Research on Cybercrime Based on BP Neural Network Algorithm

Kaihao Si [1,*,†], Xinhao Zhang [1,†], Siyuan Wu [2,†]

[1] School of Cyberspacesecurity, Northwestern Polytechnical University, Xi'an, China, 710129

[2] School of Automation, Northwestern Polytechnical University, Xi'an, China, 710129

* Corresponding Author Email: sikaihao@icloud.com

† These authors contributed equally.

Abstract. Cybercrime is a variety of criminal activities carried out using digital devices or networks. Owing to its inherent complexity, countries have developed different policies and laws to safeguard cybersecurity. This paper shows the global distribution of different indicators of cybercrime, examines the relevance of different national policies, and explores the extent to which different demographic characteristics influence cybercrime. This article first defines four main indicators to describe cybercrime, and collects data related to cybercrime from VCDB and NETSCOUT to display the global distribution of these four indicators. A model for extracting distribution characteristics of cybercrime was constructed, and Pearson coefficient was used to analyze the relationship between cybercrime and many influencing factors. Finally, it was concluded that the distribution of cybercrime is concentrated in countries with larger populations, weaker network protection technologies and infrastructure, and weaker judicial systems. At the same time, a network security policy correlation analysis model was constructed using the five network security policy indicators provided by ITU. The GUI indicators were combined with the distribution and data preprocessing of cybercrime, and linear regression algorithm was used to analyze the correlation of GUI policy indicators. By using the BP neural network algorithm to reduce the impact of nonlinear relationships, it was found that the formulation of intergovernmental cooperation policies and government capacity building policies has a more significant effect on suppressing cybercrime. Finally, this article analyzes the impact of specific national policies on cybercrime.

Keywords: Cybercrime, BP neural network algorithm, grey correlation analysis, GUI.

1. Introduction

Network based smart infrastructures are becoming common and these infrastructures are often vulnerable to cyber attacks due to poor security [1]. Fereshteh Momeni [2] describes the state of the economy and unemployment as significant factors that increase the risk of cybercrime and expresses the urgent need for stabilizing the economy in order to reduce cybercrime. Shillair R. [3] argued that the lack of cyber security awareness likewise exacerbates the problem of cyber crime and there is a need to increase security awareness among these groups through education and training. Nguyen, T., Petrov, S., & Ivanov, A [4] expresses that the relationship between political stability and cybercrime also shows some complexity, with cybercrime rates tending to be higher in politically unstable countries. Based on the above, Folorunso A [5] argues that there is a need to strengthen and regularly update the relevant laws and their enforcement to cope with the increasing cybersecurity situation. The paper defined four major indicators to reflect the factors affecting the study of cybercrime, which are: the number distribution coefficient, the success rate, the reporting rate, and the appeal rate, and thus obtained the distribution pattern of cybercrime. In addition, in order to further obtain the distribution pattern, we used the Pearson correlation coefficient [6] to analyse the relationship between the distribution of cybercrime and a number of factors, and carried out a correlation test. The paper established the Cybersecurity Policy Relevance Analysis Model, which was firstly fitted using a linear regression algorithm [7] to find the correlation between the factors and the distribution, and at the same time, in order to prevent the linear model from ignoring the nonlinear relationship of the data itself, we used a BP neural network algorithm [8] to validate the conclusions reached and to draw

final conclusions. In order to prevent the linear model from ignoring the non-linear relationships in the data itself, a BP neural network algorithm was used to validate the conclusions drawn, to draw final conclusions and to improve the generalisation of the model.

2. Research on the Distribution of Cybercrime

2.1. Key indicators of cybercrime

In order to comprehensively analyze the influencing factors and distribution characteristics of cybercrime, this article defines four core indicators, as shown in Figure 1. These four indicators form a comprehensive analytical framework across four dimensions, encompassing multi-level influencing the technical, social, and institutional aspects. This framework ensures a thorough revelation of the distribution patterns and driving mechanisms of cybercrime.

