Cybercrime And Policy Insights Based on Time-Series Forecasting and Multiple Regression

Xusheng Zhao [1, *], Xinyi Wang [2], Jinhong Qi [3]

[1] School of Science, Jiamusi University, Jiamusi, China, 154007

[2] School of Materials Science and Engineering, Jiamusi University, Jiamusi, China, 154007

[3] School of Mechanical Engineering, Jiamusi University, Jiamusi, China, 154007

* Corresponding Author Email: m19545392425@163.com

Abstract. With the rapid development of global information technology and the internet, cybercrime has become increasingly rampant, posing severe threats to the economy, society, and national security of various countries. This study aims to construct mathematical models to analyze the global distribution and influencing factors of cybercrime and evaluate the effectiveness of national cybersecurity policies. We utilized data visualization techniques to illustrate the frequency of cybercrime and its economic impact across countries. Additionally, a time-series forecasting model predicts a steady rise in cybercrime frequency over the next three years(2023-2025), underscoring the importance of continuously enhancing cybersecurity policies. Moreover, a decision tree model identified internet penetration and per capita GDP as the most crucial statistical variables in explaining variations in cybercrime rates.

Keywords: Cybercrime, Decision tree, Time-series forecasting model, Neural Network.

1. Introduction

In the Internet era, IT innovation drives network society development but also raises cybercrime risks, threatening national cybersecurity. So, countries introduce relevant policies, and the International Telecommunication Union focuses on global network security [1]. A systematic network security policy system has been initially formed, guaranteeing network security governance. Security and development are intertwined: cybersecurity capabilities matter in tech competition, while infrastructure and economic-tech development underpin them. The current policy system can prevent cybercrime, and there's an ambition to develop stronger national cybersecurity policies [2].

Wang et al. noted that cybersecurity capability, as a key part of digital competitiveness, is crucial for economic and technological development [3]. Zhang[4] compared domestic and international cyber security development, stating that China has advantages in cyber and digital development overall, yet its security capabilities can be improved. They analyzed major countries' successful practices in cybersecurity construction like strategic planning, technological innovation, etc. Based on China's situation, targeted suggestions were put forward, such as improving the organizational system, developing the technology industry, to accelerate the modernization of cybersecurity [5] system and capacity and support the goal of a strong digital country.

This paper selected appropriate indicators, collect relevant data, establish multiple linear regression model, and quantitatively analyze the relationship between cybercrime and national characteristics. Then we established a mathematical model showing the influence of national cybersecurity policies on solving cybercrimes. Correlation analysis and multiple regression analysis are used to explore the relationship between demographic characteristics and cybercrime, to reveal the main influencing factors of cybercrime, and to provide data support for the formulation of targeted cybersecurity policies.

2. Preliminary

2.1. Assumption

1.The frequency of cybercrime is positively correlated with the rate of Internet penetration.
2.GDP per capita is positively correlated with economic losses from cybercrime.
3.Education level is positively related to cybercrime frequency.
4.Cybersecurity investment is negatively related to cybercrime success rates.
5.Prosecution rates are positively correlated with cybercrime reporting rates.
6.There is a negative correlation between the overall policy score and the frequency of cybercrime.

2.2. Network security

It means that the hardware and software of the network system and the data in its system are protected from damage, alteration and leakage due to accidental or malicious reasons, and that the system operates continuously, reliably and normally, and the network service is not interrupted [6].

2.3. Cybercrime

A general term for the use of computer technology by the perpetrator, with the help of a network, to attack his system or information, to destroy it or to use the network to commit other crimes [7].

2.4. Notations

The key mathematical notations used in this paper are listed in Table 1.

Table.1. Notations

Symbol	Description
n	Sample size
x	Observed values of the first variable
y	Observed values of the second variable
t	T-statistic
r	Pearson correlation coefficient
sum xy	The sum of the products of all x and y
sum x	Sum of the first variable
sum y	Sum of the second variable
sum x^2	The sum of the squares of the first variable
sum y^2	The sum of the squares of the second variable

3. Model Preparation

3.1. Data Pre-processing

For the missing data in the network security investment/dollar data, Newton interpolation method (formula) was used to process it and the filling value of $5545771429 was obtained to prepare for the subsequent modeling and application. When we looked at the key data, the amount of cybersecurity investment (in US dollars), we found some missing values.

