J. Indones. Math. Soc.
Vol. 30, No. 03 (2024), pp. 338–351.

ANALYSIS OF UNEMPLOYED YOUTH IN INDONESIA BY PANEL DATA REGRESSION WITH MODERATING VARIABLE

Sri Subanti[1], and Nabila Tri Amanda[2]

[1]**Universitas Sebelas Maret, Surakarta, Central Java, Indonesia, sri_subanti@yahoo.co.id**
[2]**Universitas Sebelas Maret, Surakarta, Central Java, Indonesia, nabilatriamanda32@gmail.com**

Abstract. Indonesia is entering the era of demographic bonus with the productive age dominating the population. Productive age is the main focus of the government in maximizing the demographic dividend, but Indonesia has the highest percentage of Not in employment, education or training (NEET) in Asia. NEET are people on 15-24 years old who do other activities outside of school, work or training. This study aims to analyze NEET in Indonesia using panel data with moderate regression analysis. The analysis of multiple linear regression is focused on the relationship between the independent and dependent variables without taking into other possible outcomes. By inserting a moderating variable, this study explores the relationship between the independent and dependent variables differently and aims to strengthen or weaken it. Under certain conditions, the relationship between the independent and dependent variables can be explained by the moderating variable. The research data used were obtained from the employment book and the website of BPS Indonesia, in the form of 34 cross section and 5 years time series data that tends to be stationary. The dependent variable is NEET with 5 independent variables including Human development index, the open unemployment rate, labor force participation rate, proportion of individuals who own phone, and proportion of informal employment. The moderating variable is the proportion of youth aged with ICT skills. The best model in regression analysis panel data is FEM with 4 significant independent variables and 92.75% of R-square. Moderating variable can moderates the relationship of NEET with its independent variables and increased the R-square to 94.19%.

Key words and Phrases: Not In Employment Education Or Training, Moderating, Panel Data Regression

2020 Mathematics Subject Classification: 91B82
Received: 05-01-2024, accepted: 09-09-2024.

1. INTRODUCTION

Indonesia is a country that ranks fourth in the world as a country with a large population. Based on Worldometers Indonesia [13], Indonesia's population in 2020 was 3.29% of the global population, or as many as 269 million people. The findings of the National Labor Force Survey (Sakernas) in August 2021, Indonesia even entered the demographic bonus era, namely the young productive age (15-64 years) dominated the number of residents in the country. According to EUROFOUND [3], this demographic dividend can occur because the number of dependency groups is low, but this cannot be realized if the policies implemented by the government and private sector to increase labor supply and youth labor force participation are not realized.

In 2017, the International Labour Organization (ILO) defined NEETs as young individuals aged 15 to 24 who are not employed, educated, or trained, education or training. Indonesia has one of the highest NEET rates in Asia [6]. Referring to Figure 1, the percentage of NEET in Indonesia exhibited an increase

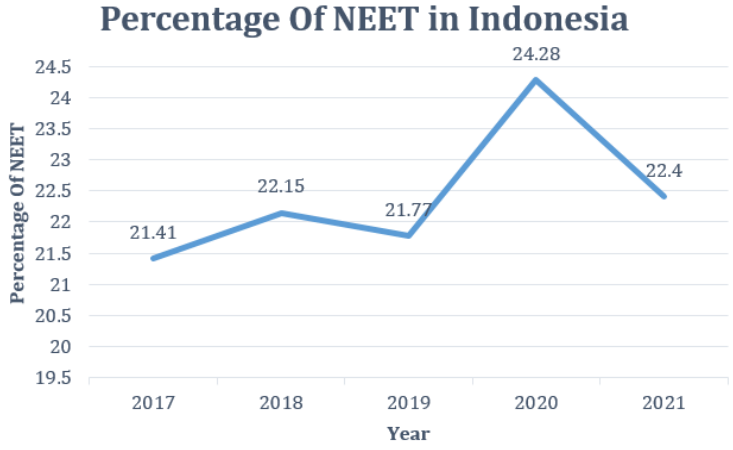

FIGURE 1. Indonesia's Percentage of NEET

in 2018. Despite a decrease to 21.77% in 2019, the highest peak occurred in 2020, this is allegedly reinforced by the Covid-19 pandemic. The NEET issue is crucial and demands attention from the government. A substantial and prolonged NEET phenomenon will hamper the economy, industrial development, and endanger the stability and development of a country.

