SCIENOVA

Modeling The Increase In Indonesian Migrant Workers (PMI) In Indonesia In 2024 Using the Multiple Linear Regression Method

Yusril Saputra[1], Salsabila Nur Athifah[1], Nurhayati[1], Muhammad Alfayed[1]

[1] Universitas Negeri Makassar
* Correspondence: yusrilsaputra010505@gmail.com

Received: Feb 1, 2026
Revised: Mar 13, 2026
Accepted: May 20, 2026
Published: Jun 30, 2026

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Abstract

Indonesian Migrant Workers (PMI) play an important role in improving household income and supporting the national economy through remittances. This study investigates the effects of the open unemployment rate and the number of poor people on the number of Indonesian Migrant Workers in Indonesia using multiple linear regression. Secondary data for 2024 were obtained from Statistics Indonesia (BPS) and the Ministry of Manpower. The regression analysis showed that the number of poor people had a significant positive effect on the number of Indonesian Migrant Workers (β = 7.3563, p < 0.001), whereas the open unemployment rate was not statistically significant (β = −661.5751, p = 0.0856). The overall regression model was statistically significant (F = 116.8, p < 0.001) and explained 88.95% of the variation in the number of Indonesian Migrant Workers (R^2 = 0.8895). These findings indicate that poverty remains the primary economic driver of labor migration in Indonesia. The results provide evidence for policymakers to prioritize poverty alleviation, employment creation, and workforce development programs while strengthening migrant worker protection policies to address the socioeconomic factors encouraging overseas migration.

Keywords: Indonesian Migrant Workers, Open Unemployment Rate, Number of Poor People, Mul-tiple Linear Regression

1. Introduction

Indonesian Migrant Workers (PMI) is the term used for Indonesian citizens who will, are currently, or have been working and receiving wages outside the territory of Indonesia. It was recorded that as many as 708,141 Indonesian citizens were working abroad [1]. There are several benefits obtained from international migration. One of them is that income abroad is higher than in the country of origin. Thus, migrants can achieve a better standard of living. In addition, they may gain access to better education, social services, and facilities, and in some countries these benefits are even provided free of charge [2]. Therefore, Indonesian citizens prefer to work and settle abroad.

At the beginning of 2025, Indonesian society was faced with social, economic, and political conditions that were unstable. In addition, the emergence of news regarding unethical behavior carried out by government officials or state institutions, which was considered detrimental to the public, made Indonesian people feel uncomfortable

with their own country. The public perceived indications of problems in the country's finances, which would ultimately affect society, particularly the lower-middle class. This was reinforced by the phenomenon of the declining middleclass population [3]. These conditions encouraged people to seek employment abroad in pursuit of a better quality of life. Communities encouraged one another to engage in international migration and promoted the hashtag “just run away first” across various social media platforms [4].

To identify the reasons behind this international migration trend, statistical analysis is needed to provide a scientific explanation of the factors influencing the growth of PMI. One factor strongly suspected of influencing PMI is poverty. This is in line with research conducted by Afrina Juliana, Akung Daeng, and Satarudin entitled “The Effect of Education, Poverty, and Unemployment Rate on the Number of Indonesian Migrant Workers (PMI) Abroad in West Nusa Tenggara Province.” The study explained that poverty was one of the significant factors affecting PMI in West Nusa Tenggara Province using the multiple linear regression analysis method.

Previous studies found that poverty had a significant effect on the increase of PMI in West Nusa Tenggara Province. Therefore, this study analyzes the effect of poverty on PMI in Indonesia and incorporates the Open Unemployment Rate as an additional factor that is suspected to influence PMI. Using the multiple linear regression method, the dependent variable is Indonesian Migrant Workers (Y), while Poverty (X_1) and Open Unemployment Rate (X_2) serve as explanatory variables. The data period used is 2024, with provinces in Indonesia serving as samples. Data were obtained from Statistics Indonesia (BPS) and related institutions.

