E-ISSN : 2963-4946
Vol. 2 No. February 5, 2024
Asian Journal of Social and Humanities
https://ajosh.org/

Regression Analysis of The National Cyber Security Index In The Southeast Asia Region

Adityan Wisnu Yuda Prasetya, Bambang Suhardjo, Rinaldy Munir
Universitas Pertahanan Republik Indonesia, Indonesia
E-mail: adityan.prasetya@tp.idu.ac.id, bambang_suharjo@tnial.mil.id, rinaldi@staff.stie.itb.ac.id

*Correspondence: adityan.prasetya@tp.idu.ac.id

KEYWORDS
national cyber security index; regression analysis

ABSTRACT
In the current era of the Industrial Revolution 4.0, information and communication technology is rapidly developing in all sectors. In the matrix of non-military defense implementation, technology is one of the dimensions that influence the achievement of a country's general defense policy. Technological development does not always have a positive impact. Cyber threats and attacks are an example of the negative impact in technological development. In the "Cyber Defense Guidelines," national cyber security is defined as all efforts to maintain the confidentiality, integrity, and availability of information and all its supporting facilities at the national level, which are cross-sectoral in nature. According to data from the National Cyber Security Index (NCSI) survey agency, a country's National Cyber Security Index (Y) is measured based on 12 indicators, including: cyber security policy development (X1), cyber threat analysis and information (X2), education and professional development (X3), contribution to global cyber security (X4), protection of digital services (X5), protection of essential services (X6), e-identification and trust services (X7), protection of personal data (X8), cyber incident response (X9), cyber crisis management (X10), fight against cybercrime (X11), and military cyber operations (X12). In this study, a regression analysis will be conducted on the 12 indicators of the National Cyber Security Index against the Total Assessment of the National Cyber Security Index (Y) in Southeast Asia, so that the type of equation in the regression analysis can be identified.

Introduction

The Industrial Revolution is a system of industrialization that evolves over time and is massive globally. With globalization, the world has now entered the era of the Industrial Revolution 4.0. (Craveli, 2017) describes the development of the Industrial Revolution from 1.0 to 4.0 as an industrial system that relies on human labor, machines and

electricity, IT, up to digitization. The use of Information and Communication Technology (ICT) in the 4.0 industrial revolution is very extensive, including the use of computers, digital communication, smart applications, smartphones, expert systems, digital money, and others (Danuri, 2019). Another example is the use of ICT in the Indonesian Navy to connect command and control, sensors, and actuators for situational responsiveness, command speed, operational tempo, combat power, endurance, and synchronization for an operational system (Sucipto & Doheir, 2023). However, the development of ICT does not always have a positive impact. There are also state actors and non-state actors who misuse it.

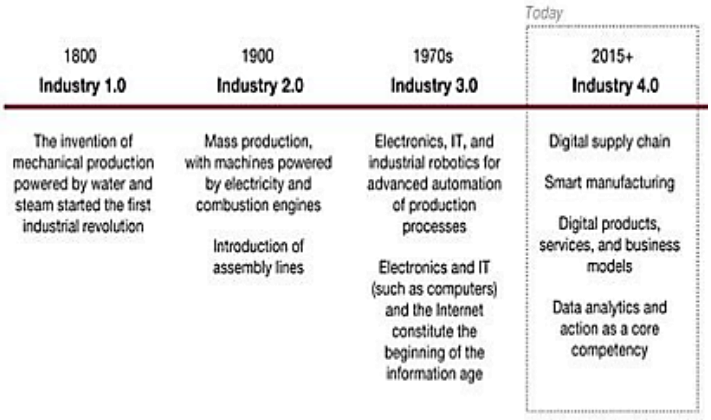

Figure 1. Development of the Industrial Revolution
Source: Craveli (2017)

There are 9 trends that are currently widespread in the era of the Industrial Revolution 4.0, one of which is cyber security (Pereshybkina et al., 2017). In the "Cyber Defense Guidelines," national cyber security is defined as all efforts to maintain the confidentiality, integrity, and availability of information and all its supporting facilities at the national level, which are cross-sectoral. National cyber security is established to confront cyber threats and attacks. These cyber threats target hardware, software, and information, which can evolve into cyber warfare. In the Cyber Security Landscape, Indonesia is one of the countries that is both a source (193,250,972 cases) and a target (539,922,976 cases) of cyber-attack anomalies (BSSN, 2022). In 2023, BSSN (2022) predicts several cyber threats and ways to confront them, including ransomware, data breaches, APT, phishing, cryptojacking, DDoS, RDP, social engineering, and web defacement (Farahbod et al., 2020).