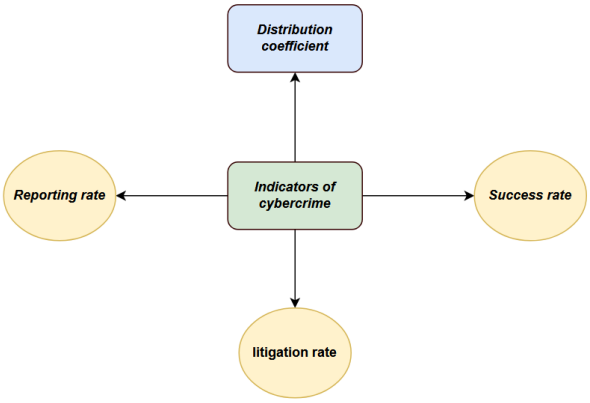

Figure 1. Four Key Indicators of Cybercrime

2.1.1 Quantitative Distribution Coefficient

The quantity distribution coefficient D_i reflects the distribution of cybercrime on a global scale. By counting the number of cybercrimes occurring in each country or region in a certain period of time, D_i can reveal the high incidence areas and patterns of cybercrimes.

2.1.2 Success rate

The success rate S_i indicates the proportion of cybercrimes that achieve their purpose after being committed. This indicator directly reflects the efficiency of cybercrime. A lower S_i usually indicates that the target country's protection system is sound, while a higher S_i may expose the problem of protection loopholes.

2.1.3 Reporting Rate

The reporting rate R_i measures the proportion of cybercrime incidents that are reported by relevant agencies or the public. This indicator reflects the transparency of cybercrime and the attitude of the public or organizations towards cybersecurity incidents.

2.1.4 Prosecution Rate

The prosecution rate P_i reflects the percentage of cybercrime incidents that can be traced and prosecuted. This indicator demonstrates the effectiveness of a country or region's legal enforcement, transnational cooperation, and law enforcement capabilities.

2.2. Analysis of Indicators

The four indicators are selected to comprehensively assess a country’s cybercrime and security landscape. They integrate technical, social, and legal dimensions, offering multi-dimensional insights. By synthesizing these, stakeholders can identify vulnerabilities, assess governance effectivenes, benchmark against global trends, and inform targeted policies.

2.2.1Data Integration

The collection of sufficient data is the basis for a complete system of indicators. Our data mainly come from VCDB (Veris Community Database), ITU (International Telecom-munication Union), NETSCOUT, World Bank, National Bureau of Statistics, etc. The database in Table 1 does not represent all the data we used, only representative data are shown here.

Table 1. Data set sources

Database	Website
VCDB	https://verisframework.org/index.html
NETSCOUT	https://www.netscout.com/threatreport
World Bank	https://data.worldbank.org
ITU	https://www.itu.int/en/ITU-D/Cybersecurity/Pages/Global-Cybersecurity-Index.aspx

This article combines data on cybercrime incidents by country to form a comprehensive dataset that includes indicators for each different type of cyber attack event. During this process, ensure that the average data of network event indicators for each country matches correctly for subsequent analysis, as shown in Table 2.

Table 2. Country Analysis

	Max Bandwidth(Gbps)	Max Throughput(Mpps)	Average Duration(Minutes)	Attack Frequency(Attacks)
China	999.93	203.37	30.67	236,951
Korea	459.7	173.69	49.56	351,049
Russian	996.03	243.43	35.5	178,643
America	939.43	335.49	45.75	210,896
India	275.9	76.79	166.74	227,357
Brazil	798.72	80.43	51.85	142,825
Canada	377.34	253.56	24.05	120,722
Japan	411.53	210.62	23.64	187,547

*For reasons of space, only representative and relevant countries are shown here.

2.2.2 Quantitative Distribution Analysis

Quantitative distribution analysis, as shown in Figure 2.

