Research on network security situation evaluation based on multivariate linear regression model

Yutong Guo [1, *], Boyuan Fu [2], Yukun Zhang [2]

[1] The School of Mechanical Engineering and Automation, Dalian Polytechnic University, Dalian, China

[2] School of Information Science and Engineering, Dalian Polytechnic University, Dalian, China

* Corresponding Author Email: 19818922108@163.com

Abstract. Technology development facilitates life, but frequent cybersecurity incidents have caused complex jurisdiction across borders. In order to solve the problem of policy effectiveness and the distribution of cybercrime, the comparison of national policies and crimes and the impact of demographic characteristics on the distribution of crimes is analyzed. This paper adopts multiple linear regression, double differential method (DiD) and stepwise regression analysis model, and combines ITU crime data for comparison and analysis; regression analysis and its visualization methods are used, including scatter plots, contrast line plots, etc., to explore demographic data. With cybercrime impact and successfully predict future crime events. The study found that national policy differences significantly affect the incidence and success rates of cybercrime. Effective implementation of policies can curb crime and improve judicial efficiency. In addition, increased Internet access rates, uneven wealth and low education levels have all aggravated cybercrime. Improving education and cybersecurity awareness is crucial to preventing and responding to cybercrime. To sum up, this study provides valuable reference for governments to formulate and optimize cybersecurity policies, and also provides solid theoretical basis and practical guidance for preventing and responding to cybercrime.

Keywords: network security, comparative analysis, factor, Multivariate regression analysis, policy, optimization.

1. Introduction

The rapid development of modern technology has made the world more interconnected, increased productivity, but also increased the risk of cybercrime [1]. To respond, countries have formulated cybersecurity [2] policies and strengthened international cooperation, but cybercrime is complex and transnational, and the challenges are arduous. Therefore, studying the distribution of cybercrime, policy effectiveness and its relationship with population characteristics is crucial to optimizing policy laws.

In the field of machine learning, the most popular supervised learning techniques are called classification and regression methods. Recently, Sarker et al. proposed BehavDT and IntruDtree [3] classification techniques, which can effectively build data-driven prediction models. Furthermore, reinforcement technology is another type of machine learning that creates its own learning experience by interacting directly with the environment, an environment-driven approach, where the environment is often expressed as a Markov decision-making process and made based on the reward function decision making [4]. Monte Carlo learning [5], Q-learning, and Deep Q Networks are the most common reinforcement learning algorithms [6]. For example, in a recent work, the authors proposed a method to detect botnet traffic or malicious network activity using reinforcement learning in combination with neural network classifiers [7].

There are also some shortcomings in the research in the field of cybersecurity data science. The policy rules used in most cybersecurity systems are static, generated by human expertise or based on ontology [8]. Although the association rule learning technique generates rules from data, there are redundant generation problems [9], which complicates the policy rule set. Moreover, most commercial products in the field of cybersecurity include signature-based intrusion detection

technology [10]. However, lack of functionality or inadequate analysis may cause these technologies to miss unknown attacks.

2. Model building

2.1. Regressive Analysis Model

Multivariate linear regression model is a statistical method used to establish a linear relationship between a dependent variable and multiple independent variables. In this model, the dependent variable (response variable) is predicted as a linear combination of multiple independent variables (explanatory variables), each independent variable has a corresponding regression coefficient that represents the independent variable versus the dependent variable degree of impact. This model is used for prediction and decision-making.

The working principle of the regression analysis model is mainly based on mathematical statistical methods, which are used to study the relationship between the dependent variable (response variable) and the independent variable (explanatory variable). It tries to find the mathematical relationship between the dependent variable and the independent variable, which is a relationship. It is usually expressed as a regression equation. This equation describes how the dependent variable changes with the independent variable and can be used to predict the value of the dependent variable or to explain the causal relationship between the dependent variable and the independent variable.

Then we do the regression analysis and we use multiple linear regression assuming that there is a linear relationship between the dependent variable Y and the independent variables X_1, X_2, X_3, X_4, X_5 ,and the model can be expressed as formula (1).

$$Y = \beta_0 + \beta_{12} * X_1 + \beta_{22} * X_2 + \beta_{32} * X_3 + \beta_{42} * (X_1 * X_2 + \beta_{52} * (X_1 * X_3) + \varepsilon_2) \ldots\ldots \quad (1)$$

In the formula (1), β_1, β_2, β_3, β_4 is the parameter to be estimated and ε is the error term.

$$Q = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 * x_{i1} \ldots\ldots\ldots - \beta_k * x_{ik})^2 \quad (2)$$

$\beta_1, \beta_2, \beta_3, \beta_4$ The least square method is used to calculate the parameters. The goal of the least square method is to find a set of parameter values that minimizes the sum of squares of error between the observed value y and the model predicted value y_i. The specific formula is formula (2), which minimizes Q_e to obtain the estimated value, such as formula (3) (4) (5).

$$U = \sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \quad (3)$$

$$Q_e = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (4)$$

$$r = \sqrt{\frac{U}{L_{yy}}} = \sqrt{\frac{U}{U+Q_e}} \quad (5)$$

2.2. Regression analysis of dual-difference method (DiD)

The dual-difference method is a quasi-experimental design that compares the variations between the treatment group (groups affected by policy or intervention) and the control group (groups not affected by policy or intervention) before and after policy or intervention implementation. This model is often used to evaluate the impact of a policy on the target group, to evaluate the effectiveness of an implementation of a project or plan, or to evaluate the impact of a social phenomenon or event on an individual or group.

The working principle of the dual-differential model (DID) is mainly based on a counterfactual framework, which is centrally lies in comparing the treatment group (the group receiving the

intervention) and the control group (the group not receiving the intervention) before and after the intervention. to estimate the net effect of intervention. This approach eliminates the impact of temporal trends and individual characteristics on outcomes, thereby more accurately assessing the effectiveness of policies or treatments.

The mathematical formula of regression analysis of the dual-differential method (DiD) is a key tool commonly used in policy evaluation, and its basic form can be expressed as formula (6):

$$Y_{it} = \alpha + \beta Treat_i + \gamma Post_t + \delta * (Treat_i * Post_t) + \varepsilon_{it} \quad (6)$$

Where Y_{it} is the result variable of individual i at time t, and $Treat_i$ is a dummy variable, and 1 for individuals in the treatment group and 0 for individuals in the control group , and $Post_t$ is a dummy variable, which is 1 for the time period after the policy is implemented and 0 for the time period before the policy is implemented, and $Treat_i * Post_t$ is an interaction term that indicates the effect of the processing group after policy implementation, and α is a constant term, and β indicates the difference between the treatment group and the control group before policy implementation, and γ represents the time effect, that is, the changes of all individuals over time are not considered when the processing effect is not considered, and δ is the coefficient of most interest, which represents the policy effect, i.e. the additional changes in the treatment group relative to the control group after policy implementation, and ε_{it} is the error term.

2.3. Multivariate linear regression model

Cyber crimes are a global issue, and it is necessary to understand its mechanism in depth. This study adopts a multi -linear regression model, combined with the interaction effect, and quantitatively analyzes the impact of Internet access rate, level of wealth and education level on the number of cyber crimes. Multi -linear regression model quantitative multi -independent variables and due to variables, the population statistics data is used as an independent variable, and the number of cyber criminal events is used as the cause variable. Including interactive effects can improve the accuracy of forecasting, comprehensively capture the complex relationship between factors, and provide more accurate predictions.

