421.

[2] A.V. Kharybin, O.N. Odaryshchenko (2006). About the approach to the decision of questions of a choice of methodology of an estimation of system reliability and survivability of information systems of critical application. Radiotechnical and computer systems. Kh.: NAU "KhAI". No. 6 (18). Pp. 61–70.

[3] Biyue Diao, Guoping Chen, Feng He, " Loudspeaker Operation Status Monitoring System based on Power Line Communication Technology", International Journal of Image, Graphics and Signal Processing, Vol.10, No.10, pp. 54-62, 2018.

[4] Naila Samad Shaikh, Affan Yasin, Rubia Fatima, "Ontologies as Building Blocks of Cloud Security", International Journal of Information Technology and Computer Science, Vol.14, No.3, pp.52-61, 2022.

[5] P. Bhandari, M. S. Gujral (2014). Ontology Based Approach for Perception of Network Security State. Proceedings of 2014 RAECS UIET Panjab University Chandigarh, 06 – 08 March.

[6] K. Bernsmed, A. Undheim, P. Hakon Meland, M. G. Jaatun (2013). Towards an Ontology for Cloud Security Obligations. International Conference on Availability, Reliability and Security.
[7] N. F. Noy, McGuinness, D. L., “Ontology development 101: “A guide to creating your first ontology". Stanford University, Stanford, CA, 94305, 2001
[8] Hakan KEKÜL, Burhan ERGEN, Halil ARSLAN, " Estimating Missing Security Vectors in NVD Database Security Reports", International Journal of Engineering and Manufacturing, Vol.12, No.3, pp. 1-13, 2022.
[9] P. Mell, K. Scarfone, and S. Romanosky (2007). A Complete Guide to the Common Vulnerability Scoring System Version 2.0. FIRSTForum of Incident Response and Security Teams. Available at: https://www.first.org/cvss/cvss-v2-guide.pdf (accessed Jan. 01, 2021).
[10] G. Spanos, A. Sioziou, and L. Angelis (2013). WIVSS: A New Methodology for Scoring Information Systems Vulnerabilities. Proceedings of the 17th Panhellenic Conference on Informatics. Pp. 83–90.
[11] Hakan Kekül, Burhan Ergen, Halil Arslan, " A New Vulnerability Reporting Framework for Software Vulnerability Databases", International Journal of Education and Management Engineering, Vol.11, No.3, pp. 11-19, 2021.
[12] Muhammad Noman Khalid, Muhammad iqbal, Kamran Rasheed, Malik Muneeb Abid, "Web Vulnerability Finder (WVF): Automated Black- Box Web Vulnerability Scanner", International Journal of Information Technology and Computer Science, Vol.12, No.4, pp.38-46, 2020.
[13] Abhinandan H. Patil, Neena Goveas, Krishnan Rangarajan,"Regression Test Suite Prioritization using Residual Test Coverage Algorithm and Statistical Techniques", International Journal of Education and Management Engineering, Vol.6, No.5, pp.32-39, 2016.
[14] R. Ranjan, G. Sahoo (2014). A new clustering approach for anomaly intrusion detection. International Journal of Data Mining & Knowledge Management Process (IJDKP). Vol. 4. No. 2. Pp. 29–38.
[15] Serhii Zybin, Yana Bielozorova, "Risk-based Decision-making System for Information Processing Systems", International Journal of Information Technology and Computer Science, Vol.13, No.5, pp.1-18, 2021.
[16] I. Parkhomey, S. Gnatyuk, R. Odarchenko, T. Zhmurko et al, “Method For UAV Trajectory Parameters Estimation Using Additional Radar Data”, Proceedings of the 2016 4th International Conference on Methods and Systems of Navigation and Motion Control, Kyiv, Ukraine, October 18-20, 2016, pp. 39-42.
[17] Falaye Adeyinka A, Etuk Stella Oluyemi, Adama Ndako Victor, Ugwuoke Cosmas Uchenna, Olujimi Ogedengbe, Seun Ale,"Parametric Equation for Capturing Dynamics of Cyber Attack Malware Transmission with Mitigation on Computer Network", International Journal of Mathematical Sciences and Computing, Vol.3, No.4, pp.37-51, 2017.
[18] Yaser Ghaderipour, Hamed Dinari. " A Flow-Based Technique to Detect Network Intrusions Using Support Vector Regression (SVR) over Some Distinguished Graph Features ", International Journal of Mathematical Sciences and Computing, Vol.6, No.4, pp.1-11, 2020.
[19] Peltier, Thomas R. (2016). Information Security Policies, Procedures, and Standards: guidelines for effective information security management. Auerbach Publications.
[20] Nazareth, Derek L., and Jae Choi (2015). A system dynamics model for information security management. Information & Management. Vol. 52 (1). Pp. 123-134.
[21] Layton, Timothy P. (2016). Information Security: Design, implementation, measurement, and compliance. Auerbach Publications.
[22] Joshi, Chanchala, and Umesh Kumar Singh (2017). Information security risks management framework–A step towards mitigating security risks in university network”. Journal of In formation Security and Applications. Vol. 35. Pp. 128-137.
[23] Soomro, Zahoor Ahmed, Mahmood Hussain Shah, and Javed Ahmed (2016). Information securi ty management needs more holistic approach: A literature review. International Journal of Information Management. Vol. 36 (2). Pp. 215-225.
[24] Grudzień, Łukasz, and Adam Hamrol (2016). Information quality in design process documenta tion of quality management systems. International Journal of Information Management Vol. 36 (4). Pp. 599-606.
[25] E. Lavrov, A. Tolbatov, N. Pasko, V. Tolbatov (2017). Cybersecurity of distributed information systems. The minimization of damage caused by errors of operators during group activity, Proceedings of 2017 2nd International Conference on Advanced Information and Commu nication Technologies, AICT 2017. Pp. 83-87.
[26] Ilya Livshitz, Pavel Lontsikh and Sergey Eliseev (2017). The optimization method of the inte grated management system security audit. 2017 20th Conference of Open Innovations Association (FRUCT). IEEE.
[27] Jacobs, Stuart (2015). Engineering information security: The application of systems engineering concepts to achieve information assurance. John Wiley & Sons.
[28] Yoon, Junseob, and Kyungho Lee (2016). Advanced assessment model for improving effective ness of information security measurement. International Journal of Advanced Media and Communication. Vol. 6 (1). Pp. 4-19.
[29] Kasliono, Suprapto, Faizal Makhrus, "Point Based Forecasting Model of Vehicle Queue with Extreme Learning Machine Method and Correlation Analysis", International Journal of Intelligent Systems and Applications, Vol.13, No.3, pp.11-22, 2021.
[30] A. Anbarasa Pandian, R. Balasubramanian,"Analysis on Shape Image Retrieval Using DNN and ELM Classifiers for MRI Brain Tumor Images", International Journal of Information Engineering and Electronic Business, Vol.8, No.4, pp.63-72, 2016.
[31] Muhammad Resa Arif Yudianto, Tinuk Agustin, Ronaldus Morgan James, Firstyani Imannisa Rahma, Arham Rahim, Ema Utami, " Rainfall Forecasting to Recommend Crops Varieties Using Moving Average and Naive Bayes Methods", International Journal of Modern Education and Computer Science, Vol.13, No.3, pp. 23-33, 2021.