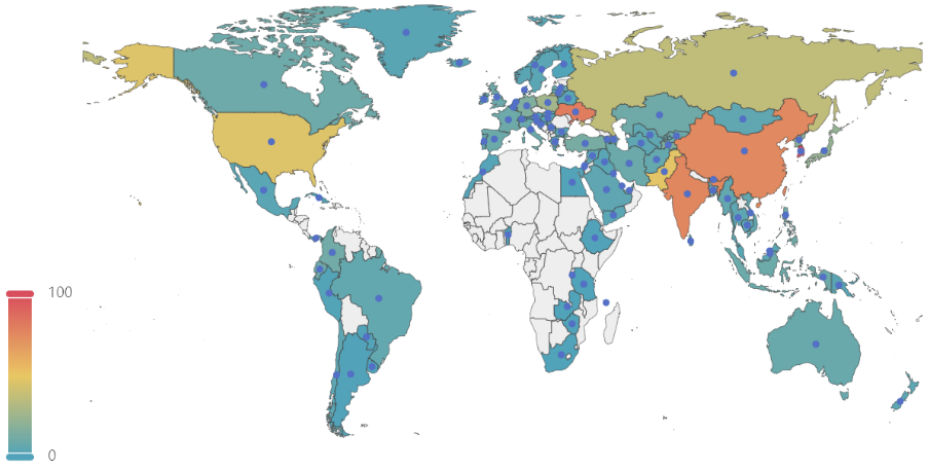

Figure 2. Quantitative Distribution Heatmap

The volume distribution coefficients show significant differences on a global scale. China and India has the largest number of Internet users in the world, and its fast-growing economy and extensive data-intensive industries make it a major target for cybercrime. Meanwhile, weak network security protection and insufficient user security awareness in some regions further push up D_i.

The U.S. is technologically advanced with strong cybersecurity protection, but as a global economic center, its high-value targets attract sophisticated cybercrime, keeping D_i. in the middle of the pack.

Russia's complex geopolitical environment makes Ukraine a target for transnational cyberattacks. The combination of a weak economy and insufficient protection capabilities, coupled with high levels of criminal activity in some parts of the country, has led to high D_i..

Europe has well-developed cybersecurity policies, advanced technology and strong public security awareness. Extensive international cooperation and high investment in the EU have resulted in a low success rate of cybercrime, keeping D_i. at a low level overall.

In order to further explore the relationship between the quantity distribution coefficients and other variables, we randomly selected a number of countries and analyzed the quantity distribution coefficients of these countries in conjunction with their population sizes.

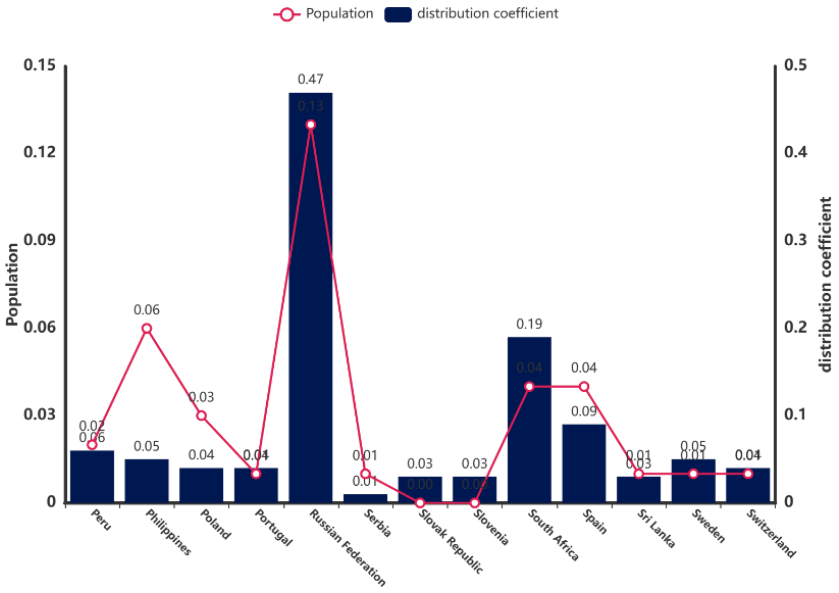

Figure 3. Relevance Analysis Chart

As shown in Figure 3, the coefficient of quantity distribution is correlated with population size. In countries with a large population base, the coefficient of quantity distribution is usually higher.