The working principle of a multivariate linear regression model is based on statistics and is used to establish a linear relationship between multiple independent variables and dependent variables. It tries to predict a cause through multiple independent variables (also known as explanatory variables or predictors). The value of a variable (also known as a response variable or result variable). The relationship between these independent variables and the dependent variable is assumed to be a linear relationship, i.e. the change of the dependent variable can be expressed as a linear combination of independent variables plus an error term.

The model is set up as follows:

Let Y be the rate of cybercrime (or success rate, reporting rate, prosecution rate, etc., depending on the analysis target), X1, X2,... Xn is the n factors that affect cybercrime (such as education index, GDP, human development index, various dimensions of GCI index, different dimensions of cybersecurity policy, etc.). Then the multiple linear regression model can be expressed as formula (7):

$$Y = \beta_0 + \beta_{12}X_1 + \beta_{22}X_2 + \beta_{32}X_3 + \beta_{42}(X_1 \, x \, X_2) + \beta_{52}(X_1 \, x \, X_3) + \varepsilon^2 \quad (7)$$

Where Y is dependent variable, which represents indicators related to cybercrime (such as incidence), and X_0,X_1,...,X_n is the intercept term represents the expected value of the dependent variable when all independent variables are 0, and β_0 is the interception item indicates that when all the independent variables are 0, the expected value of the variable, and $\beta_1, \beta_2, ..., \beta_n$ is the regression coefficient indicates the degree of influence of each variable on the variable, and ε is error items indicate part of the part that cannot be explained in the model.

3. Results and analysis

3.1. Establishing a Data framework

We set up a data framework, we set up a framework with country name, cybercrime incidents, cybercrime success rate, cybercrime prosecution rate, cybercrime reporting rate, cybersecurity policy release year, and then we do data visualization. We use a scatter plot to do data visualization, showing cybercrime incidents, cybercrime success rate, cybercrime prosecution rate, and so on. The relationship between the reporting rate of cyber crimes and the release year of cyber security policies. Let's take the United States as an example and assume that the United States released its cyber security policy in 2000 while China released its cyber security policy in 2010. We will link the number of cyber crimes in each country with the release time of the cyber security policies and establish a table.

Data collection is a key starting point when studying the relationship between cybercrime and cybersecurity policy. We carefully collected important data from the International Telecommunication Union (ITU) on the number of secure Internet servers in each country and the number of people using the network in each country. These data provide a solid foundation for further analysis of the situation of cybercrime. Then, the collected cybercrime data of each country is closely integrated with the cybersecurity policies implemented by that country in order to explore the internal links between the two in a comprehensive and multi-dimensional way.

3.2. Data Processing

First, the raw data is carefully processed. Comprehensively check the integrity of the data, and adopt statistical methods such as mean filling and median filling to properly deal with the possible missing values to ensure the accuracy and availability of the data. At the same time, the abnormal values are checked and identified by scientific means such as box plot, and these abnormal data are processed reasonably to avoid interference to the subsequent analysis results. In addition, in order to unify the scales of different variables and facilitate subsequent calculation and analysis, Z-score standardization and other methods are used to standardize or normalize the data to make the data more comparable.

After completing the data processing, proceed to establish the data framework. Construct a framework that covers key information such as country name, number of cybercrime incidents, cybercrime success rate, cybercrime prosecution rate, cybercrime reporting rate, and year of cybersecurity policy release. This frame diagram can systematically and intuitively present the logical relationship between various data, and provide a clear idea and structure for the subsequent analysis.

3.3. Data Visualization

Then, the data is visualized. We choose scatter plot as the main visualization tool to clearly show the relationship between cybercrime incidents, cybercrime success rate, cybercrime prosecution rate, cybercrime reporting rate and the year of cybersecurity policy release. Take the United States and China as an example, suppose that the United States released its network security policy in 2000, while China released its network security policy in 2010, the number of cyber crimes in each country is related to the time of the release of network security policy, and a detailed Table I is established. In this way, the potential relationship between the time of policy release and the number of cybercrimes can be more directly observed, providing intuitive data support for further research and analysis.

Table 1. Data framework

Country	Cybercrime Incidents	Success Rate	Report Rate	Prosecution Rate	Policy Year
United States	10000	60	90	80	2000
China	9000	80	75	75	2010
India	14000	75	80	60	2015
Germany	7000	65	95	80	2005
Brazil	9000	70	85	65	2012

3.4. Time sequence analysis

After the establishment, we analyzed the relationship between the policy release time and cybercrime by using the visualization technology scatter plot. First, we compared the cybercrime incidents, the success rate of cybercrime, the prosecution rate of cybercrime, the reporting rate of cybercrime and the year of the release of the cybersecurity policy in various countries by using time series analysis, and then visualized it so that the changing trend could be seen more clearly.

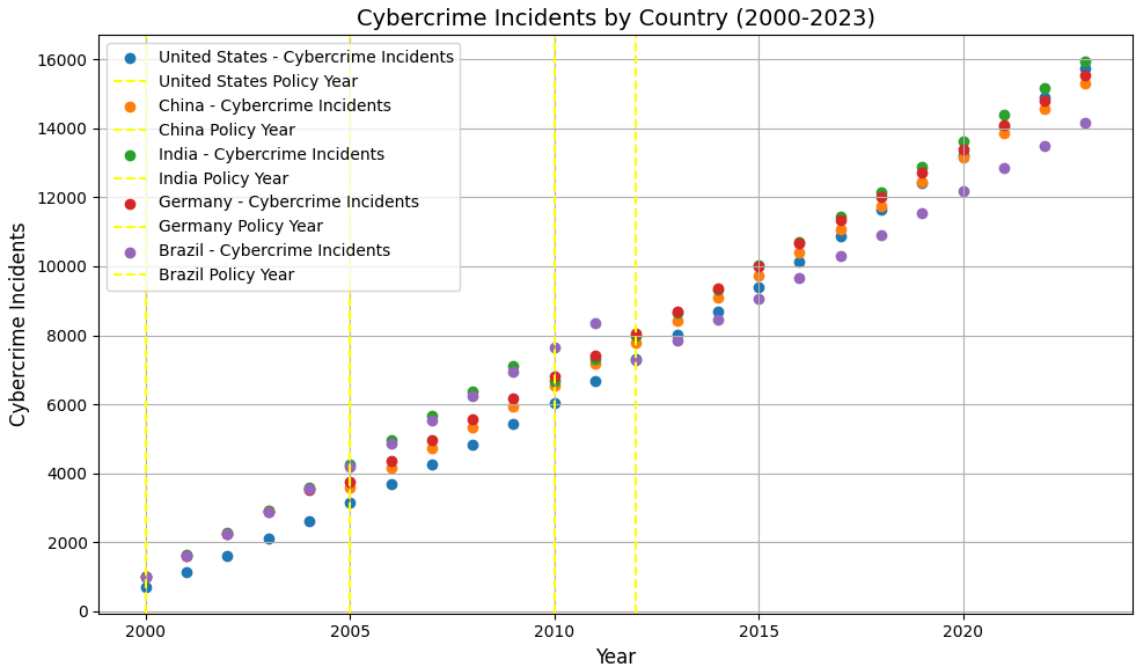

Figure 1. Scatter plot of the number of cybercrimes in each country

As can be seen from Figure 1, this is a scatter plot showing the number of cybercrime incidents in the United States, China, India, Germany, and Brazil from 2000 to 2023. The colored dots represent the number of cybercrime incidents in different countries: blue for the United States, orange for China, green for India, red for Germany, and purple for Brazil.