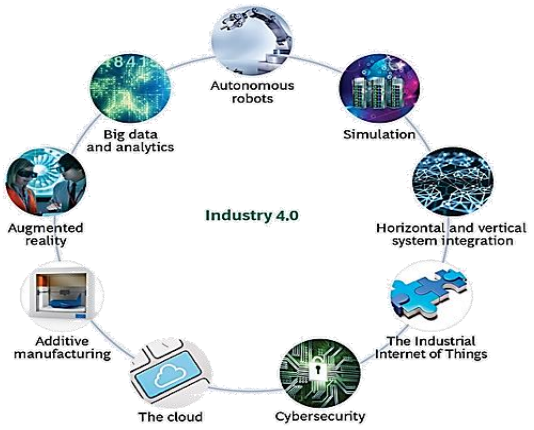

Fig 2. Trends of the Industrial Revolution 4.0
Source: (Pereshybkina et al., 2017)

Indonesia is a country in Southeast Asia with a population of 258.7 million, a land area of 1.9 million km², and a GDP per capita of $13,100 (NCSI, 2023). The National Cyber Security Index (NCSI) measures the national cyber security index based on 12 indicators. From 2019 to 2023, NCSI stated that Indonesia ranked 85th out of 164 countries globally or 6th out of 10 countries in Southeast Asia as a country with an inadequate national cyber security index. This certainly requires improvements and enhancements in the NCSI assessment indicators. According to (Koibichuk & Gerasymenko, 2022), future national cyber security should be built on 5 basic areas, namely:

1. Legal certainty;
2. Technical and procedural actions;
3. Organizational structure;
4. Capacity building and user education; and
5. International cooperation.

The development of these 5 basic areas in national cyber security (Rio & Anggraini, 2019) is expected to prevent the emergence of cyber threats at a global level with the aim of:

1. Building military and civilian geometry;
2. Preventing the spread of cybercrime;
3. Implementing policies integrated with regional and global cyber institutions; and
4. Creating a structure for Indonesia's National Cyber Defense and Security.

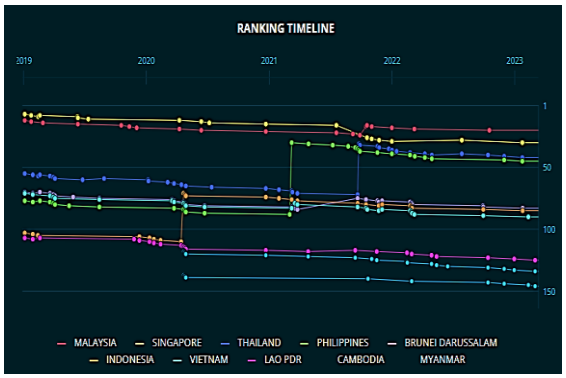

Fig 3. Percentage of Indonesia's National Cyber Security Index
Source: NCSI (2023)

Regression Analysis

Statistics is a science used for conducting experiments that include the stages of collection, organization, presentation, analysis, interpretation, and decision-making based on the data (Tyas et al., 2020). In statistics, there is a method often used to analyze the influence between a dependent variable (Y) and an independent variable (X), commonly known as the regression method. Regression analysis is divided into two types: linear regression analysis and nonlinear regression analysis. Linear regression can be defined as a form of regression equation where X is raised to the power of 1, while nonlinear regression contains X raised to a power greater than 1 (Bolpagni, 2022). Linear regression itself can also be divided into simple linear regression (containing only 1 independent variable) and multiple linear regression (containing more than 1 independent variable) (Ahmad, 2021). Apart from linear regression, data can also be analyzed using various nonlinear regressions. Nonlinear regression analysis is used to estimate regression equation models with better accuracy than linear regression because the model estimation uses iteration algorithms (Syafira & Hatta, 2023).