Although previous studies have examined the determinants of Indonesian Migrant Workers (PMI), most have focused on specific provinces or limited regional contexts, making it difficult to generalize their findings to the national level. Furthermore, relatively few studies have simultaneously examined the effects of poverty and the open unemployment rate using the most recent provincial data from across Indonesia. This study addresses this research gap by analyzing provincial-level data from 2024 using a multiple linear regression model, thereby providing updated empirical evidence on the socioeconomic determinants of labor migration in Indonesia.

Regression analysis is a statistical method that explains the relationship between two variables, namely the response variable (Y) and the explanatory variable (X). Regression analysis is divided into two types: simple linear analysis and multiple linear analysis. Simple linear regression is an analytical technique that discusses the relationship between a response variable and an explanatory variable, whereas multiple linear regression analysis is a regression model that contains more than one explanatory variable. One of the advantages of multiple linear regression analysis is its ability to predict future conditions through the measurement of several explanatory variables (X) and a response variable (Y).

According to [5], poverty is a condition in which an individual is unable to fulfill minimum basic needs such as food, clothing, health care, housing, and education. These minimum basic needs are translated into financial resources in the form of money. The value of these minimum requirements is known as the poverty line. A person is considered poor if their income falls below the poverty line. Poverty is viewed as an individual's economic inability to meet basic needs and other necessities, measured from the expenditure side. A population is classified as poor when its average monthly expenditure is below the poverty line [6].

According to [7], open unemployment is created due to a substantial increase in the labor force while available employment opportunities remain limited.

Furthermore, according to [8], open unemployment refers to individuals who do not have a job and are actively seeking employment, as well as those who already have a job but have not yet started working.

This study aims to determine whether poverty and the open unemployment rate have a significant effect on PMI, both partially and simultaneously. It is expected that this study will provide valid and reliable information through the application of multiple linear regression analysis.

The results of this study are expected to be beneficial to the general public from several perspectives. From a social perspective, this study provides insights for the government regarding the economic factors influencing the increase in the number of PMI so that more effective policies can be formulated. It also provides empirical data that can be utilized by academics for further research related to labor migration and employment policies. Furthermore, it assists policymakers and relevant institutions in understanding labor migration trends and their impact on the national economy.

2. Literature Review

2.1. Poverty

According to [5], poverty is a condition in which an individual is unable to fulfill basic needs, including food, clothing, health care, housing, and education. These basic needs are measured in terms of financial value, known as the poverty line. The poverty line serves as an important indicator in determining whether an individual belongs to the poor population or not. If an individual's income falls below this threshold, that person is categorized as poor because they do not have sufficient resources to meet their daily needs.

Poverty is not only viewed from the income perspective but is also measured based on expenditure. According to [6], an individual is considered poor if their average monthly expenditure for living necessities falls below the poverty line. This means that economic inability is reflected in an individual's limited purchasing power to meet basic and other essential needs. Therefore, expenditure is one of the important indicators in identifying the level of community welfare.

2.2. Open Unemployment

Stated that open unemployment occurs when the number of available workers increases but is not accompanied by a corresponding increase in employment opportunities [9]. This imbalance causes many individuals who are ready to work to remain unabsorbed by the labor market. This phenomenon reflects a structural problem in the economy, where the growth of the economic sector is not sufficiently rapid to create adequate new employment opportunities.

According to [8], open unemployment is defined as individuals who do not have a job and are actively seeking employment. In addition, it also includes individuals who have already obtained a job but have not yet started working. This condition indicates that unemployment not only includes those who are completely unemployed but also those who are in the transition process toward entering the workforce. Open unemployment serves as an important indicator for assessing labor market conditions in a country.

The testing method employed in this study is Multiple Linear Regression. In order for the test results to be properly interpreted, the displayed results must be unbiased.

2.3. Multiple Linear Regression

Regression is a measurement tool that can also be used to determine the presence or absence of correlation among variables. If we have two or more variables, it is appropriate to study how these variables are related or can be predicted [10].

Regression analysis is a statistical method that explains the relationship between two variables, namely the response variable (Y) and the explanatory variable (X). Regression is generally divided into two types, namely simple linear regression analysis and multiple linear regression analysis. Simple linear regression is an analytical technique that examines the relationship between a dependent variable and an independent variable, whereas multiple linear regression analysis is a regression model that includes more than one independent variable [11].