[32] Volodymyr Lytvynenko, Olena Kryvoruchko, Irina Lurie, Nataliia Savina, Oleksandr Naumov, Mariia Voronenko, "Comparative Studies of Self-organizing Algorithms for Forecasting Economic Parameters", International Journal of Modern Education and Computer Science, Vol.12, No.6, pp. 1-15, 2020.
[33] Nor Hamizah Zulkifley, Shuzlina Abdul Rahman, Nor Hasbiah Ubaidullah, Ismail Ibrahim, " House Price Prediction using a Machine Learning Model: A Survey of Literature", International Journal of Modern Education and Computer Science, Vol.12, No.6, pp. 46-54, 2020.
[34] Muhammad Resa Arif Yudianto, Tinuk Agustin, Ronaldus Morgan James, Firstyani Imannisa Rahma, Arham Rahim, Ema Utami, " Rainfall Forecasting to Recommend Crops Varieties Using Moving Average and Naive Bayes Methods", International Journal of Modern Education and Computer Science, Vol.13, No.3, pp. 23-33, 2021.
[35] Gbadamosi Babatunde, Adeniyi Abidemi Emmanuel, Ogundokun Roseline Oluwaseun, Oladosu Bukola Bunmi, Anyaiwe Ehiedu Precious,"Impact of Climatic Change on Agricultural Product Yield Using K-Means and Multiple Linear Regressions", International Journal of Education and Management Engineering, Vol.9, No.3, pp.16-26, 2019.
[36] O.Ye. Korystin & O.O. Korystin (2022). Threats in the sphere of cyber security in Ukraine. Nauka i pravookhoronna. Vol. 1. Pp. 127–131.
[37] Goldammer, P., Annen, H., Stöckli, P. L., & Jonas, K. (2020). Careless responding in questionnaire measures: Detection, impact, and remedies. The Leadership Quarterly. Vol. 31 (4). 101384.
[38] Oleksandr Korystin, Nataliia Svyrydiuk, Alexander Vinogradov (2021). The Use of Sociological Methods in Criminological Research. Proceedings of the International Conference on Social Science, Psychology and Legal Regulation (SPL 2021). Series: Advances in Social Science, Education and Humanities Research. Vol. 617. 18 December. Pp.1-6.
[39] ISO 31000:2018 - RISK MANAGEMENT. Available at: https://www.iso.org/ru/publication/PUB100464.html
[40] O.Korystin, N. Svyrydiuk (2020). Methodological principles of risk assessment in law enforcement activity. Nauka i pravooxoronna. No. 3. Pp. 191-197.
[41] Kasliono, Suprapto, Faizal Makhrus, "Point Based Forecasting Model of Vehicle Queue with Extreme Learning Machine Method and Correlation Analysis", International Journal of Intelligent Systems and Applications, Vol.13, No.3, pp.11-22, 2021.
[42] Mohamed Zaim Shahrel, Sofianita Mutalib, Shuzlina Abdul-Rahman, " PriceCop–Price Monitor and Prediction Using Linear Regression and LSVM-ABC Methods for E-commerce Platform", International Journal of Information Engineering and Electronic Business, Vol.13, No.1, pp. 1-14, 2021.
[43] Oleksandr Korystin and Nataliia Svyrydiuk (2021). Activities of Illegal Weapons Criminal Component of Hybrid Threats. Proceedings of the International Conference on Economics, Law and Education Research (ELER 2021). Series: Advances in Economics, Business and Management Research. Vol. 170. 22 March. Pp. 86-91.
[44] Convention on Cybercrime. Budapest, 23.XI.2001.
[45] Creation of a global culture of cybersecurity. UN. General Assembly (57th sess. : 2002-2003). Available at: https://digitallibrary.un.org/record/482184
[46] The Directive on security of network and information systems (NIS Directive). Available at: https://ec.europa.eu/digital-single-market/en/network-and-information-security-nisdirective.
[47] Joint Communication to the European Parliament and the Council Joint Framework on countering hybrid threats a European Union response (2016).
[48] Joint Report to The European Parliament and the Council on the Implementation of the Joint Framework on countering hybrid threats - a European Union response (2017).
[49] Joint Report to the European Parliament, the European Council and the Council on the Implementation of the Joint Framework on countering hybrid threats from July 2017 to June 2018. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=JOIN:2018:014:FIN
[50] Joint Staff Working Document Report on the implementation of the 2016 Joint Framework on countering hybrid threats and the 2018 Joint Communication on increasing resilience and bolstering capabilities to address hybrid threats.
[51] Oleksandr Korystin, Nataliia Svyrydiuk, Volodymyr Tkachenko (2021). Fiscal Security of the State Considering Threats of Macroeconomic Nature. Proceedings of the International Conference on Business, Accounting, Management, Banking, Economic Security and Legal Regulation Research (BAMBEL2021). Series: Advances in Economics, Business and Management Research. Vol. 188. 27 August. Pp. 65-69.