These missing data can adversely affect subsequent accurate data analysis, modeling, and model-based practical applications, potentially resulting in biased analysis results and inaccurate model predictions. In order to properly deal with these missing data, after careful consideration and technical evaluation, we decided to use Newton interpolation method for data filling. Newton interpolation is a kind of numerical analysis method which approximates the unknown data by constructing polynomials based on the known data points [8]. Its core formula is

$$NN (x) = f (x0) + f[x0 ,x1 \quad](x - x0) + f[x0 , x1, x2](x - x0)(x - x1) + ... + f[x] \tag{1}$$

3.2. The global distribution of cybercrime

We can see that the cybercrime rate varies significantly from country to country, with developed countries such as the United States and Germany having higher cybercrime frequencies of 65% and 55%, respectively, which may be related to their large Internet user bases and levels of technological development. On the other hand, some developing countries such as Vietnam and the Philippines have lower cybercrime rates of 13% and 2% respectively, which may be related to their lower Internet penetration rate and insufficient awareness of cybersecurity [9]. Overall, the frequency of cybercrime is closely related to a country's Internet penetration rate, cybersecurity measures, enforcement of laws and regulations, and the public's awareness of cybersecurity, so countries should strengthen cybersecurity education, improve laws and regulations, and enhance their technical protection capabilities to cope with the growing threat of cybercrime.

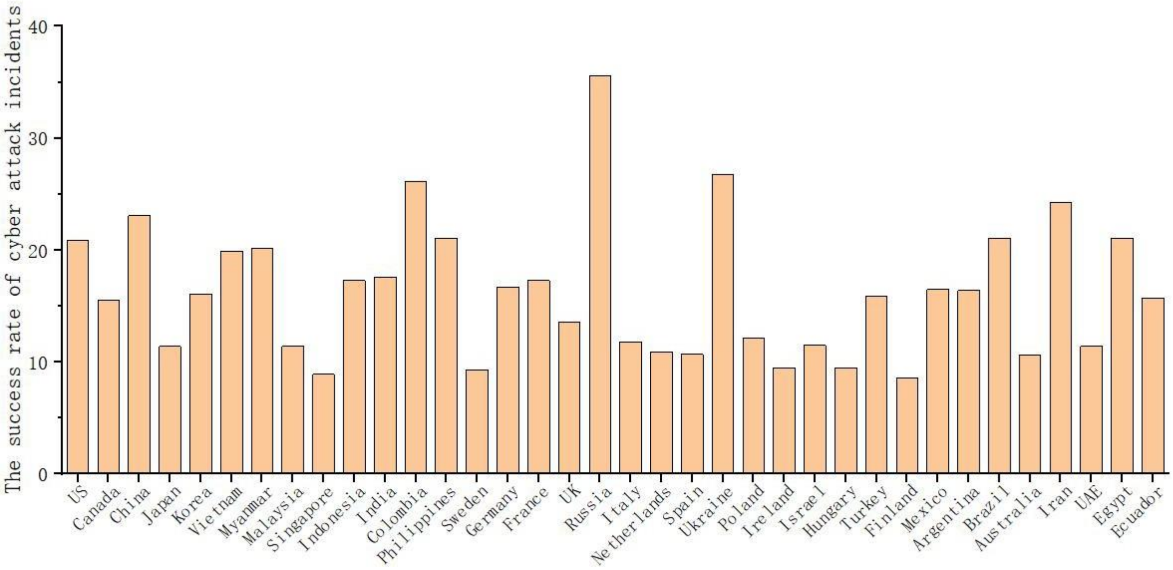

Figure 1. Success and failure

Figure 1 shows the success rate of cyberattack incidents in various countries, with the vertical axis for the success rate percentage and the horizontal axis listing countries. The bar height represents the success rate in each country.

There are marked differences among countries. For instance, Russia and Egypt have high - success - rate bars, likely due to weak cybersecurity defenses. In contrast, the US and Germany have low success rates, indicating effective defenses.