The multiple linear regression model is formulated as follows [12]:

$$\mathrm{Y} = \beta_0 + \beta_1 \mathrm{X}_1 + \beta_2 \mathrm{X}_2 + \cdots + \beta_k \mathrm{X}_k + e$$

Where:

Y = The value of the response variable (to be modeled)
X_1 = The value of the first explanatory variable
X_2 = The value of the second explanatory variable
X_3 = The value of the third explanatory variable
X_k = The value of the k-th explanatory variable
β_0 = The ordinate of the point where the straight line intersects the Y-axis.
β_1 = Regression coefficient for X_1
β_2 = Regression coefficient for X_2
β_3 = Regression coefficient for X_3
β_k = Regression coefficient for X_k
e = Residual

The multiple linear regression (straight-line) technique is used when we want to analyze the effect of or predict k explanatory variables, namely $X_1, X_2, \cdots, X_k$ with one response variable, namely Y. To calculate $b_0, b_1, \cdots, b_k$, the following normal equations are obtained:

$$\begin{aligned}
&\mathrm{b0n} &&+ \mathrm{b1}\textstyle\sum \mathrm{X1} &&+ \mathrm{b2}\textstyle\sum \mathrm{X2} &&+ \ldots \mathrm{bk}\textstyle\sum \mathrm{Xk} &&= \textstyle\sum \mathrm{Y} \\
&\mathrm{b_0}\textstyle\sum \mathrm{X_1} &&+ \mathrm{b_1}\textstyle\sum \mathrm{X_1^2} &&+ \mathrm{b_2}\textstyle\sum \mathrm{X_1X_2} &&+ \ldots \mathrm{b_k}\textstyle\sum \mathrm{X1X_k} &&= \textstyle\sum \mathrm{X1Y} \\
&\mathrm{b_0}\textstyle\sum \mathrm{X_2} &&+ \mathrm{b_1}\textstyle\sum \mathrm{X_2X_1} &&+ \mathrm{b_2}\textstyle\sum \mathrm{X_2^2} &&+ \ldots \mathrm{b_k}\textstyle\sum \mathrm{X_2X_k} &&= \textstyle\sum \mathrm{X_2Y} \\
&\quad . &&\quad . &&\quad . &&\quad . &&\quad . \\
&\quad . &&\quad . &&\quad . &&\quad . &&\quad . \\
&\quad . &&\quad . &&\quad . &&\quad . &&\quad . \\
&\mathrm{b_0}\textstyle\sum \mathrm{X_k} &&+ \mathrm{b_1}\textstyle\sum \mathrm{X_kX_1} &&+ \mathrm{b_2}\textstyle\sum \mathrm{XkX_2} &&+ \ldots \mathrm{b_k}\textstyle\sum \mathrm{X}^{\,2} &&= \textstyle\sum \mathrm{X_kY}
\end{aligned}$$

For $k = 2, Y = b_0 + b_1 X_1 + b_2 X_2$, one dependent variable (Y), and two independent variables $(X_1$ and $X_2)$ so b_0, b_1 and b_2 are calculated from the following normal equations:

$$\begin{aligned}
&\mathrm{b_0n} &&+ \mathrm{b_1}\textstyle\sum \mathrm{X_1} &&+ \mathrm{b_2}\textstyle\sum \mathrm{X_2} &&= \textstyle\sum \mathrm{Y} \\
&\mathrm{b_0}\textstyle\sum \mathrm{X_1} &&+ \mathrm{b_1}\textstyle\sum \mathrm{X}_1^{\,2} &&+ \mathrm{b_2}\textstyle\sum \mathrm{X_1X_2} &&= \textstyle\sum \mathrm{X_1Y} \\
&\mathrm{b_0}\textstyle\sum \mathrm{X_2} &&+ \mathrm{b_1}\textstyle\sum \mathrm{X_2X_1} &&+ \mathrm{b_2}\textstyle\sum \mathrm{X_2^2} &&= \textstyle\sum \mathrm{X_2Y}
\end{aligned}$$

(There are three equations with three unknown variables, namely b_0, b_1, and b_2).