Authors' Profiles

Oleksandr Korystin

DSc, PhD, Professor. In 2009 he received DSc degree in information law from NAIA. In 2014 he received Professor degree. Honored Academic of Science and Technology of Ukraine.

Chief Research Scientist of the Criminological Research Laboratory of the State Scientific Research Institute of the Ministry of Internal Affairs of Ukraine. In 2014–2016 – Rector of the Odesa State University of Internal Affairs. Member of the Expert Council of the Ministry of Education and Science of Ukraine on legal sciences. Research interests: criminology; economic security; cybersecurity; intelligence; methodology of strategical (SWOT-analysis; risks assessment); counteraction to the hybrid threat.

Svyrydiuk Nataliia

DSc, PhD, Professor. In 2016 she received DSc degree in information law from OSUIA. In 2021 she received Professor degree.

Deputy head of the Criminological Research Laboratory of the State Scientific Research Institute of the Ministry of Internal Affairs of Ukraine. Research interests: criminology; economic security; cybersecurity; intelligence; methodology of strategical (SWOT-analysis; risks assessment); counteraction to the hybrid threat.

Olena Mitina

PhD Philological Sciences, Associate Professor, Head of the English Philology and Translation Studies Dpt., Odesa National Polytechnic University. Research interests: English lexicology; interactive technologies for foreign languages; criminology; cybersecurity.

How to cite this paper: Oleksandr Korystin, Svyrydiuk Nataliia, Olena Mitina, "Risk Forecasting of Data Confidentiality Breach Using Linear Regression Algorithm", International Journal of Computer Network and Information Security(IJCNIS), Vol.14, No.4, pp.1-13, 2022. DOI:10.5815/ijcnis.2022.04.01