The percentage of young NEET in Indonesia is assumed to be influenced by several variables. This study aims to analyze the independent variables that affect the young NEET with the moderating variables. Multiple linear regression analysis allows the independent variable to impact the dependent variable without considering any other influences. The aim of this study is to gain a more complete

understanding of the phenomenon being investigated by incorporating variables that are presumed to enhance or weaken the relationship between the independent and dependent variables. The moderating variable is able to explain the possibilities and conditions under which the relationship between the independent and dependent variables can change. The independent variables include the human development index (HDI), open unemployment rate, the proportion of individuals who own mobile phones, labor force participation rate, and the proportion of informal employment. The moderating variable in this study was the proportion of adolescents and adults aged 15-24 years with information and computer technology (ICT) skills. Based on the previous explanation, the percentage of young NEET will be analyzed using panel data regression with moderation variables. This study is expected to provide an overview of young NEET in Indonesia, in order to provide insights for the government and the public in minimizing the number of NEET in Indonesia.

2. THEORETICAL BASIS

2.1. **Not in Employment, Education or Training (NEET).**

The NEET percentage is calculated using Equation . The Central Statistics Agency defines Not in Employment, Education or Training (NEET) as young individuals who participate in other activities outside of school, work, or training. The population in the young age category is the male and female population aged 15-24 years 1. The NEET percentage is calculated using Equation (1).

$$\mathrm{PrNEET} = \left(\mathrm{PTB}_{15-24} + \mathrm{PTS}_{15-24} + \mathrm{PTT}_{15-24}\right) / \mathrm{P}_{15-24} \times 100\%, \tag{1}$$

where

PrNEET : percentage of young people who are not in school, work, or training (%);
PTS_{15-24} : the number of young people who do not go to school (people);
PTB_{15-24} : the number of young people who do not work (people);
PTT_{15-24} : the number of young people who do not attend training (people);
P_{15-24} : total population aged 15-24 years (people).

2.2. **Regression Analysis of Panel Data.**

Based on Gujarati [5], panel regression is a regression with a panel data structure that combines cross-sectional data and time-series data. Cross-sectional data are observations collected from one or more units at a single point in time, while time-series data are observations collected over time for one or more variables. The panel data regression equation is shown in Equation (2).

$$Y_{it} = \alpha_{it} + \beta' X_{it} + \mu_{it}, \tag{2}$$

where

Y_{it} : individual unit response variable at period t;

α_{it} : intercept coefficients of the i-th individual units and t-th time periods slope;

β' : slope coefficient vector;

β_1 : coefficient of independent variable;

β_2 : coefficient of moderating variable;

β_3 : coefficient of interaction;

X_{it} : predictor variable of i-th individual unit at period t;

M_i : moderating variable;

μ_{it} : residual component with mean 0 and variance σ^2.

Belonging to Baltagi [2], panel data regression has three approaches, namely Common Effect Model (CEM), Fixed Effect Model (FEM), and Random Effect Model (REM).

2.2.1. *The Common Effect Model (CEM).*

Combines data without regard to individuals and time. Refers to Baltagi [2], the CEM regression equation is shown in Equation (3).

$$Y_{it} = \alpha + \beta' X_{it} + \mu_{it}. \tag{3}$$

2.2.2. *Fixed Effect Model (FEM).*

FEM is an estimation of panel data that can be differentiated by individual and time. The slope coefficient of the FEM model is constant, but the intercept coefficient changes with each individual and time. Appropriate to Gujarati [5], the FEM equation is shown in Equation (4).

$$Y_{it} = \alpha + \mu_i + \lambda_t + \beta' X_{it} + \mu_{it}. \tag{4}$$

2.2.3. *Random Effect Model (REM).*

The correlation between error terms due to changes in time and individuals is involved in REM, Greene [4]. The REM Model is often called the generalized least square with Equation 5.

$$Y_{it} = \alpha_{it} + \beta' X_{it} + \omega_{it}. \tag{5}$$

2.3. **Panel Data Regression Model Selection.**

In choosing the panel data regression model, two tests are performed: the Chow Test and Hausman test. CEM and FEM are compared to select the best model in the Chow Test. This test can be done by looking at the significance of FEM model using the F-statistical test. Based on Widarjono [12], the Hausman test is used to test between two models, fixed effect model, and random effect model. If both tests are performed, and the p-value is smaller than the significance level, then the null hypothesis is rejected, and the chosen model is the fixed effect model (FEM).