To determine the values of $b_0, b_1,$ and b_2, the following steps are performed:

$$X^T X = \begin{vmatrix} n & \sum X_1 & \sum X_2 \\ \sum X_1 & \sum_1^2 & \sum X_1 X_2 \\ \sum X_2 & \sum X_1 X_2 & \sum_2^2 \end{vmatrix}$$

$$(X^T X)^{-1} = \frac{1}{\det(X^T X)} \begin{vmatrix} C_{11} & -C_{12} & C_{13} \\ -C_{21} & C_{22} & -C_{23} \\ C_{31} & -C_{32} & C_{33} \end{vmatrix}$$

In determining the cofactors, the following steps are performed:

For C_{11}, "11" indicates the elimination of the first row and the first column of the matrix. The difference between the product of the variables marked in red and those marked in blue represents the value of C_{11}

$$C_{11} = \begin{vmatrix} n & \sum X_1 & \sum X_2 \\ \sum X_1 & \sum X_1^2 & \sum X_1 X_2 \\ \sum X_2 & \sum X_1 X_2 & \sum X_2^2 \end{vmatrix} = (\sum X_1^2 \times \sum X_2^2) - (\sum X_1 X_2 - \sum X_1 X_2)$$

$$\vdots$$

$$C_{33} = \begin{vmatrix} n & \sum X_1 & \sum X_2 \\ \sum X_1 & \sum X_1^2 & \sum X_1 X_2 \\ \sum X_2 & \sum X_1 X_2 & \sum X_2^2 \end{vmatrix} = (n \times \sum X_1^2) - (\sum X_1 - \sum X_1)$$

In determining the determinant, the following steps are performed:

$$\det(X^T X) = \begin{vmatrix} n & \sum X_1 & \sum X_2 & n & \sum X_1 \\ \sum X_1 & \sum X_1^2 & \sum X_1 X_2 & \sum X_1 & \sum X_1^2 \\ \sum X_2 & \sum X_1 X_2 & \sum_2^2 & \sum X_2 & \sum X_1 X_2 \end{vmatrix}$$

$$\det(X^T X) = (\sum X_2 \times \sum X_1^2 \times \sum X_2) - (\sum X_1 X_2 \times \sum X_1 X_2 \times n) - (\sum X_2^2 \times \sum X_1 \times \sum X_1)$$

$$(X^T Y) = \begin{vmatrix} \sum Y \\ \sum X_1 Y \\ \sum X_2 Y \end{vmatrix}$$

$$(X^T X)^{-1} \times (X^T Y) = \begin{vmatrix} b0 \\ b1 \\ b2 \end{vmatrix}$$

2.4. Classical Assumptions of Multiple Linear Regression

Before drawing conclusions from a regression model, classical assumption tests are required to ensure that the results obtained do not lead to misleading interpretations. If classical assumption tests are not conducted, there will be uncertainty as to whether the resulting regression equation provides accurate, unbiased, and consistent estimates [13]. The classical assumption tests for multiple linear regression include:

2.4.1 Asumsi Normalitas

The normality test is intended to determine whether the residual values are normally distributed or not [14]. The parameter tested in the normality test of a regression model is the residual. The residual is formulated as follows:

$$e = Y - \hat{Y}$$

$$e = Y - (b_0 + b_1 X_1 + b_2 X_2)$$

According to [15], the tests used to assess the normality assumption include the Lilliefors test, Shapiro–Wilk test, Skewness and Kurtosis test, D'Agostino–Pearson test, and Jarque–Bera test. In this study, the researcher focuses on testing normality using the Shapiro–Wilk test method.