2.2.3 Success Rate Analysis

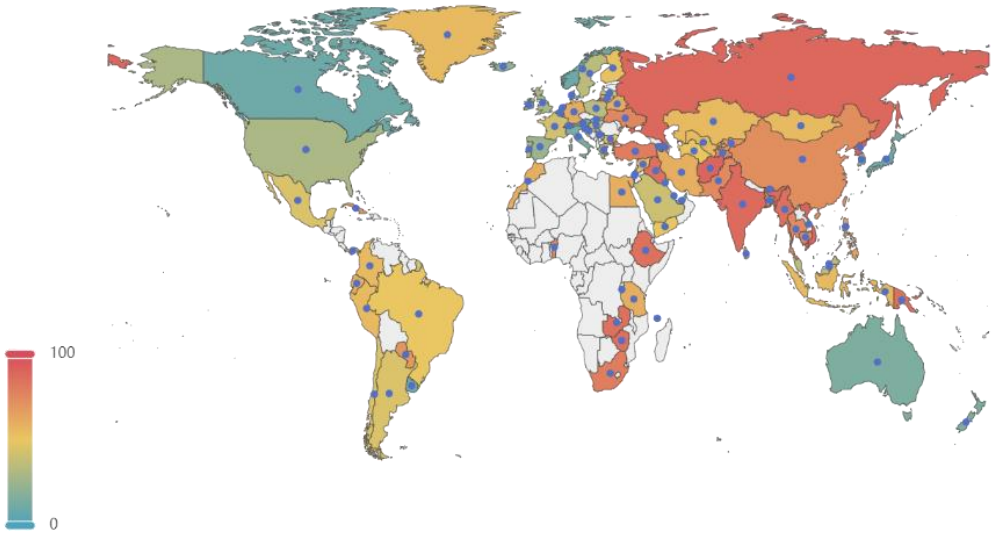

Figure 4. Success Rate Heatmap

As shown in Figure 4, the high success rate of cybercrime in some Asian countries is mainly due to inadequate law enforcement and uneven network protection technology and infrastructure in some regions. In addition, the huge Internet user group, coupled with rapid digital transformation, has not fully matured the network security system, providing more opportunities for offenders. The higher level of some Eastern European countries is limited by weak enforcement of legal systems, as well as weak network regulation and cross-border cooperation mechanisms. Western European countries have high-level law enforcement and strong network protection technologies, as well as strong regulatory mechanisms and attention to data protection. The lower level of the United States is due to its highly advanced network protection technology and strict law enforcement system.