Overall, the chart reflects the uneven global cybersecurity level. Some countries should strengthen defenses, especially for critical infrastructure and sensitive data. Those with low success rates offer lessons in cybersecurity technology and policies. It serves as an important basis for analyzing the effectiveness of countries' cybersecurity strategies. It is shown in Figure 2.

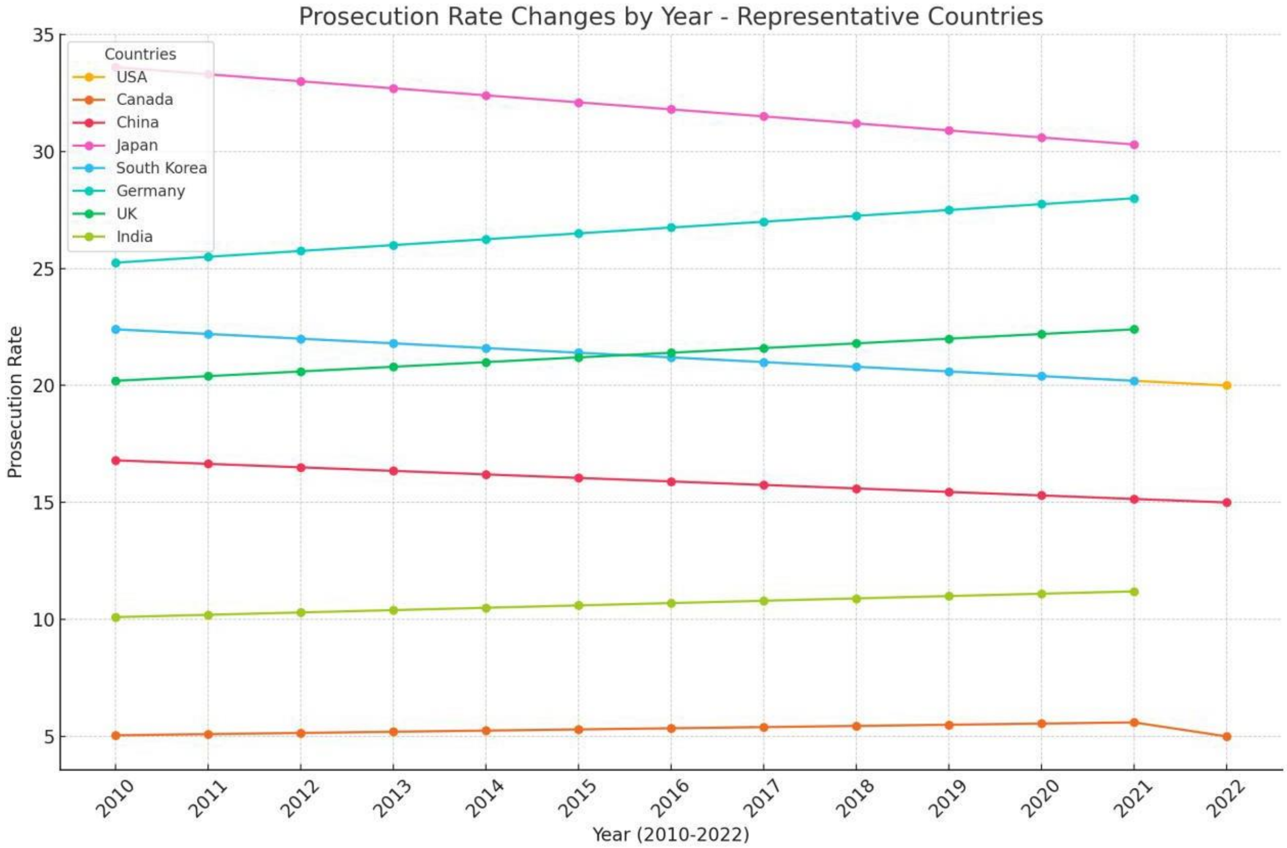

Figure 2. Success and failure

The trends in cybercrime prosecution rates from 2010 to 2022 show significant differences among representative countries. Japan had the highest prosecution rate, with a slight annual decline but still higher than others, indicating strong law enforcement. Conversely, the United States and Canada had lower rates that further decreased, possibly due to increased difficulty in tracking offenders or changes in law enforcement priorities. China and Germany maintained relatively stable rates with a slight increase, showing stable enforcement capacity. India had the lowest rate but saw yearly increases, suggesting gradual improvement in its cybersecurity enforcement system [10]. Overall, these trends are related to countries' cybersecurity policies, technological development, and legal enforcement strength, providing a basis for studying governance strengths and weaknesses and highlighting the need for improvement in low-prosecution-rate countries.