The dashed yellow line shows the year each country introduced its cyber policy. As can be seen from the figure, India, Brazil and other countries are greatly affected by policies, but it is difficult for some countries to directly judge the clear impact of policies on the number of cyber crime incidents from the figure, which may be because it takes some time for the effect of policies to appear, or the situation of cyber crime is complex, and policy is only one of the influencing factors.

Overall, the chart reflects the development of cybercrime in major countries in the world over the past two decades, highlighting the severity and complexity of the cybersecurity problem.

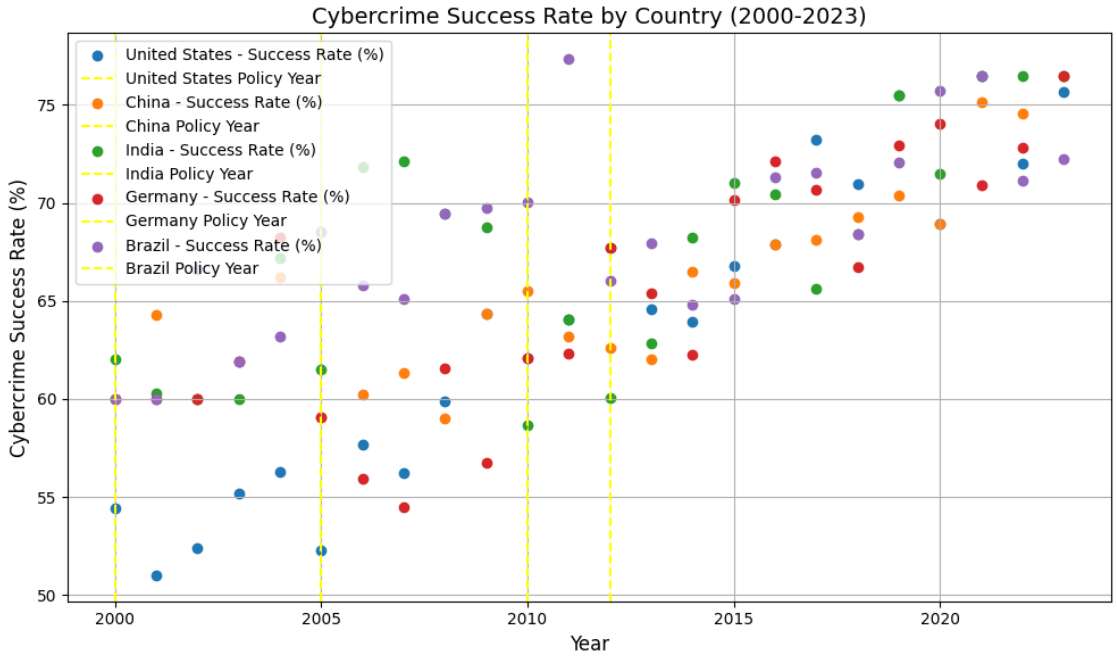

Figure 2. Scatter plot of crime success rate

It can be seen from Figure 2 that the success rate of cybercrime in various countries fluctuates during 2000-2023. In some years, the success rate varies significantly between countries, such as the relatively low success rate in the United States around 2000, and the high success rate in Brazil in some years.

The yellow dotted line clearly marks the key time nodes of the strategy introduction. From this chart data, we can observe that many countries such as India and China have been significantly affected after the implementation of policies. However, due to various complex factors, such as insufficient policy adaptability, lag in relevant laws and regulations, and various challenges at the implementation level, some countries seem to be unable to implement policies. Together, these factors make it difficult to directly and accurately evaluate the specific impact of policy on the success rate of cybercrime. Therefore, when exploring the effectiveness of policies, we need to comprehensively consider a variety of factors, and we cannot simply use the policy to be introduced as the criterion for judgment.

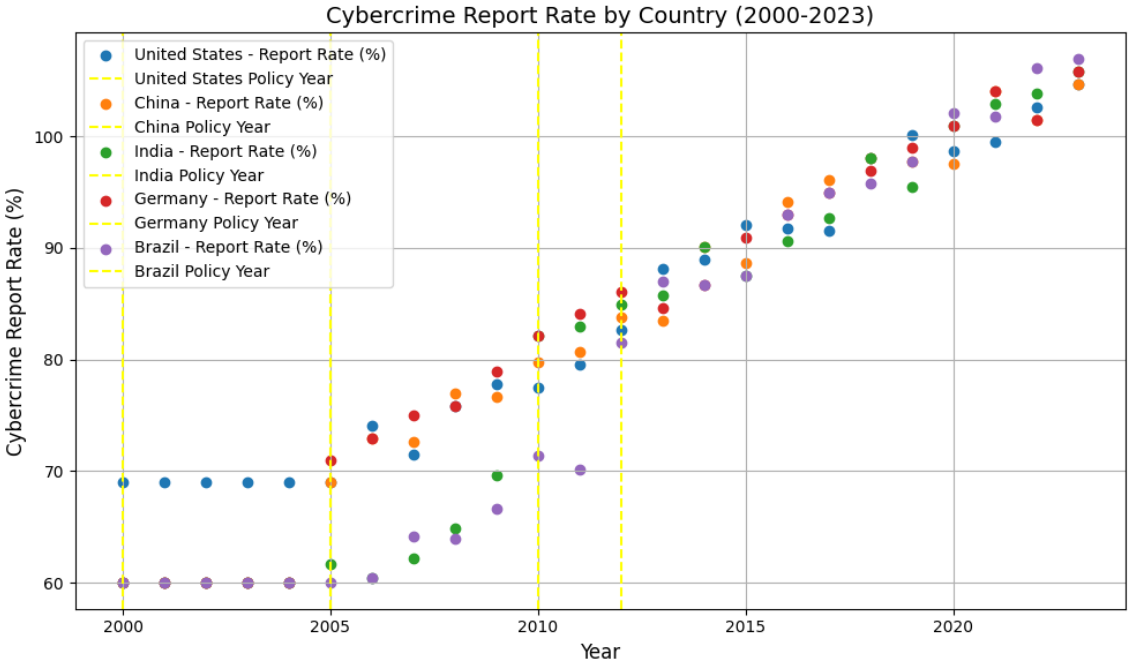

Figure 3. Scatter plot of crime reporting rate

Figure 3 shows a scatter plot of reported rates of cybercrime in five countries: the United States, China, India, Germany, and Brazil from 2000 to 2023. The different colored dots in the graph represent the rate of cybercrime reporting in different countries: blue for the United States, orange for China, green for India, red for Germany, and purple for Brazil. The dashed yellow line shows the year each country introduced its cyber policy.

As can be seen from the graph, overall, the reported rate of cybercrime by country is on the rise between 2000 and 2023. Around 2000, most countries had a reporting rate of 70% or less, and by around 2020, some countries had a reporting rate of more than 95%.