In regression analysis, several stages need to be carried out, including:
1. Creating a regression equation;
2. Simulating the regression equation graph;
3. Determining the coefficient of determination;
4. Conducting hypothesis testing; and
5. Making interpretations.

Based on the background above, this research will discuss the Regression Analysis of the National Cyber Security Index in the Southeast Asian Region. The expected outcome of this study is to identify the appropriate type of simple regression equation for each independent variable in relation to the dependent variable. Thus, this equation can be used as a forecasting tool for the independent variable in relation to the subsequent assessment of the dependent variable (Waseem et al., 2019).

Variable Definition

This study refers to the survey data from NCSI on the National Cyber Security Index of 10 countries in the Southeast Asian region (Malaysia, Singapore, Thailand, Philippines, Brunei Darussalam, Indonesia, Vietnam, Laos, Cambodia, and Myanmar) with the following 12 indicators:
1. Cyber security policy development (X_1);
2. Cyber threat analysis and information (X_2);
3. Education and professional development (X_3);
4. Contribution to global cyber security (X_4);
5. Protection of digital services (X_5);
6. Protection of essential services (X_6);
7. E-identification and trust services (X_7);
8. Protection of personal data (X_8);
9. Cyber incident response (X_9);
10. Cyber crisis management (X_10);
11. Fight against cybercrime (X_11); And
12. Military cyber operations (X_12) as well
13. Total National Cyber Security Index Assessment (Y).

Regression Equations

The regression equation is a mathematical equation used to model the relationship between one or more independent variables (X) and the dependent variable (Y). Regression equations can be used to predict Y values based on given X values or to understand the relationship between variables X and Y (Watters et al., 2012). Non-linear regression equations are difficult to find constant values and regression coefficients analytically. For this reason, it is necessary to transform from non-linear to linear.

Table 1. Transformation of Regression Equations

Pers Regresi	Model Non Linear	Model Linear
Linear	$Y = a + bX$	-
Logarithmic	$Y = a + bln(X)$	-
Inverse	$Y = a + \frac{b}{X}$	-
Quadratic	$Y = a + bX + cX^2$	-
Cubic	$Y = a + bX + cX^2 + dX^3$	-
Compound	$Y = ab^X$	$\ln(Y) = \ln(a) + Xln(b)$

Pers Regresi	Model Non Linear	Model Linear
Power	$Y = aX^b$	$\ln(Y) = \ln(a) + bln(X)$
S	$Y = e^{a+\frac{b}{X}}$	$\ln(Y) = a + \frac{b}{X}$
Exponential	$Y = ae^{bX}$	$\ln(Y) = \ln(a) + bX$
Logistic	$Y = \frac{1}{ab^X}$	$\ln(Y) = \ln(\frac{1}{a}) + Xln(\frac{1}{b})$

Source: Anggi (2021)

The constants and coefficients of the regression equation can be found using the formula:(Onumo et al., 2017)

1. Regression Linear

$$Y = a + bX$$

with values with values a and b obtained from:

$$b = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2}$$

$$a = \overline{Y_i} - b\overline{X_i}$$

2. Regression Logarithmic

$$Y = a + b.\ln(X)$$

with values with values a and b obtained from:

$$b = \frac{n\sum \ln(X_i) Y_i - \sum \ln(X_i) \sum Y_i}{n\sum (\ln(X_i))^2 - (\sum \ln(X_i))^2}$$

$$a = \overline{Y_i} - b\overline{\ln(X_i)}$$

3. Regression Inverse

$$Y = a + b\left(\frac{1}{X}\right)$$

with values with values a and b obtained from:

$$b = \frac{n\sum \frac{1}{X_i} Y_i - \sum \frac{1}{X_i} \sum Y_i}{n\sum \left(\frac{1}{X_i}\right)^2 - \left(\sum \frac{1}{X_i}\right)^2}$$

$$a = \overline{Y_i} - b\overline{\frac{1}{X_i}}$$

4. Regression Quadratic

$$Y = a + bX + cX^2$$

with values a, b and c obtained from:

$$\sum Y_i = na + b\sum X_i + c\sum X_i^2$$

$$\sum X_i Y_i = a\sum X_i + b\sum X_i^2 + c\sum X_i^3$$

$$\sum X_i^2 Y_i = a\sum X_i^2 + b\sum X_i^3 + c\sum X_i^4$$

5. Regression Cubic

$$Y = a + bX + cX^2 + dX^3$$

with values a, b and c obtained from:

$$\sum Y_i = na + b\sum X_i + c\sum X_i^2 + d\sum X_i^3$$

$$\sum X_i Y_i = a\sum X_i + b\sum X_i^2 + c\sum X_i^3 + d\sum X_i^4$$

$$\sum X_i^2 Y_i = a\sum X_i^2 + b\sum X_i^3 + c\sum X_i^4 + +d\sum X_i^5$$

$$\sum X_i^3 Y_i = a\sum X_i^3 + b\sum X_i^4 + c\sum X_i^5 + +d\sum X_i^6$$

6. Regression Compound

$$Y = ab^X$$

with values with values a and b obtained from:

$$ln(b) = \frac{n\sum X_i ln(Y_i) - \sum X_i \sum ln(Y_i)}{n\sum X_i^2 - (\sum X_i)^2}$$

$$ln(a) = \overline{ln(Y_i)} - ln(b)\overline{X_i}$$

8. Regression S

$$Y = e^{a+\frac{b}{X}}$$

with values with values a and b obtained from:

$$b = \frac{n\sum \frac{1}{X_i} ln(Y_i) - \sum \frac{1}{X_i} \sum ln(Y_i)}{n\sum \frac{1}{X_i}^2 - \left(\sum \frac{1}{X_i}\right)^2}$$

$$a = \overline{ln(Y_i)} - b\frac{\overline{1}}{X_i}$$

9. Regression Logistic

$$Y = \frac{1}{ab^X}$$

with values with values a and b obtained from:

$$ln(\frac{1}{b}) = \frac{n\sum X_i ln(Y_i) - \sum X_i \sum ln(Y_i)}{n\sum X_i^2 - (\sum X_i)^2}$$

$$ln(\frac{1}{a}) = \overline{ln(Y_i)} - ln(\frac{1}{b})\overline{X_i}$$

10. Regression Exponential.

$$Y = ae^{bX}$$

with values with values a and b obtained from:

$$b = \frac{n\sum X_i ln(Y_i) - \sum X_i \sum ln(Y_i)}{n\sum X_i^2 - (\sum X_i)^2}$$

$$ln(a) = \overline{ln(Y_i)} - b\overline{X}_i$$

By definition:
X_i:= X value in the i-th data
Y_i:= Y value in the ith data
a:= Regression constant
b:= Regression coefficient
n:= Number of data

Coefficient of Determination

The coefficient of determination (r^2) is a statistical measure used to evaluate the degree to which a linear regression model fits the observed data (Makridis & Smeets, 2019). The coefficient of determination is the percentage of Y that can be explained by The coefficient of determination can be calculated by the formula:

$$SSR = \sum_{i=1}^{n} \left(\widehat{Y}_i - \bar{Y}\right)^2$$

$$SST = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

$$r^2 = \frac{SSR}{SST}$$

By definition:
SSR:= Sum of Squares of Regression
SST:= Sum of Squares of Total

Hypothesis Testing

The hypothesis tests that will be used in this research are the T Test and F Test. The T Test is used to test the truth of the hypothesis that there is no significant difference between the means of two samples taken from X and Y. Meanwhile, the F Test is used to find out whether X simultaneously influences Y. The significance level used is α=5%. Hypothesis testing steps can be carried out as follows:

1. Determine H_0 and H_1 (for t test and f test);
2. Determine the level of significance (α);
3. Looking for calculated statistics (t_count and f_count);
4. Determine table statistics (t_table and f_table);
5. Comparing calculated values with table values; And
6. Draw a conclusion.

Research Methods

This study used qualitative descriptive research method. Qualitative research is a research method that aims to understand and explain a phenomenon or event in their natural context, without using numbers or quantitative data. This research involves collecting descriptive data consisting of words, images, or objects, and requires in-depth

interpretation and analysis of the data. The focus of qualitative research is on a deep understanding of the reasons, motivations, and feelings behind a behavior or event, as well as the social, cultural, and historical context in which the phenomenon occurs (Kusumastuti & Khoiron, 2019). The data collection technique in this study is by literature study obtained from Google Schoolar. The type of data used in this study is secondary data. The collected data is then analyzed in three stages, namely data reduction, data presentation and conclusion drawing.