2.4. **Classical Assumption Test.**

Normality tests are used to determine if variables that disrupt or remain have a normal distribution or not in regression models. A good regression model should have a normal or near-normal distribution. To test whether the data is normally distributed or not, the Jarque-Bera test (J-B) can be employed, Baltagi [2].

Multicollinearity test is necessary to determine whether or not there are independent variables that share similarities between them in a regression model. If correlation exists, it is indicated that the regression model is experiencing a multicollinearity issue, Gujarati [5].

The multicollinearity test is done by examining at the value of tolerance and the value of the variance inflation factor (VIF). Multicollinearity problems are not present when the VIF value is below 10.

To test heteroscedasticity, the glacier test can be applied. The Glejser test is conducted by making a regression equation of the absolute value of the residue with the independent variable [5].

The autocorrelation test is aimed at establishing a correlation in a linear regression model between the disturbance error (residue) in the period t with the error in the period $t-1$ (previous) [2]. If there is a correlation, then there is a problem of autocorrelation. A regression model is considered good when detached from autocorrelation. The Durbin-Watson test can be used to determine whether autocorrelation is present or absent.

2.5. **Moderated Regression Analysis (MRA).**

Based on Liana [8], Moderated Regression Analysis (MRA) is a special application of multiple linear regression in which the regression equation contains interaction (multiplication of two or more predictor variables). The moderated regression model is shown in Equation (6):

$$Y_i = \alpha + \beta_1 X_i + \beta_2 M_i + \beta_3 X_i \cdot M_i + \varepsilon. \tag{6}$$

The classification of moderation variables is determined by the significance of the parameters in the regression model of equation 6, which is explained in Table 1. Table 1 shows the classification of moderation types based on the significance of coefficients.

Table 1. Classification of Moderation

Moderation Type	Coefficient
Absolute Moderation	β_1 and β_2 insignificant
	β_3 significant
Pure Moderation	β_2 insignificant
	β_3 significant
Quasi Moderation	β_2 significant
	β_3 significant
Homologiser Moderation	β_2 insignificant
	β_3 insignificant
Prediktor Moderation	β_2 significant
	β_3 insignificant

2.6. **Model Goodness Test.**

Model goodness test consists of 3 tests, namely simultaneous test (F-test), Partial Test (t-test), and coefficient of determination. Hypothesis testing research which includes F-test is done to determine the presence or absence of the influence of predictor variables with response variables, Misbahuddin and Hasan [7]. If the value of $F_{\text{count}} > F_{\text{table}}$ or p-value is smaller than the significant level, the null hypothesis is rejected, indicating that the independent variables jointly affect the dependent variable.

Partial test is used to test the coefficient (slope) regression individually. This test is done for each coefficient separately [7]. If the value of $t_{\text{count}} > t_{\text{table}}$ or p-value is less than the significant level, then the null hypothesis is rejected. This means that the independent variables individually affect the dependent variable.

3. METHODOLOGY

3.1. **Study Area and Data Sources.**

The study was conducted on the percentage of Not in Employment Education Training (NEET) young people (15-24 years) based on provinces in Indonesia in 2017-2021. The Labor book in the third edition of 2021 is where we got the data for this study, the Satu Data Ketenagakerjaan website, and the official website of the Central Statistics Agency (BPS). The observation units used are the 34 provinces in Indonesia.

3.2. **Study Variables.**

Variables used in this study consisted of a dependent variable, five independent variables, and one moderating variable. The explanation of variables is shown in Table 2.

3.3. **Research Methodology.**

The software used is Microsoft Excel and EViews 10. The steps taken in this study are as follows:

TABLE 2. Variables used in this research

Category	Name of Variable
Y	Percentage of Not In Employment Education Or Training (NEET) Of Youth (15-24 Years old)
X_1	Human Development Index (HDI)
X_2	The open unemployment rate
X_3	The labor force participation rate
X_4	Individuals who own a mobile phone
X_5	The proportion of informal employment
M (moderating variable)	The proportion of youth aged 15-24 years old with information technology and computer skills

(1) Determining the dependent, independent, and moderating variables
(2) Building the regression models: Common Effect Model (CEM), Fixed Effect Model (FEM), and Random Effect Model (REM)
(3) Performing model selection tests, namely Chow test and Hausman Test
(4) Conducting assumption tests consisting of heteroskedasticity and multicollinearity tests
(5) Performing Moderated Regression Analysis on predictor variables to response variables with the moderating variables
(6) Checking the goodness of fit of the MRA model panel data using Partial Test (t-test), simultaneous test (F-test), and coefficient of determination (R^2)
(7) Model interpretation and drawing conclusions

4. RESULT AND DISCUSSION

4.1. **Regression model estimation.**

Analysis regression model for the NEET variable, found that the X4 variable did not significantly affect the NEET variable, so the X4 variable was excluded from the model. The estimation results of the CEM regression parameters are shown in Table 3.