3.3. Correlation analysis

The correlation heat map shows relationships among variables in cybersecurity. The number of victims is highly correlated with the total and successful attack numbers, meaning more attacks result in more victims. Economic loss has a weak positive correlation with victims and attacks. Cybersecurity investment is negatively correlated with the number of victims, indicating that more investment reduces victim numbers. The cybersecurity index is negatively correlated with the attack success rate, showing higher security levels lower success rates. Internet penetration benefits cybersecurity, while population size has little correlation with other variables. Overall, the heat map indicates that increasing cybersecurity investment and the security index are effective ways to lessen the impact of cyber-attacks. It is shown in Figure 3.

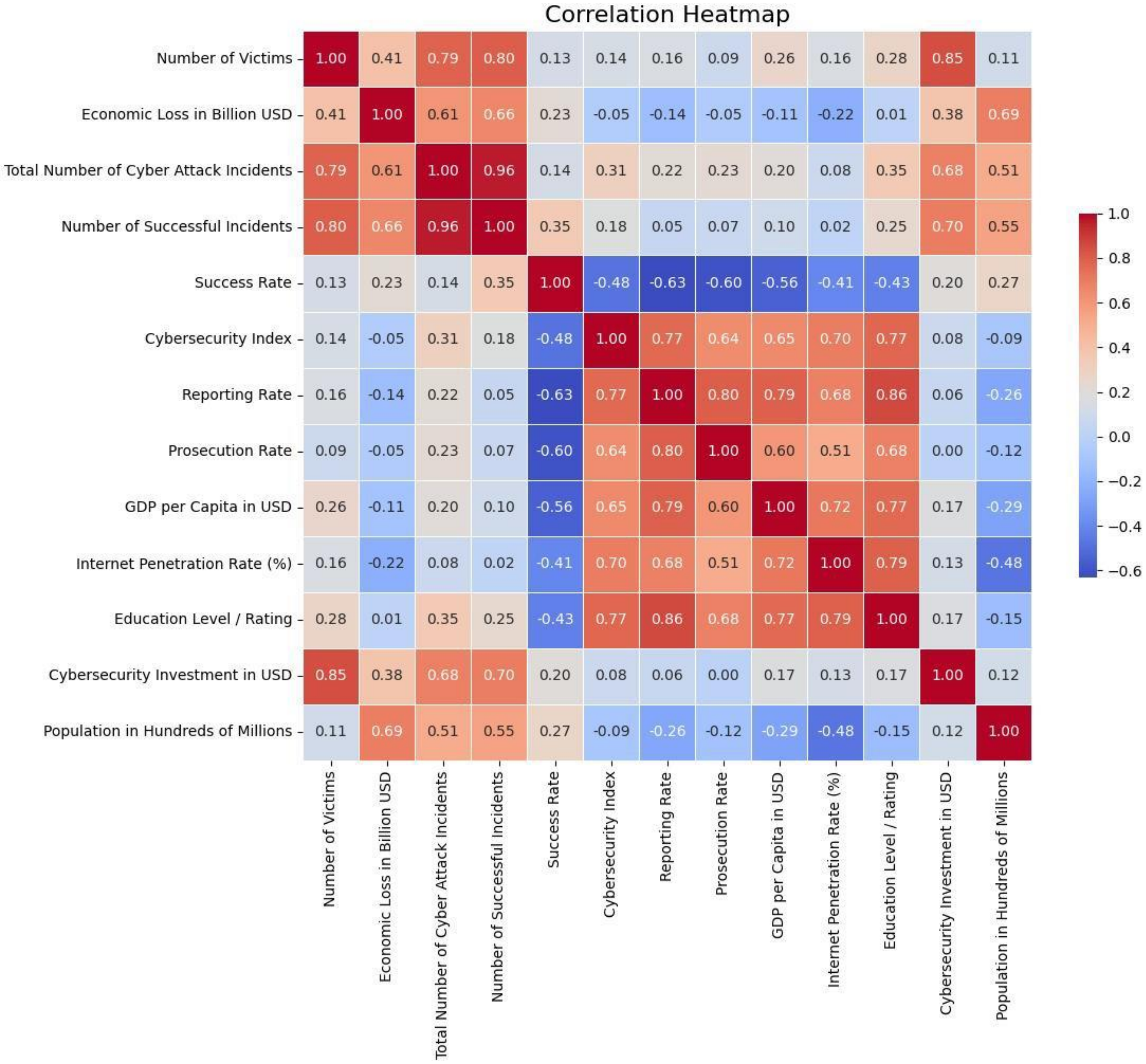

Figure 3. Correlation heat map

4. Integrated assessment of policy effectiveness

Entropy Weight Method) is used to determine the weight of each evaluation indicator. By calculating the information entropy of each indicator, it reflects the discrete degree of the indicator, the larger the entropy value, the smaller the amount of information, the lower the weight, and vice versa.

TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) a multi-indicator decision analysis method, which comprehensively evaluates the advantages and disadvantages of the program by calculating the distance between each program and the ideal solution and negative ideal solution.