Results and Discussions

National Cyber Security Index Data for Countries in the Southeast Asia Region

Table 2. National Cyber Security Index Data for Countries in the Southeast Asia Region

Country	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	Y
Malaysia	6	5	9	5	4	6	7	4	3	2	7	3	61
Singapore	6	5	9	5	0	6	3	4	3	5	7	2	55
Thailand	6	5	8	1	1	3	5	4	6	3	4	4	50
Philippines	6	4	6	2	1	1	5	4	5	3	9	3	49
Brunei D	0	2	7	1	0	0	5	4	5	1	7	0	32
Indonesia	0	1	4	1	1	0	8	1	4	1	7	2	30
Vietnam	2	0	6	2	4	3	1	0	6	3	1	0	28
Laos	0	1	2	1	0	0	4	1	3	1	1	0	14
Kamboja	0	1	0	1	0	0	1	0	5	0	4	0	12
Myanmar	0	0	2	1	0	0	1	0	3	1	0	0	8

Source: NCSI (2023)

With the help of the SPSS application, the data will be processed and to obtain the results of the regression equation, coefficient of determination and hypothesis test values as follows:

B. Regression Equations

1. X_1 against Y
   $Y = 19{,}2 + 3{,}721X_1 + 0{,}34X_1^2$
   (Quadratic)
2. X_2 againts Y
   $Y = 17{,}516 - 2{,}559X_2 + 5{,}478X_2^2 -$
   $0{,}692X_2^3$ (Cubic)
3. X_3 againts Y
   $Y = 9{,}89 + 1{,}084X_3 + 0{,}67X_3^2 - 0{,}023X_3^3$
   (Cubic)
4. X_4 againts Y
   $Y = 24{,}285 + 20{,}88\ln(X_4)$ (Logarithmic)

5. X_5 againts Y
   $Y = 24{,}2 + 23{,}375X_5 - 4{,}575X_5^2$
   (Quadratic)
6. X_6 againts Y
   $Y = 19{,}2 + 48{,}333X_6 - 20{,}844X_6^2 +$
   $2{,}311X_6^3$ (Cubic)
7. X_7 againts Y
   $Y = e^{3{,}885 - \frac{1{,}242}{X_7}}$ (S)
8. X_8 againts Y
   $Y = 16 + 5{,}217X_8 + 0{,}783X_8^2$
   (Quadratic)
9. X_9 againts Y
   $Y = \frac{1}{(0{,}056)(0{,}899)^{X_9}}$ (Logistic)
10. X_{10} againts Y
    $Y = 8{,}377 + 18{,}385X_{10} - 1{,}874X_{10}^2$
11. X_{11} againts Y
    $Y = 12{,}904e^{0{,}165X_{11}}$ (Exponential)
12. X_{12} againts Y
    $Y = 18{,}8 - 2{,}033X_{12} + 11{,}425X_{12}^2 -$
    $2{,}242X_{12}^3$ (Cubic)

From the processed data, it shows that the regression analysis of X_i against Y tends to use quadratic, cubic, logarithmic, S, logistic and exponential regression equations which are influenced by thesize of the coefficient of determination.

C. Coefficient of Determination

Table 3. Coefficient of Determination X_i against Y

X_i	r^2	X_i	r^2
X_1	0,824	X_7	0,462
X_2	0,855	X_8	0,749
X_3	0,86	X_9	0,035
X_4	0,537	X_{10}	0,57
X_5	0,289	X_{11}	0,544
X_6	0,772	X_{12}	0,746

The coefficient of determination above shows the level of influence of the variable X_i on Y. With the largest level of influence of X_3 on Y of 0.86 and the smallest level of

influence of X_9 on Y of 0.035. The size of the level of influence can be influenced by other factors outside the data.

D. t Test and f Test

For n=10 and α=0.05, then t_{table}=2.262 and f_{table}=4.256. The regression analysis of X_i against Y will be significant if $t_{count}>t_{table}$ and will be simultaneous if $f_{count}>f_{table}$.