TABLE 3. Estimation of Common Effect Model

Var	Coeff	Prob
C	113.94	0.00
X1	-0.73	0.00
X2	0.32	0.01
X3	-0.83	0.00
X5	0.17	0.00
M	0.08	0.00

The estimation results of the FEM regression parameters are shown in Table 4. Fixed effect model has fixed Intercept values that are used between individuals.

Table 4. Estimation of Fixed Effect Model

Var	Coeff	Prob
C	-48.21	0.076
X1	1.11	0.007
X2	0.44	0.005
X3	-0.29	0.003
X5	0.26	0.00
M	-0.04	0.12

The estimation results of the REM regression parameters are shown in Table 5. Random effect model has intercept value that is not considered a constant, but is considered a random variable with an average value.

Table 5. Estimation of Random Effect Model

Var	Coeff	Prob
C	74.10	0.000
X1	-0.52	0.000
X2	0.59	0.000
X3	-0.47	0.000
X5	0.21	0.000
M	0.06	0.001

Based on Tables 4, 5, and Table 3 it can be seen that the variables X_1, X_2, X_3, and X_5 are significant in all three models, but the moderation variable has a p-value > 0.05 in the estimated fixed effect model (FEM) which means that the moderation variable is not significant simultaneously to the relationship between the variable Y and the predictor variable. Judging from the value of R-square, common effect model is worth 71%, fixed effect model is worth 92%, and random effect model is worth 50%, with this FEM model has the ability to explain the dependent variable highest among 2 other models.

4.2. **Selection of Model Estimates.**

The results of the Chow and Hausman tests are shown in Table 6.

Table 6. The Chow and Hausman Test

Effects Test	Statistic	Prob.
Cross-section F	11.981	0.0000
Cross-section Chi-square	236.446	0.0000
Cross-section random	40.982	0.0000

Based on Table 6, it can be seen that in the Chow Test, a probability value of $0.00 < \alpha(0.05)$ is obtained. Therefore, the fixed effect model has been proven to be better than the common effect model. The testing continued to the Hausman

test with a probability value of $0.00 < \alpha(0.05)$. Thus, the test is stopped, and the selected model is the fixed effect model.

4.3. **Panel Data Regression Model Equation.**

The selected model through testing is the fixed effect model shown in Table 7.

Table 7. Fixed Effect Model (FEM)

Variable	Coefficient	Prob.
C	-48.21	0.076
X1	1.116	0.007
X2	0.446	0.005
X3	-0.29	0.003
X5	0.262	0.000
M	-0.043	0.121

The panel data regression equation with moderation variables for the fixed effect model (FEM) as the estimation of the selected model is written as follows:

$$\text{NEET} = -48.21 + 1.11X_1 + 0.44X_2 - 0.29X_3 + 0.26X_5 - 0.04M + \varepsilon.$$

From the regression model formed, the interpretation obtained is that every increase in one unit of HDI with other variables constant, then NEET in Indonesia will increase by 1.11%. If there is an increase of one unit of the open unemployment rate with other variables constant, then NEET in Indonesia will increase by 0.4%. Next, for each increase in the labor force participation rate with other variables constant, then NEET in Indonesia will decrease by 0.2%, and each increase of one unit of the proportion of informal employment with other variables constant, then NEET in Indonesia will increase by 0.2%.

4.4. **Classical Assumption Test.**

Heteroscedasticity test results are shown in Table 8.

Table 8. Heteroscedasticity Test Results

Variable	Coefficient	t-Stat	Prob.
C	12.197	0.854	0.394
X1	-0.086	-0.397	0.691
X2	0.042	0.517	0.605
X3	-0.091	-1.764	0.079
X5	-0.002	-0.081	0.934
M	0.011	0.744	0.458

The heteroscedasticity test is done by the Glejser method, which is to regress the independent variable to the absolute value of the residual. Based on Table

8, it is known that the value of p-value $= 0.00 < \alpha(0.05)$; the residuals have a homogeneous variety.