The yellow dotted line shows the time node when the policy was introduced, which can be used to observe the change of the reporting rate of cyber crimes after the policy was introduced. Generally speaking, the reporting rate has increased rapidly.

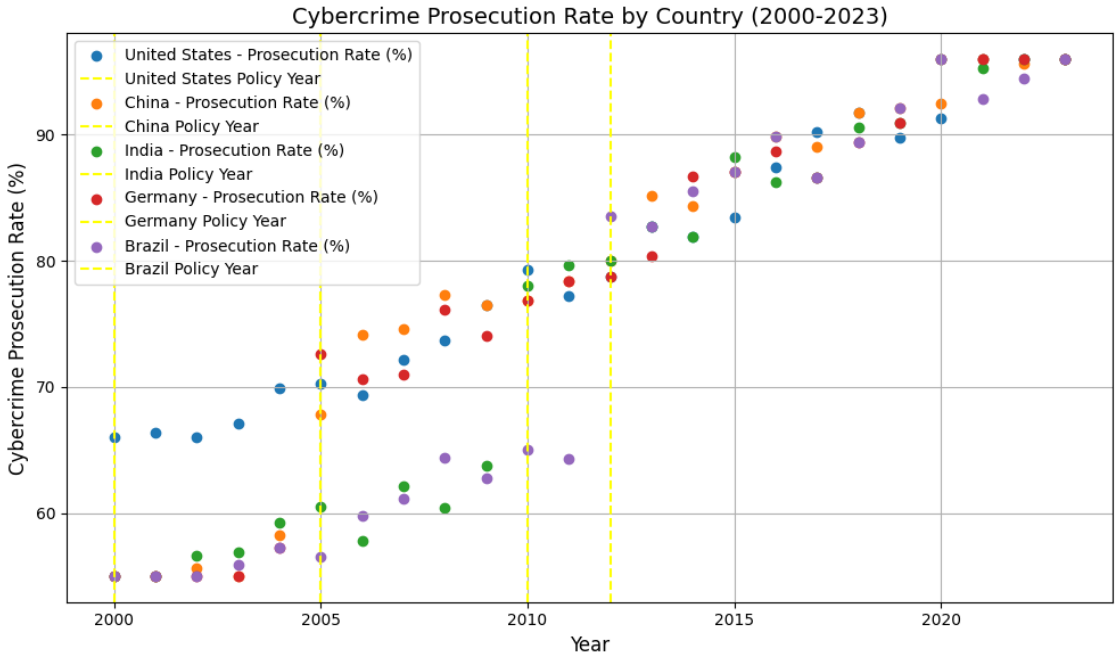

Figure 4. Scatter plot of network prosecution rate

Trends in Figure 4: Over the period 2000-2023, the overall rate of cybercrime prosecutions in the five countries showed an upward trend, indicating that the legal prosecution of cybercrime in each country has gradually strengthened.

In the early stage, the gap between countries' prosecution rates is relatively obvious, and in the later stage, the gap will narrow, and by 2020, the prosecution rate in most countries will be 90% or more. The prosecution rate in the United States and Germany was at a higher level in the later period; Prosecutions in China, India and Brazil are also increasing.

The yellow dotted line represents the policy year. It can be seen from the figure that the United States, Germany and other countries are greatly affected by policies, but it is difficult to directly see that there is a clear and immediate correlation between the introduction of policies and the change of prosecution rate in some countries, which may take time to show the effect of policies, or the prosecution of cyber crimes is affected by multiple factors.

3.5. Regression analysis results of double difference method (DiD)

Not only that, in order to verify and analyze from another perspective, we also carried out regression analysis of DiD with the help of the statsmodels library function in Python. This method can evaluate the changes before and after the implementation of the policy in a more comprehensive way, effectively control the interference of other confounding factors, and thus reveal the effect of policy issuance on cybercrime more accurately, such as Table 2 and Figure 5.

Table 2. OLS Regression Results

OLS Regression Results						
Dep.Variable	Cybercrime Incidents		R-squared		0.093	
Model	OLS		Adj.R-squared		0.086	
Method	Least Squares		F-statistic		12.16	
Date	Sun, 26 Jan 2025		Prob (F-statistic)		0.000686	
Time	17:27:31		Log-Likelihood:		-935.50	
No. Observations	120		AIC		1875.	
Df Residuals	118		BIC:		1881.	
Df Model	1					
Covariance Type	nonrobust					
	coef	std err	t	P>|t|	[0.025	0.975]
Post_t	213.4774	61.212	3.488	0.001	92.262	334.693
Treated_i	1303.6352	104.837	12.435	0.000	1096.030	1511.241
Post_t_Treated_i	213.4774	61.212	3.488	0.001	92.262	334.693
Omnibus	1.038		Durbin-Watson		0.287	
Prob(Omnibus)	0.595		Jarque-Bera JB		0.287	
Skew	0.212		Prob(JB)		0.576	
Kurtosis	2.799		Cond. No.		4.31e+15	