Based on the needs of the problem and the available data, we chose law and technology as the first-level indicators for assessing the effectiveness of the policy, prosecution rate, reporting rate and cybercrime success rate as the second-level indicators for law, and Internet penetration, the degree of sophistication of technological protection measures and the level of cyberinfrastructure as the second-level indicators for technology.

Figure 4 shows the detail of indexes.

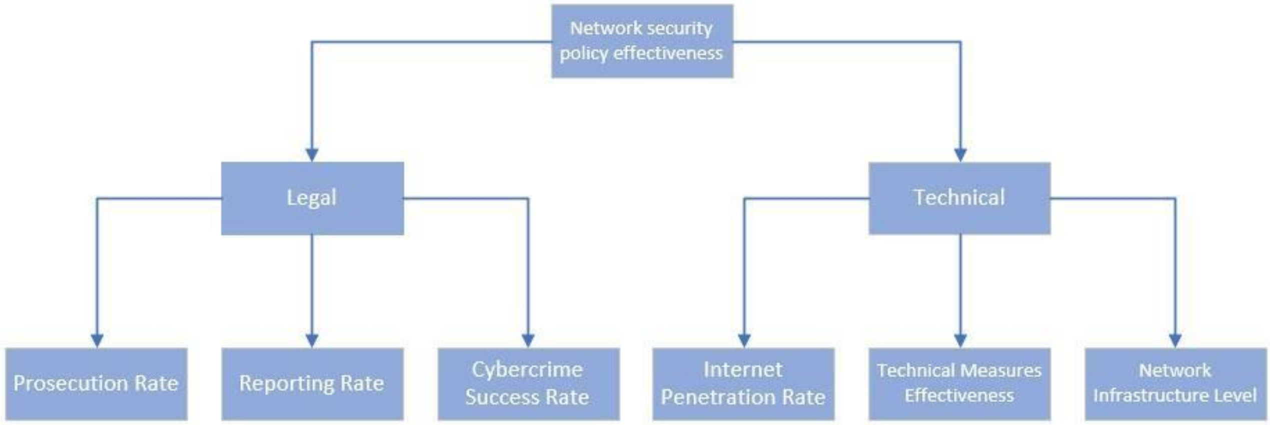
Network security policy effectiveness
Legal
Technical
Prosecution Rate
Reporting Rate
Cybercrime Success Rate
Internet Penetration Rate
Technical Measures Effectiveness
Network Infrastructure Level

Figure 4. Hierarchy of primary and secondary indicators

4.1. Policy effectiveness results analysis

Due to the large amount of data, we obtained the first level indicator weights by python program, the weight of the indicator for law is 0.8501 and the weight of the indicator for technology is 0.1499. The second level indicator weights are shown in the table 2.

Table.2. Weighting of secondary indicators

norm	weights
prosecution rate	0.471737
reporting rate	0.207993
success rate	0.170335
Internet penetration	0.035668
Network Security Index	0.094956
Level of network facilities	0.019311

By calculating the composite score and ranking the policy effectiveness based on the proximity of the score to 1, the results were obtained as shown in Figure 5.