Table 4. T_{count} and f_{count} values

X_i	t_{hitung}	Sig	f_{hitung}	Sim
X_1	4,735	V	16,366	V
X_2	2,831	V	11,83	V
X_3	1,156	X	12,323	V
X_4	4,517	V	3,043	X
X_5	2,97	V	1,422	X
X_6	3,853	V	6,773	V
X_7	14,213	V	6,875	V
X_8	2,559	V	10,424	V
X_9	5,096	V	0,293	X
X_{10}	1,745	X	4,632	V
X_{11}	3,36	V	9,547	V
X_{12}	3,576	V	5,879	V

The results of the t test and f test show that:

- The regression analysis of X_(1,2,6,7,8,11 and 12) on Y is significant and simultaneous.
- The regression analysis of X_(4,5 and 9) on Y is significant and not simultaneous.
- The regression analysis of X_(3 and 10) against Y is simultaneous and not significant.

Data that is not significant or not simultaneous can be caused by other factors outside variable X_i that influence variable Y.

Conclusion

Based on the results of the regression analysis of the National Cyber Security Index for Countries in the Southeast Asia Region, it can be concluded that regression analysis uses quadratic, cubic, logarithmic, S, logistic and exponential regression equations. The largest level of influence of variable Xi on Y is 0.86 and the smallest is 0.035. Hypothesis testing shows that not all of the variables Xi to Y are significant and simultaneous, but there are also some that are neither significant nor simultaneous

References

Ahmad, M. (2021). Optimization of Learning Effectiveness through the Learning Environment and Learning Motivation for Students. *Al-Ishlah: Jurnal Pendidikan*, *13*(3), 1546–1555.

Bolpagni, M. (2022). Cyber risk index: a socio-technical composite index for assessing risk of cyber attacks with negative outcome. *Quality & Quantity*, *56*(3), 1643–1659.

Craveli, A. (2017). *Department of Business and Management Course of Digital Transformation*.

Danuri, M. (2019). Perkembangan dan transformasi teknologi digital. *Jurnal Ilmiah Infokam*, *15*(2).

Farahbod, K., Shayo, C., & Varzandeh, J. (2020). Cybersecurity indices and cybercrime annual loss and economic impacts. *Journal of Business and Behavioral Sciences*, *32*(1), 63–71.

Koibichuk, V., & Gerasymenko, V. (2022). Effectiveness of National Cyber Security: DEA analytics. *Herald of Economics*, *3*, 8–21.

Kusumastuti, A., & Khoiron, A. M. (2019). *Metode penelitian kualitatif*. Lembaga Pendidikan Sukarno Pressindo (LPSP).

Makridis, C. A., & Smeets, M. (2019). Determinants of cyber readiness. *Journal of Cyber Policy*, *4*(1), 72–89.

Onumo, A., Cullen, A., & Ullah-Awan, I. (2017). An empirical study of cultural dimensions and cybersecurity development. *2017 IEEE 5th International Conference on Future Internet of Things and Cloud (FiCloud)*, 70–76.

Pereshybkina, A., Conde, M. E. C., & Kalyesubula, T. (2017). How will the industry 4.0 transformations affect SMEs in Germany by 2030. *Hochschule Furtwangen University*.

Sucipto, A., & Doheir, M. (2023). Dijkstra-based Official Motorcycle Repair Shop Application for Determining the Shortest Route. *Journal of Applied Intelligent System*, *8*(2), 227–236.

Syafira, P. A., & Hatta, M. I. (2023). Pengaruh Self Determination terhadap Work Engagement pada Mahasiswa Its yang Mengikuti Magang. *Jurnal Riset Psikologi*, 75–84.

Tyas, E. H., Sunarto, S., & Naibaho, L. (2020). Building superior human resources through character education. *TEST Engineering & Management*, *83*, 11864–11873.

Waseem, A., Rashid, Y., Warraich, M. A., Sadiq, I., & Shaukat, Z. (2019). Factors affecting E-commerce potential of any country using multiple regression analysis. *Journal of Internet Banking and Commerce*, *24*(3), 1–28.

Watters, P. A., McCombie, S., Layton, R., & Pieprzyk, J. (2012). Characterising and predicting cyber attacks using the Cyber Attacker Model Profile (CAMP). *Journal of Money Laundering Control*, *15*(4), 430–441.