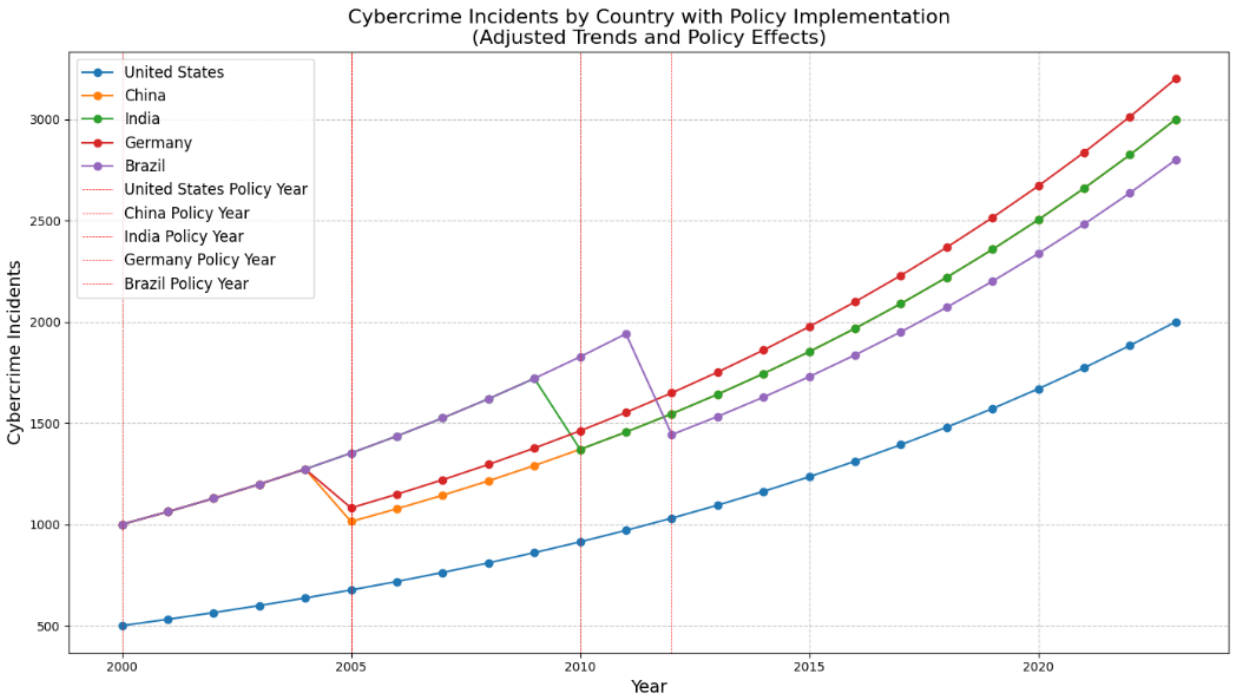

Figure 5. The impact of policy implementation on national crime incidents

Figure 5 shows the number and policy impact of cybercrime incidents in the United States, China, India, Germany and Pakistan from 2000 to 2020. Overall, the number of cybercrime incidents in various countries has generally increased. In terms of policy implementation, after the policy year marked by the red dotted line, the crime trends in some countries have been adjusted. For example, after the implementation of Brazil's policy, the growth trend of crime fluctuates, while China is relatively stable. Germany's cybercrime grew significantly, surpassing other countries in the later period. Regression analysis showed that the year of policy release had a significant impact on cybercrime, with the regression coefficient of 2198.97 and the interaction term coefficient of 2198.93, indicating that the policy release was inversely proportional to the number of cybercrimes, and $p<0.05$, with a significant linear relationship. In order to intuitively compare the policy effects of various countries, a bar chart is used for visualization. To sum up, policies have an important impact on

cybercrime. Different countries have different policies and different policies, and targeted cybersecurity policies need to be formulated and implemented to deal with the growth of crime.

3.6. Comparison of policy effects in different countries

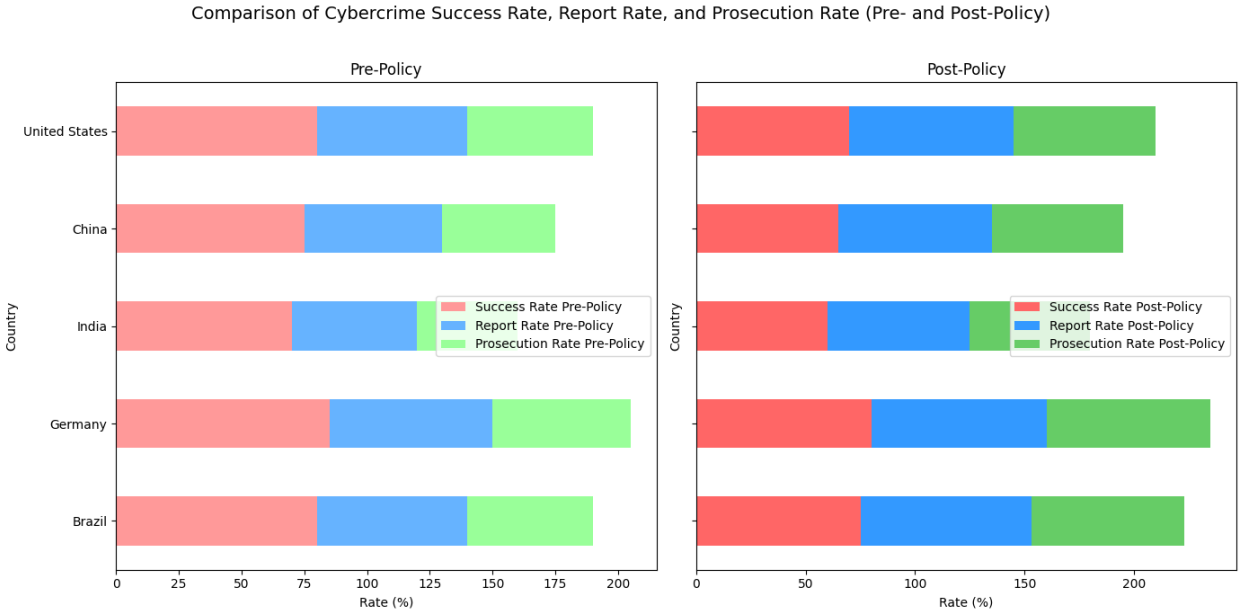

Figure 6. Comparison of Policy Implementation

Figure 6 compares the success rate, reporting rate and prosecution rate of cybercrime before and after the implementation of the policy. We use the United States, China, India, Germany and Brazil as examples. The chart is divided into two parts, the left side is before the policy implementation (Pre-Policy), and the right side is after the policy implementation (Post-Policy).

3.7. The relationship between Internet wiring rate and cyber crime

Assume that the Internet wiring rate is related to cybercrime and control variables has nothing to do with other factors. We take the data of 2023 as an example. We have established a scattered point map and fitting a scattered point diagram and linear linear between the number of Internet crimes and the number of network crimes. The regression relationship curve is Figure 7.

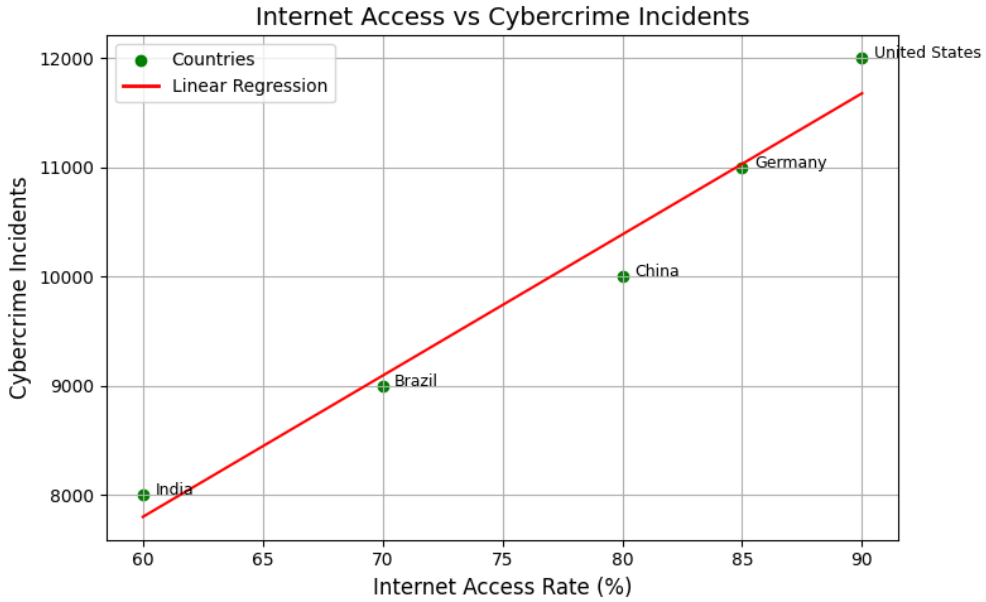

Figure 7. Scatter plot and linear regression plot of the relationship between Internet access rate and cybercrime incidents

It can be seen from Figure 7 that the X-axis represents Internet access rate (Internet Access Rate), the unit is percentage (%), and the range ranges from 60%to 90%. The Y axis represents the number of cyber criminal incidents (CyberCrime Incidents), from 8000 to 12000. The green scattered point represents the data points of different countries, and it involves the United States, Germany, China (China), Brazil, and India. Red straight line represents linear regression, showing the linear relationship trend between the Internet access rate and the number of cybercrime events.