2.2.4 Reporting Rate Analysis

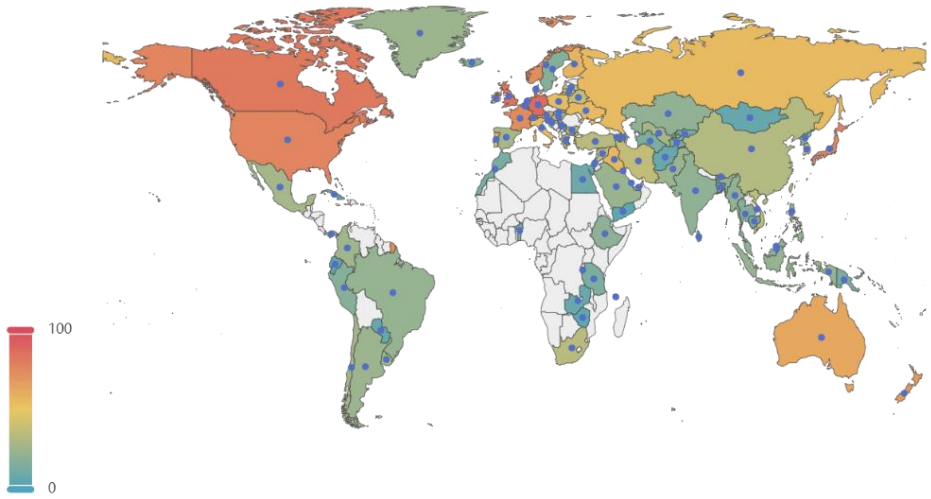

Figure 5. Reporting Rate Heatmap

According to Figure 5, in developed countries such as the United States and Western Europe, citizens who have suffered from cybercrime are more likely to report receiving treatment, with a higher proportion. On the contrary, in Southeast Asia, such as Myanmar and Thailand, citizens in some areas with lower levels of education may not be able to identify cybercrime or may not know how to report these crimes. Citizens of countries with higher education usually have a better understanding of the law, cybercrime and its prevention, and are more sensitive to dealing with cybercrime; At the same time, high-level education ensures that citizens have a strong awareness of protecting personal information security and privacy, and are more familiar with ways to report crimes.

2.2.5 Prosecution Rate Analysis

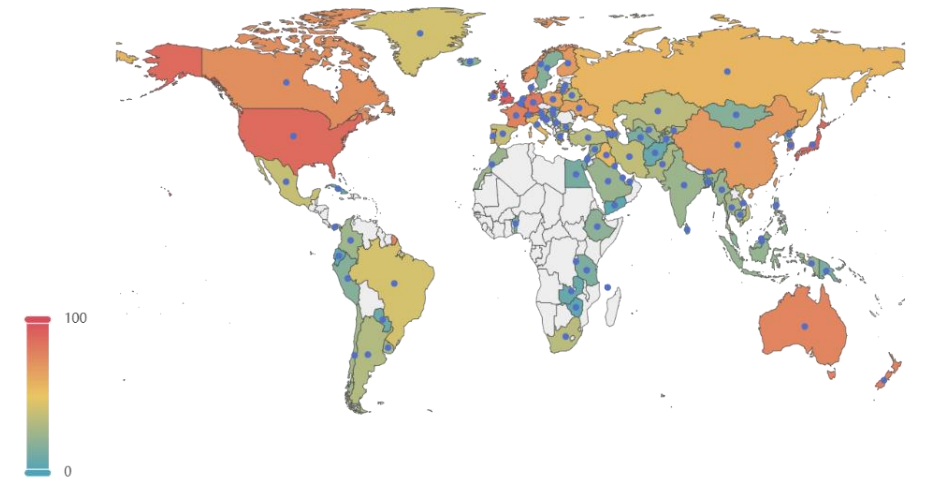

Figure 6. Prosecution Rate Heatmap

As shown in Figure 6, judicial authorities in countries such as the United States, Canada, and the United Kingdom tend to handle reported cases quickly and properly, resulting in a high appeal rate for cybercrime. In Latin America, where the judicial system is underdeveloped, it is more difficult to classify and handle cybercrime cases, resulting in lower appeal rates. Regions with better judicial systems usually have more complete and detailed plans for cybercrime cases, respond faster to cybercrime cases, and can respond to cybercrime cases in a more complete and rapid manner.