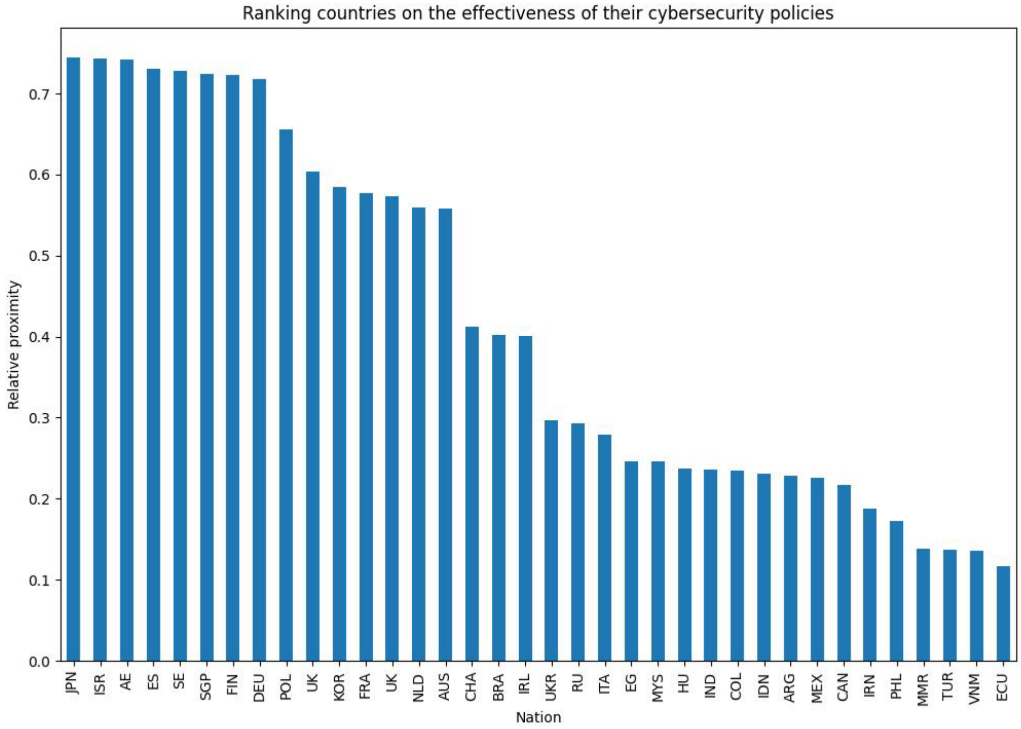

Figure 5. Ranking of countries on the effectiveness of their cybersecurity policies

This chart shows how countries are ranked in terms of the effectiveness of their cybersecurity policies, based on relative proximity (*C_i*). Top-ranked countries, such as the United States, Sweden, Germany, and Japan, are considered to have the most effective cybersecurity policies, reflecting their significant strengths in legal and technical measures. Moderately ranked countries such as Finland, the United Kingdom and Poland show better policy implementation but still have room for further improvement. Lower-ranked countries. such as China, Russia and the Philippines have less effective cybersecurity policies and may need to invest more resources in upgrading cyber protection, strengthening legal enforcement and improving technical infrastructure. Overall, there are significant differences in the effectiveness of cybersecurity policy implementation among different countries, underscoring the importance of strengthening international cooperation and optimizing the allocation of resources to enhance global cybersecurity.

4.2. Time series prediction

Based on the characteristics of the data we choose ARIMA (Autoregressive Integral Sliding Average Model) for time series forecasting because of its applicability to time series data with linear trend and no seasonality. We select a suitable time series model, such as ARIMA, through a python program and use automated tools (e.g., auto_arima from pmdarima) to determine the optimal parameters (*p, d, q*) of the model. In the model training phase, the data are divided into a training set (80%) and a test set (20%), the model is fitted through the training set and the model performance is evaluated using the test set, which in this case uses the mean square error (MSE) for metrics evaluation. After confirming that the model performs well, the trained model is used to forecast policy effectiveness from 2023 to 2025. Finally, the prediction results are combined with historical data to optimize the decision support by visualizing the historical trend and future development of policy effectiveness in each country, as seen in the figure 6.