With the increase of the Internet access rate, the number of cyber crimes has also shown an upward trend. For example, the United States has the highest Internet access rate, close to 90%, and the number of cyber crimes is the largest, exceeding 12,000; India's Internet access rate is the lowest, about 60%, and the number of cyber crime incidents is relatively small, about 8,000. The linear return line further shows that there is a positive correlation between the two, that is, the higher the Internet access rate, the more the number of cybercrime events may be.

3.8. The relationship between wealth level and cyber crime

We represent the level of wealth as a per capita GDP. In a richer region, we may attract higher -tech crimes. We draw a scattered dot diagram and linear regression map to explain the relationship between wealth and cybercrime incidents as Figure 8.

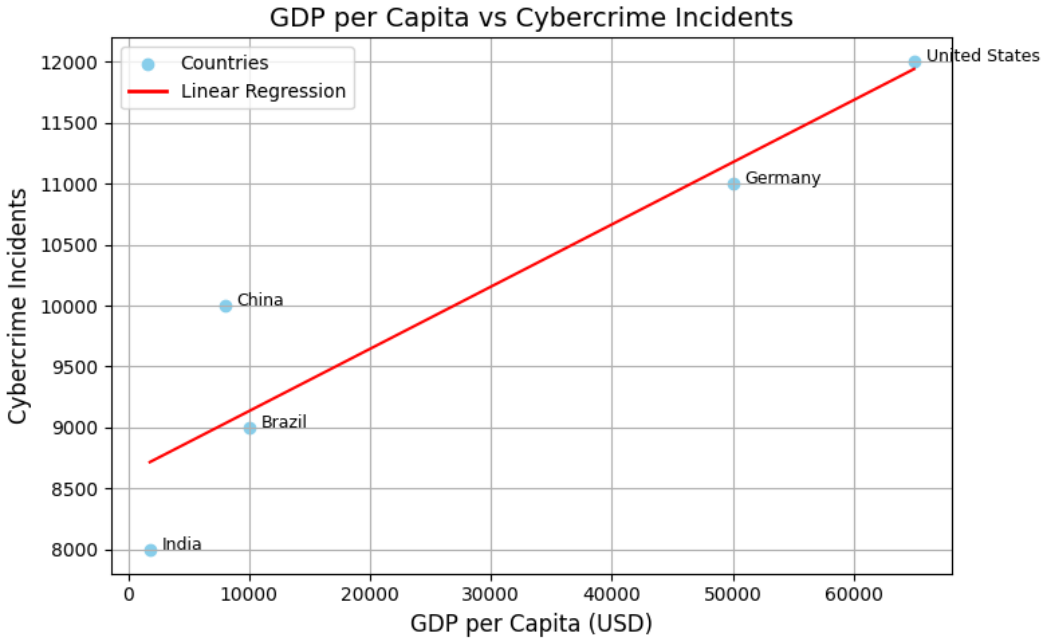

Figure 8. Scatter plot and linear regression plot of GDP per capita and cybercrime incidents

3.9. The relationship between education level and cyber crime

The level of education has a direct impact on the prevention and legal enforcement of cyber crimes. The impact of education level on the prevention of cyber crime enhances the importance of the public with a higher awareness of prevention and ability. The implementation is smooth. Increase the report rate and prosecution rate of crime. Therefore, the relationship between education level and network crime report rate and prosecution rate is worthy of in -depth analysis. Visualization: The folding diagram can help us compare the changes in the report rate and prosecution rate before and after the implementation of the policy, and then understand the impact of education level.

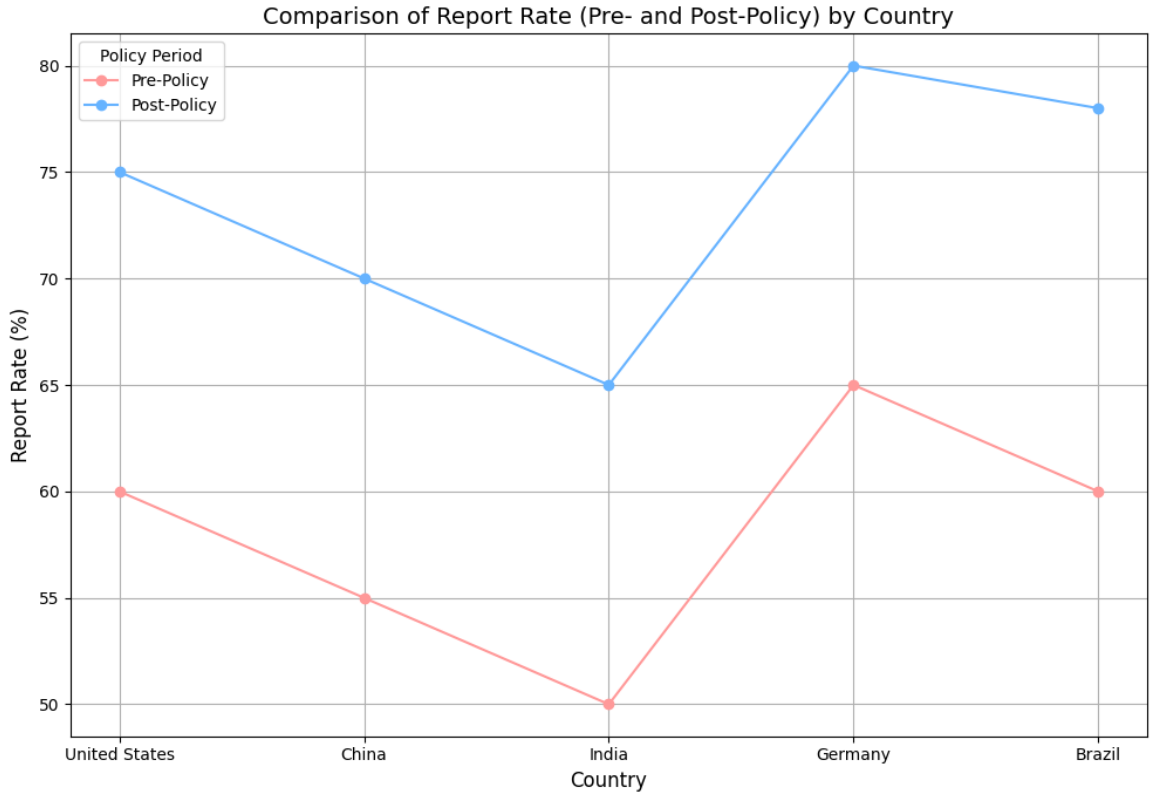

Figure 9. The impact of education on the reporting rate of cybercrime incidents

Figure 9 can be seen from the figure that the previous report rate of the United States was about 60%before the implementation of the policy, and the report rate after the policy was implemented was about 75%. The previous report rate in China was about 55%, and the report rate after the policy was implemented was about 70%. India's previous reporting rate was about 50%, and the report rate after the policy was implemented was about 65%. The pre -implementation report rate in Germany was about 65%, and the report rate after the policy was implemented was about 80%. The previous report rate of Brazil in the policy implementation was about 60%, and the report rate after the policy was implemented was about 78%.

Overall, the reporting rate of various countries after policy implementation is higher than that of the policy implementation, showing that the policy may have a positive impact on improving the report rate. Further we can know that increasing education level can increase the report rate of cybercrime.