Multicollinearity test results are shown in Table 9.

Table 9. Multicollinearity Test Results

Variable	VIF	Result
X1	3.091	No multicollinearity
X2	1.864	No multicollinearity
X3	1.725	No multicollinearity
X5	1.298	No multicollinearity
M	2.822	No multicollinearity

Based on Table 9, it can be seen that the Variance Inflation Factor (VIF) values are all less than 10, which means that there is no multicollinearity. The results show that the independent variables used and the dependent variable or related variables are not in a linear relationship.

4.5. **Moderated Regression Analysis.**

Moderated Regression Analysis results are shown in Table 10.

Table 10. Moderated Regression Analysis

Variable	Coefficient	t-Statistic	Prob
C	-149.165	-3.871	0.0002
X1	1.917	3.996	0.0001
X2	0.151	0.225	0.8223
X3	0.272	0.939	0.3492
X5	0.426	3.763	0.0003
M	1.662	4.458	0.0000
X1*M	-0.0137	-4.176	0.0001
X2*M	0.008	1.118	0.2654
X3*M	-0.009	-2.380	0.0188
X5*M	-0.003	-1.915	0.0577

Based on Table 10, the panel data regression model with moderation variables can be written as follows:

$$\text{NEET} = -149.165 + 1.917X_1 + 0.0003X_5 + 1.662M - 0.0137X_1M - 0.009X_3M.$$

From the equation, the interpretation is as follows: every increase of one unit of HDI variable with other variables constant, then NEET will increase by 1.917 percent. Every increase of one unit of the informal employment variable proportion with other variables constant, then NEET will increase by 0.0003 percent. Each increase of one unit variable proportion of young age with ICT skills with other

variables constant, then NEET will increase by 1,662 percent. Every increase of one unit of interaction variable between HDI with ICT skills proportion with other variables is constant, then NEET will decrease by 0.0137 percent. Each increase of one unit of interaction variable between the labor force participation rate with ICT skills proportion with other variables is constant, then NEET will decrease by 0.009 percent.

4.6. **Test Goodness Test of Model MRA Panel Data.**

Goodness test of the panel data MRA model is conducted by examining the simultaneous and partial significance of moderation variables and their interactions. Additionally, the comparison between the coefficient of determination of the MRA model and panel data regression without interaction is observed.

4.6.1. *Simultaneous Influence Significance Test (F Test).*

The F test is conducted to determine the presence or absence of the influence of independent variables together against the dependent variable.

TABLE 11. F-test

F-statistic	**Prob**
49.075	0.000

Based on Table 11, moderated regression analysis shows that the value of Prob (F-stat) is $0.000 < 0.05$. This indicates that the variables X_1, X_2, X_3, X_4, and moderating variables together affect the percentage of NEET variables in Indonesia.

4.6.2. *Partial Significance Test (t-test).*

Partial testing is performed to determine the nature of the moderating variables in the relationship between dependent and independent variables.

TABLE 12. Partial test of moderation against X_1

Variable	**Coeff**	**Prob**
C	-42.033	0.305
X_1	0.884	0.149
M	0.063	0.814
Interaksi	-0.0006	0.861

Based on Table 12, the coefficient values of β_1, β_2, and β_3 are not significant because they have a probability > 0.05. Thus, the moderation variable is categorized as homologiser moderation for the relation of NEET and HDI.

TABLE 13. Partial test of moderation against X_2

Variable	Coeff	Prob
C	18.108	0.000
X_2	-0.108	0.838
M	-0.010	0.741
Interaksi	0.012	0.045

Based on Table 13, it is known that the value of the coefficients β_3 is significant with a probability of < 0.05, which is equal to 0.045, but the value of coefficients β_1 and β_2 is not significant because it has a probability > 0.05, so the moderation variable is categorized as pure moderation of the relations between the open unemployment rate with the percentage of NEET.

TABLE 14. Partial test of moderation against X_3

Variable	Coeff	Prob
C	16.367	0.3762
X_3	-0.011	0.9660
M	0.406	0.1315
Interaksi	-0.004	0.2221

Based on Table 14, it is known that the coefficient values of β_1, β_2, and β_3 are not significant because they have a probability > 0.05, so the moderation variable is homologiser moderation for the relation of NEET and labor force participation rate.