2.3. Cybercrime Distribution Feature Extraction Model

Cybercrime is globally diversified by geographical differences [9]. Based on the above information, possible distributional coefficients were obtained for the following variables: government effectiveness, number of Internet users, population size, level of the rule of law, number of secure Internet servers, data on educational indicators and average GDP, and, in order to characterize the correlation between the above variables, the Pearson correlation coefficient was used to obtain the correlation between the distribution of cybercrime and the above factors, using the following formula:

$$\mathrm{r} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \tag{1}$$

Where: r is the correlation coefficient, x_i and y_i represent the value of the i variable x as well as y, respectively, $\bar{x}$ and $\bar{y}$ represent the mean value of the variable x as well as y, respectively, and n is the sample size.

The correlation coefficient r has a range of values of [-1, 1]:

$\mathrm{r} = 1$: indicates that the two variables are completely positively correlated with a linear positive relationship between the values;

$\mathrm{r} = -1$: indicates that the two variables are completely negatively correlated with a linear inverse relationship between the values;

$\mathrm{r} = 0$: indicates that there is no linear correlation between the two variables, but there may be a non-linear relationship.

A correlation test was performed using equation (1), for the Pearson correlation coefficient, using the test statistic p:

$$p = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \tag{2}$$

Where, $p < 0.01$ means that the correlation coefficient is highly significant, $p < 0.05$ means that the correlation coefficient is significant, $p < 0.1$ means that the correlation coefficient has some significant relationship and $p > 0.1$ means that the correlation coefficient is not significant.

2.4. Evaluation Result

This article obtained the calculation results through a programming model, as shown in Table 3:

Table 3. The correlation coefficient table

	distribution coefficient
Government Effectiveness	0.203(0.014***)
Individuals using the Internet	0.089(0.278)
Population	0.326(0.000***)
Rule of Law	0.132(0.109)
Secure Internet servers	0.346(0.000***)
GDL - Educational - index -- data	0.095(0.247)
GDP_Average	0.122(0.138)

In order to better demonstrate the correlation, this article uses a heatmap to visualize the correlation of these coefficients, as shown in Figure 7:

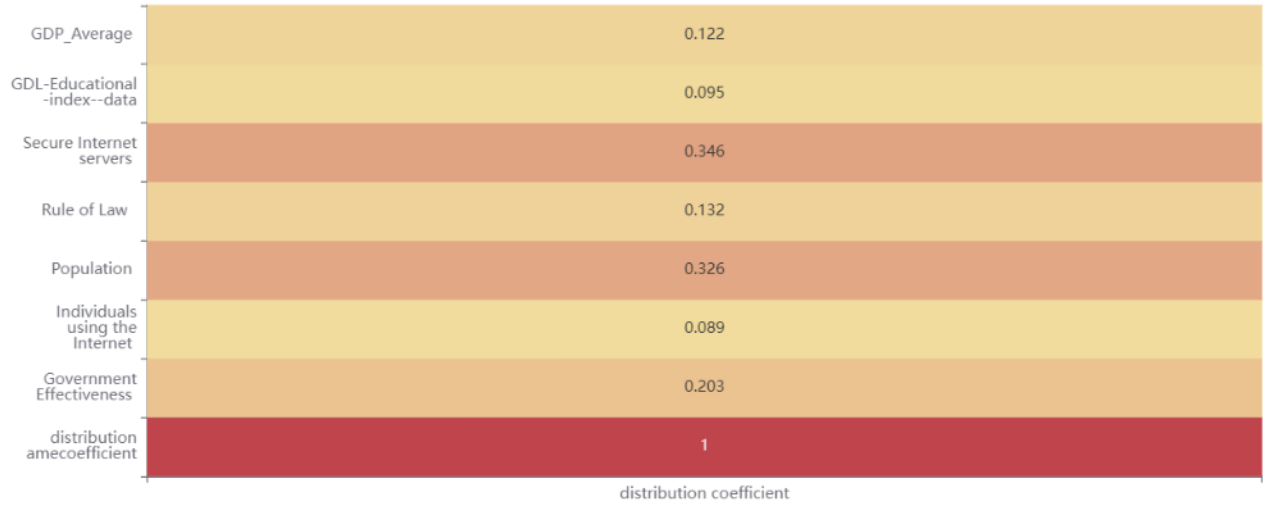

Figure 7. Correlation Coefficient Heatmap

In summary, the distribution of cybercrime is concentrated in countries with larger populations, weaker cyberprotection technologies and infrastructures and weaker judicial systems [10]. The correlation between the coefficient of distribution of cybercrime and other variables (e.g. individuals using the Internet, gdl-education and average gdp) is weak, suggesting that these variables are not directly related to the coefficient of distribution.