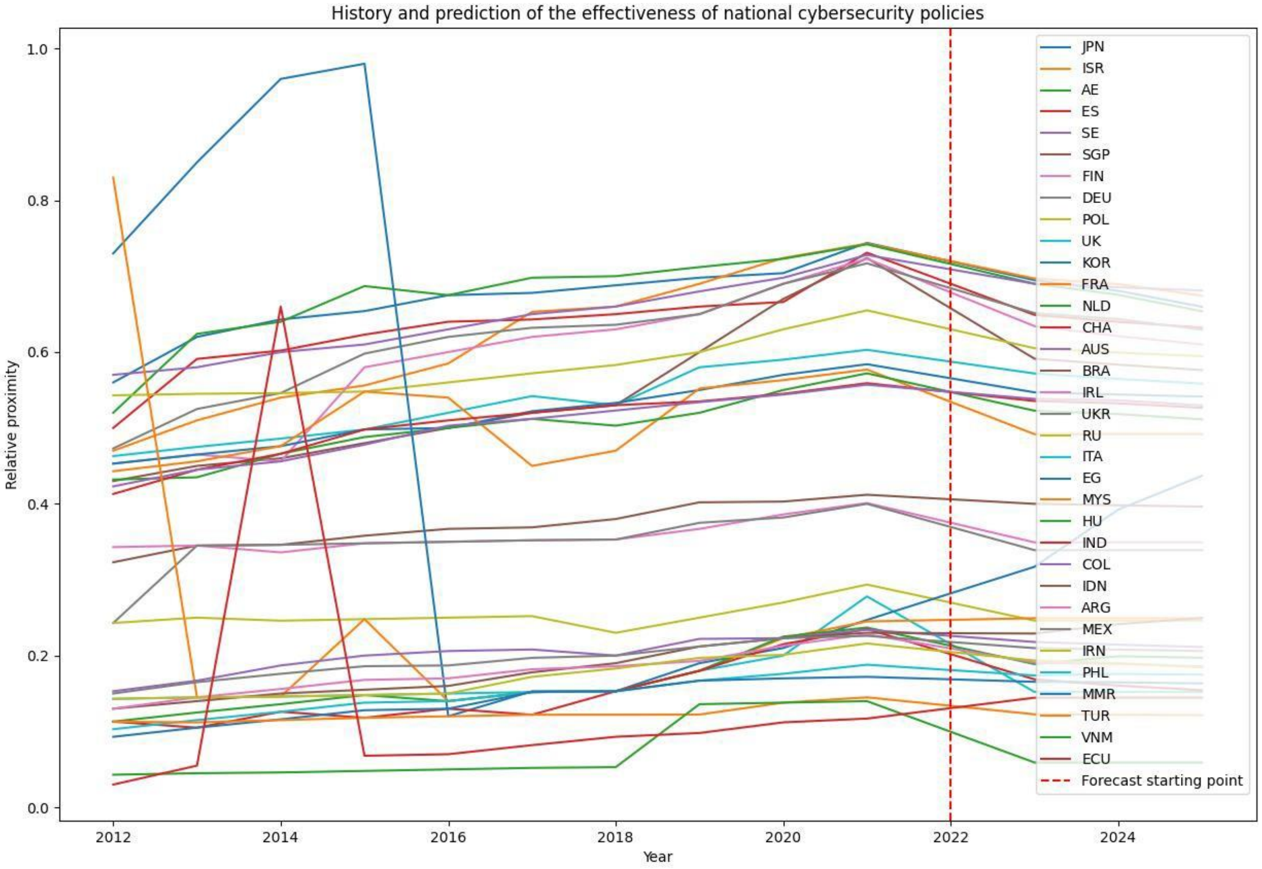

Figure 6. Analysis of time series model results

This graph shows the historical and projected effectiveness of national cybersecurity policies in different countries between 2012 and 2024. The vertical axis represents relative effectiveness, ranging from 0 to 1. It can be observed that certain countries (e.g., Japan, Israel, and Finland) have maintained a high level of cybersecurity policy effectiveness over this time period and have shown an upward trend. While some other countries (e.g., Brazil and India) show lower effectiveness with higher fluctuations. The red dotted line in the figure marks the starting point of the projections, indicating projected data from 2022 onwards. Overall, while some countries have made significant progress on cybersecurity policies, many still face challenges, demonstrating the variability and complexity of global cybersecurity policies.