3.10. Multivariate linear regression analysis results

The main purpose of return analysis is to solve the data problem in R language, and to interpret and check the data through a series of statistical methods. After completing the analysis, in order to display the results more intuitively, we can make visual charts, such as the residual graph and prediction comparison graph shown in Figure 3 and Table 3, which can clearly show the parameters of the model and its performance.

Table 3. Parameters

coef	std err	t	P>|t|	[0.025	0.975]	
const	2371.688	468.907	7.648	0.083	9544.369	3586.3405
Internet Access (%)	98.1186	4.304	22.797	0.028	43.431	152.806
GDP per Capita (USD)	0.0235	0.003	9.349	0.068	0.008	0.055
Education Level (%)	-23.0421	6.223	-3.703	0.168	102.110	56.026
		nan		Durbin-Watson		1.753
Prob(Omnibus)		nan		Jarque-Bera (JB)		0.568
Skew		0.544		Prob(JB)		0.753
Kurtosis		1.758		Cond. No.		7.25e+05

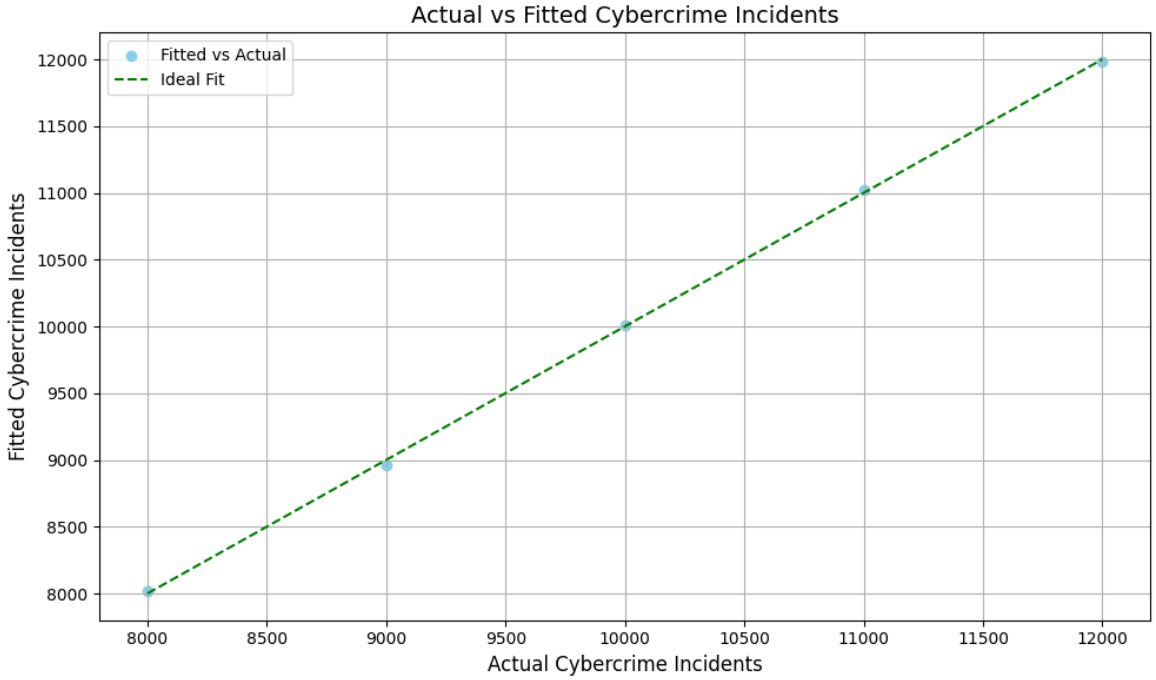

Figure 10. Actual cybercrime events vs. fitting cybercrime events

Figure 10 The horizontal axis represents the number of actual cybercriminal events (Actual CyberCrime Incidents), and the vertical axis represents the number of fat crime incidents. There are two signs in Figure 11. Blue dots represent "Fitted VS Actual" (comparison with actual comparison), and the green dotted line represents "Ideal Fit".

The blue dot represents the corresponding point of the actual data and the fitting data. From the figure, it can be seen that these points are roughly distributed near the green dotted line, but it is not completely overlapped. The green dotted line indicates the ideal fit state, that is, if the fit is completely accurate, the data point should be completely on this line. In practice, there is a certain deviation between the data point and the ideal fitting line, indicating that there are differences between the fitting model and the actual situation but acceptable. Evaluating the fitting effect of the statistical model on the actual data. In the research of cybercrime, it can help analyze the accuracy and reliability of the model built in predicting the number of cybercrime events.

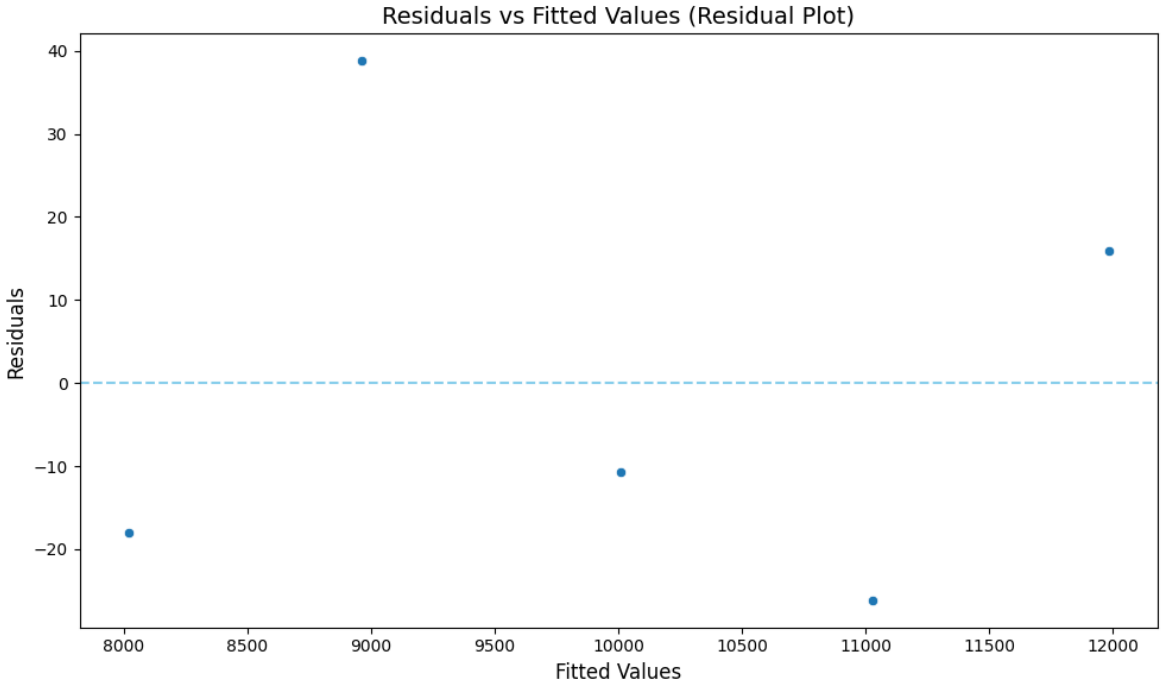

Figure 11. Residual analysis diagram

Figure 11 shows the relationship between Residuals and Fitted Values. In the regression analysis, the residue is the difference between the fitting value of the observation value and the model prediction. Residual charts are often used to evaluate the fitting effect of the regression model and whether the assumptions of the test model are established.