TABLE 15. Partial test of moderation against X_5

Variable	Coeff	Prob
C	-2.433	0.6729
X_5	0.517	0.0001
M	0.129	0.1006
Interaksi	-0.002	0.1157

Based on Table 15, it is known that the coefficient values of β_1, β_2, and β_3 are not significant because they have a probability > 0.05, so the moderation variable is homologiser moderation for the relation of NEET and the proportion of informal employment.

4.7. **Coefficient of Determination.**

Tabel 16 shows the coefficient of determination for models with and without moderator variables.

Table 16. Coefficient of Determination

R-squared with MRA	R-squared without MRA
0.941	0.928

From the test results, obtained coefficient of determination (R^2) of 0.941. This means that the ability of the model can explain the diversity of the dependent variable by 94%. According to this analysis, the model's ability to explain the dependent variable can be improved by 1.3% by using panel data regression with moderation variables.

5. CONCLUDING REMARKS

5.1. **Conclusion.**

Based on the results of the analysis, it was concluded that NEET analysis using panel data regression has the best model of Fixed Effect Model with the following equations:

$$NEET = -10.9486 + 0.55X_1 + 0.48X_2 - 0.28X_3 + 0.23X_5 + \varepsilon.$$

The Human Development Index variable, open unemployment rate, labor force participation rate, and proportion of informal employment influence the percentage of the NEET variable in Indonesia with an R-square of 92%. The equation Model, including moderation variables, is considered better because it can increase the value of R-square to 94% with the following equation:

$$NEET = -149.16 + 1.918X_1 + 0.426X_5 + 1.66M - 0.013X_1 \cdot M - 0.009X_3 \cdot M + \varepsilon.$$

Moderating variable, the proportion of young age with ICT skills, moderates the relationship of the NEET percentage variable with its independent variable. The interaction between the variable X_1 and the moderating variable is a homologizer moderation for the relationship between Y with X_1. The interaction between the variable X_3 and the moderating variable is also a homologizer moderation for the relationship between Y with X_3.

5.2. **Suggestion.**

Based on the results and discussion of this research, there are several suggestions that can be carried out for further research, namely the need to add more independent variables that influence NEET. Researchers can also compare interaction tests with regression analysis which is moderated by several independent variables with other independent variables. In addition, there is a need for various methods in analyzing the NEET variable in the case of moderating variables, for example using the logistic regression method with moderating variables. This research is also open to constructive criticism and suggestions, in order to perfect the analysis more clearly.

REFERENCES

[1] Badan Pusat Statistik. https://www.bps.go.id/. Downloaded in October 5th 2022.

[2] Baltagi, B. H., *Econometric Analysis of Panel Data* (3rd ed.), Chichester, England: John Wiley & Sons Ltd, 2005.

[3] Eurofound. (2012). *NEETs a Young people not in employment, education or training: Characteristics, costs and policy responses in Europe*. European Foundation for the Improvement of Living and Working Conditions. https://doi.org/10.2806/41578

[4] Greene, W. H., *Econometric Analysis (5 ed.)*, New Jersey: Prentice Hall International, 2012.

[5] Gujarati, D.N., *Basic Econometrics*, Tata McGraw-Hill Education, New Delhi, 2009.

[6] ILO. (2017). *Laporan Ketenagakerjaan Indonesia 2017: Memanfaatkan Teknologi untuk Pertumbuhan dan Penciptaan Lapangan Kerja*. Kantor Jakarta.

[7] Misbahuddin, & Hasan, I., *Analisis Data Penelitian dengan Statistik Edisi Ke-2*, Jakarta: Bumi Aksara, 2013.

[8] Liana, L., (2009). *Liana Lie Mra dengan SPSS*. XIV(2), 90-97.

[9] Maguire, S., Cockx, B., Dolado, J. J., Felgueroso, F., Jansen, M., StyczyÅska, I., Rinne, U. (2013). Youth unemployment, *Intereconomics*, 48(4), 196-235. https://doi.org/10.1007/s10272-013-0464-9

[10] Sirusa Badan Pusat Statistik. https://sirusa.bps.go.id/sirusa/. Access in October 20th 2022

[11] UNICEF. (2013). *NEET youth dynamics in Indonesia and Brazil: A cohort analysis, March 2013, Understanding Childrens Work Programme Working Paper Series*, (March), 1-22.

[12] Widarjono, A., *Ekonometrika*, Yogyakarta: UPP STIM YKPN, 2013.

[13] Worldometers Indonesia. 2020. *Indonesia Population*. Downloaded in October 1st 2022, from https://www.worldometers.info/world-population/indonesiapopulation