4.3. Sensitivity analysis results

The effect of maximum depth on model accuracy has been plotted as follows: as depth increases, model accuracy first improves and then stabilizes. The model performs best at depth 4 (~89% accuracy). Further increases in depth may lead to overfitting. It is shown that the grouping criteria have a large impact on the classification results of the model, with the default median grouping working best.

The maximum depth of the decision tree determines the complexity that the model is able to learn and directly affects the model's fitting ability. A smaller depth may result in an underfitting model that is unable to capture complex patterns in the data, while a larger depth may result in an overfitting model that is too sensitive to the details of the training data and difficult to generalize to the test data. Therefore, it is important to gradually vary the maximum depth and observe the predictive accuracy of the model in the sensitivity analysis.

Through experimentation we have found that:

1. When the depth is small (e.g.,2 or 3), the model is less accurate, indicating underfitting. At this point, the decision tree is unable to capture enough data features, showing a lack of ability to classify complex relationships.

2. When the depth gradually increases (e.g., depth of 4), the model accuracy reaches the highest value, showing that the model complexity and data characteristics have reached a good balance at this point. This stage is the optimal complexity of the model, which is able to fit the data effectively and also has a good generalization ability.

3. When the depth continues to increase (e.g., when it exceeds 6), the accuracy begins to fluctuate or decline, indicating model overfitting and a decrease in generalization ability due to overlearning of training data noise or extreme values.

By looking at the performance at different depths, we were able to determine an optimal depth, which not only helped us optimize the model performance, but also increased confidence in the stability of the model. If the model performance is stable over a range of depths, this indicates that the model is more robust to parameter selection and the results are reliable.

5. Conclusions

This paper provides data-driven decision support for policymakers, recommending that alongside improving economic conditions, internet accessibility, and education, countries should optimize legal and technical measures to establish a robust cybersecurity framework. Our findings advocate for a holistic approach to cybersecurity policy-making, emphasizing the need to balance economic development, internet accessibility, and education with the optimization of legal and technical measures.

References

[1] ZHANG Fanrui,SHI Tuo,YOU Hui. Research on spatial statistical distribution and risk characteristics of telecommunication network fraud victim group-taking brush order rebate category as an example[J/OL]. Research World,1-9[2025-01-28].

[2] Hu Jiang,Wang Weijian. Cybercrime motivation: from moral backlash to technological neutrality[J]. Journal of Jiangsu Police College,2023,38(04):88-97.

[3] HUANG XIAO TENG, JIN LIANGPING, WU BIN, et al. Research on the prevention mechanism of college students' cybercrime and ideological education[J]. Legal Expo,2023, (12):46-48.

[4] Wang Shiqi. Reflection and regulation on the expansion application of the crime of helping information network criminal activities[D]. Jiangxi University of Finance and Economics, 2023.

[5] Milon M N U, Ghose P, Pinky T C, et al.An in-depth PRISMA based review of cybercrime in a developing economy: Examining sector-wide impacts, legal frameworks, and emerging trends in the digital era[J]. 2024.

[6] Cretu-Adatte C, Zbinden R, Brunoni L, et al.How do Ivorian Cyberfraudsters Manage Their Criminal Proceeds?[J].European Journal on Criminal Policy & Research, 2024, 30(3).

[7] Lai C M, Guo Y H. Tracking of Disinformation Sources: Examining Pages and URLs[J]. IEEE transactions on computational social systems, 2024(5):11.

[8] Faye S, Abdulrahman J, Talb R A, et al.Cybersecurity in Aviation: A Case-Based Approach to Preparedness[J].International Journal of Information Security & Cybercrime, 2024, 13(2).

[9] Al-Dahasi E M, Alsheikh R K, Khan F A, et al.Optimizing fraud detection in financial transactions with machine learning and imbalance mitigation[J].Expert Systems, 2025, 42(2).

[10] Almutawa A, Ikuesan R A, Said H. Towards a Comprehensive Metaverse Forensic Framework Based on Technology Task Fit Model[J]. 2024.