It can be seen from the figure that with the increase of the fitting value, the distribution of residuals does not show obvious regularity. This means that the prediction error of the model does not increase or decrease with the increase of the fit value. This is a good phenomenon, because it indicates that the predictive performance of the model's predictive values is stable. And you can observe the volatility of the residues near zero. If the residual is randomly distributed near zero and there is no obvious pattern or trend, this usually means that the model is suitable and there is no systemic error. However, if the residues present a certain model (such as increased with the increase of the fitting value Adding or reduction) may indicate that the model has a certain form of deviation or needs to be adjusted further.

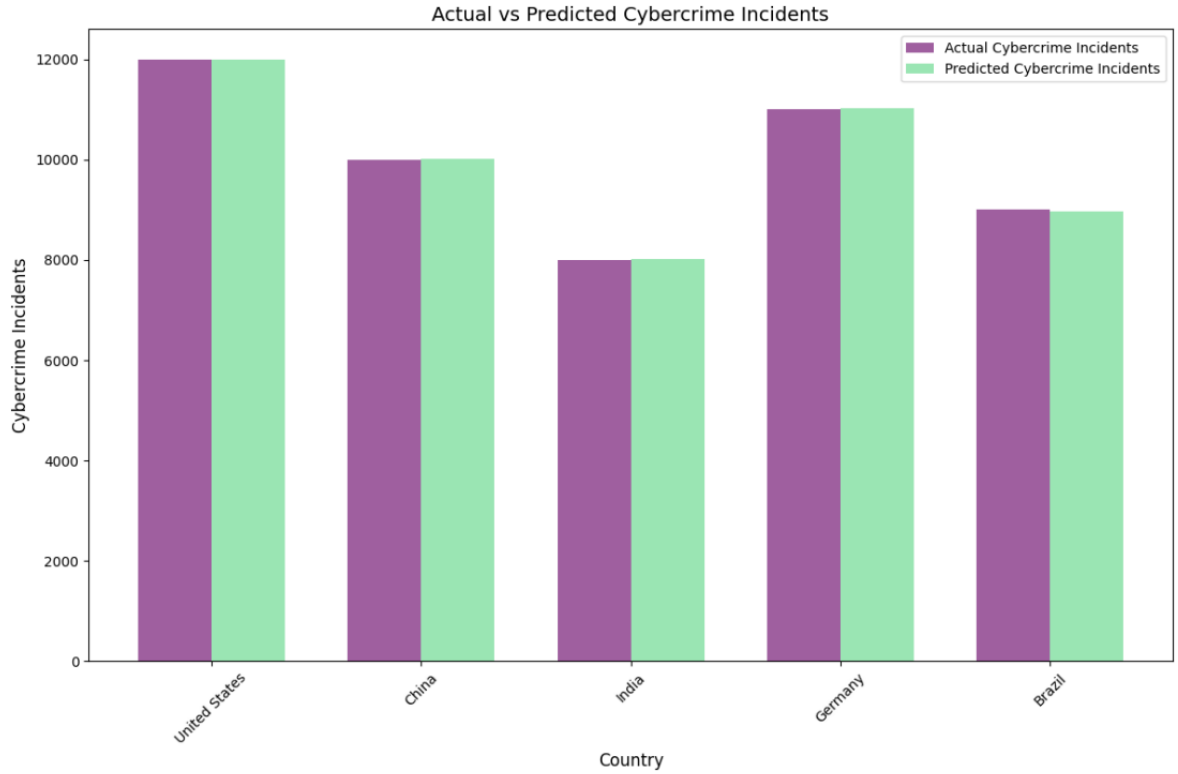

Figure 12. Histogram comparing the original and predicted data of predicted cybercrime

In Figure 12, the actual cyber criminal incident is represented by a purple column, which reflects the number of cyber criminal events actual in different years. It is expressed by a green column diagram to predict cyber criminal incidents, showing the number of cybercrime events predicted by countries in different years.

In addition, the figure also provides a magnitude of magnitude, from 2000 to 12000, with 2000 as the interval, which helps the audience to more accurately understand the magnitude of the number represented by each pillar diagram. Among them, the figure in the figure only lists the comparison of some representative national cyber crimes, and has a certain degree of limitations. From the figure, we can see that the results of the return evaluation prediction are the same as the original data prediction. For some countries, it may be further optimized.

4. Conclusions And Outlooks

This study explores global distribution and policy optimization for cybersecurity incidents and finds that areas of high incidence of cybercrime overlap with areas with weak Internet popularity, economic development, or security, which leads to an increased risk of economic losses and data breaches. Specific industries have been attacked and ransomware incidents have risen. The study further pointed out that socio-economic factors also have an impact on cybercrime. For example, the increase in Internet access rates provides more opportunities for crime, while uneven distribution of wealth and low education levels have exacerbated the incidence of crime. Therefore, improving education and strengthening cybersecurity education are seen as key measures to prevent and respond to cybercrime.

This study explores the global distribution and policy optimization of cybersecurity incidents, and found that areas with high incidence of cybercrime overlap with areas with Internet popularization, lagging economic development or weak security, resulting in economic losses and data leakage. Socio-economic factors such as uneven Internet popularization and low education levels have also aggravated cybercrime. Improving education and strengthening cybersecurity education are the key.

Future research directions can focus on refining the cybersecurity risk assessment model to more accurately predict and prevent cyber attacks. At the same time, exploring the application potential of

emerging technologies such as blockchain and quantum computing in network security, as well as how to build a stronger network security defense system through international cooperation, is also an important topic worthy of in-depth research.

References

[1] Wall D S. Cybercrime: The transformation of crime in the information age [M]. John Wiley & Sons, 2024.

[2] Sarker I H, Kayes A S M, Badsha S, et al. Cybersecurity data science: an overview from machine learning perspective [J]. Journal of Big data, 2020, 7: 1 - 29.

[3] Sarker I H, Abushark Y B, Alsolami F, et al. Intrudtree: a machine learning based cyber security intrusion detection model [J]. Symmetry, 2020, 12 (5): 754.

[4] Eschmann J. Reward function design in reinforcement learning [J]. Reinforcement learning algorithms: Analysis and Applications, 2021: 25 - 33.

[5] Barbu A, Zhu S C. Monte carlo methods [M]. Singapore: Springer Singapore, 2020.

[6] Icarte R T, Klassen T Q, Valenzano R, et al. Reward machines: Exploiting reward function structure in reinforcement learning [J]. Journal of Artificial Intelligence Research, 2022, 73: 173 - 208.

[7] Shakya A K, Pillai G, Chakrabarty S. Reinforcement learning algorithms: A brief survey [J]. Expert Systems with Applications, 2023, 231: 120495.

[8] Haim N, Vardi G, Yehudai G, et al. Reconstructing training data from trained neural networks [J]. Advances in Neural Information Processing Systems, 2022, 35: 22911 - 22924.

[9] Sarker IH, Salim FD. Mining user behavioral rules from smartphone data through association analysis. In: Proceedings of the 22nd Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), Melbourne, Australia. New York: Springer; 2018. p. 450 – 61.

[10] Liao H-J, Lin C-HR, Lin Y-C, Tung K-Y. Intrusion detection system: a comprehensive review. J Netw Comput Appl. 2013; 36 (1): 16 – 2.