3. Conclusion

This study analyzes the distribution patterns of cybercrime by defining four key indicators: the quantitative distribution coefficient, success rate, reporting rate, and prosecution rate. Through Pearson correlation analysis, linear regression, and BP neural network validation, several key findings emerge. First, the quantitative distribution coefficient is significantly correlated with population size, secure internet servers, and government effectiveness, indicating that cybercrime tends to concentrate in countries with large populations, weaker cybersecurity infrastructure, and less efficient judicial systems. For example, regions with rapid digital transformation but immature security systems exhibit higher success rate due to protection loopholes, while developed areas with strong law enforcement and public awareness show lower success rate and higher reporting rate and prosecution rate. The study also confirms that factors like internet user volume, education levels, and average GDP have weak direct correlations with cybercrime distribution, suggesting they are not primary drivers.

The dataset, sourced from platforms like VCDB and ITU, lacks full global coverage, especially for smaller or less transparent regions, which may skew regional analyses. While the BP neural network validates linear results, using more diverse machine learning methods could better capture non-linear relationships. Additionally, the study does not examine the temporal evolution of cybercrime distribution, limiting trend prediction.

Future research should address these gaps by expanding data collection to underrepresented regions and incorporating real-time threat data. Integrating multi-dimensional variables—such as cultural cybersecurity attitudes, regional economic stability, and cross-border legal cooperation—would strengthen the model's explanatory power. Longitudinal studies tracking how policy measures impact the four indicators over time may offer actionable insights.

References

[1] Lehto M. Cyber-attacks against critical infrastructure [M] // Cyber security: Critical infrastructure protection. Cham: Springer International Publishing, 2022: 3-42.

[2] Fereshteh Momeni. The impact of social, cultural, and individual factors on cybercrime. Educational Administration: Theory and Practice, 2024.30 (5), 10152–10159.

[3] Shillair R, Esteve-González P, Dutton W H, et al. Cybersecurity education, awareness raising, and training initiatives: National level evidence-based results, challenges, and promise [J]. Computers & Security, 2022, 119: 102756.

[4] Nguyen T A, Koblandin K, Suleymanova S, et al. Effects of 'Digital' Country's Information Security on Political Stability [J]. Journal of Cyber Security and Mobility, 2022: 29-52.

[5] Folorunso A, Wada I, Samuel B, et al. Security compliance and its implication for cybersecurity [J]. World Journal of Advanced Research and Reviews, 2024, 24 (01): 2105-2121.

[6] Schober P, Boer C, Schwarte L A. Correlation coefficients: appropriate use and interpretation [J]. Anesthesia & analgesia, 2018, 126 (5): 1763-1768.

[7] Priya K S. Linear regression algorithm in machine learning through MATLAB [J]. Int J Res Appl Sci Eng Technol, 2021, 9 (12): 989-995.

[8] Song S, Xiong X, Wu X, et al. Modeling the SOFC by BP neural network algorithm [J]. International Journal of Hydrogen Energy, 2021, 46 (38): 20065-20077.

[9] Chen S, Hao M, Ding F, et al. Exploring the global geography of cybercrime and its driving forces [J]. Humanities and Social Sciences Communications, 2023, 10 (1): 1-10.

[10] Mishra A, Alzoubi Y I, Anwar M J, et al. Attributes impacting cybersecurity policy development: An evidence from seven nations [J]. Computers & Security, 2022, 120: 102820.