The normality test was developed by Samuel Shapiro and Martin Wilk in 1965. This test does not require the prior determination of the mean and variance and is valid for testing normality with sample sizes ranging from 3 to 50 observations [16]. The Shapiro–Wilk (SW) test statistic is formulated as follows [17]:

$$SW = \frac{\left[\sum_{i=1}^{k} \mathrm{a}_3 \, (\mathrm{x}_{(n-i+1)} - \mathrm{x}_{(i)}\right]^2}{\sum_{i=1}^{n} (\mathrm{X}_i - \bar{X})^2}$$

Where $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ are the ordered observations that are free from scale and location effects. The distance from the symmetry of the data values around the central value is measured by:

$$(x_{(n-i+1)} - x_{(i)}); \text{ for } i = 1, 2, \ldots, k$$

where $k = \frac{(n+1)}{2}$ for odd n, and $k = \frac{(n)}{2}$ for all other values of n.

Using the Shapiro–Wilk test statistic, the following hypotheses are tested:

$H_0 : \bar{X} = 0$, The data are drawn from a normally distributed population.

$H_1 : \bar{X} \neq 0$, The data are drawn from a non-normally distributed population

The null hypothesis is rejected at the significance level α if the value of the test statistic is less than the critical value given in the Shapiro–Wilk quantile table. The null hypothesis cannot be rejected if the value of the test statistic is greater than the corresponding value in the Shapiro–Wilk percentile table.

2.4.2 Non-Autocorrelation Assumption

Autocorrelation is a condition in which, within a regression model, there is a correlation between the residual at period (t) and the residual at the previous period $(t-1)$. A good regression model is one that does not exhibit autocorrelation [13].

Autocorrelation is a violation of the OLS assumption which states that there is no autocorrelation between the error terms of different observations. Autocorrelation may occur in studies where the order of observations carries meaningful information. Although autocorrelation can be useful in everyday applications, it becomes a major issue that must be addressed in regression analysis using OLS. OLS assumes that the errors are independent random variables (uncorrelated so that the estimators possess the BLUE property). The essence of autocorrelation is that the error in one time period systematically depends on the error in another time period.

According to [18], several factors may cause the occurrence of autocorrelation problems, including:

a) The existence of inertia

One of the prominent characteristics of some economic time series data is inertia, such as national income data, consumer price index data, production data,

employment data, and unemployment data, which exhibit cyclical patterns. In such situations, observations from the previous period and the current period are highly likely to exhibit interdependence.

b) Model Specification Bias Due to Omitted Variables

This condition is caused by the omission of variables that, according to economic theory, play an important role in explaining the dependent variable. When this occurs, the disturbance term (error term), u_i, will reflect a systematic pattern among the disturbance terms themselves, resulting in a situation of autocorrelation among the disturbance terms.

c) The Existence of the Cobweb Phenomenon

The emergence of the cobweb phenomenon primarily occurs in the supply of agricultural commodities. In this sector, the supply response to price changes takes place only after a time lag (*gestation period*).

d) Data Manipulation (*Manipulation of Data*)

In empirical analysis, especially in time series data, data manipulation often occurs. This happens because the required data are not available.

The Durbin–Watson test is used to determine first-order autocorrelation in the error term of a regression equation. The Durbin–Watson test is applicable when its underlying assumptions are satisfied, namely:

1. The regression model must include an intercept and must not pass through the origin. For a two-variable model $Y_t = \beta_0 + \beta_1 x_t + e_t$, it follows $\beta_0 \neq 0$.
2. The residua ε_i is obtained from a first-order autocorrelation process $e_t = \rho e_{t-1} + V_t$. The Durbin–Watson statistic at observation t is given as follows:

$$d = \sum_{t=2}^{n} (e_t - e_{t-1})^2 \,/\, \sum_{t=1}^{n} {e_t}^2$$

When:

d = Durbin–Watson statistic value to be calculated

n = Number of observations or samples in the regression model

e_t= Residual value at time (t)

It represents the ratio of the sum of squared differences between e_t and e_{t-1}. Note that in the numerator of the Durbin–Watson statistic, the number of observations is only $n - 1$.

The hypotheses used in this statistical test are two-tailed hypotheses:

$H_0 : \rho = 0$

$H_1 : \rho \neq 0$

If $d < d_L$ and $d > 4 - d_L$ reject H_0, if $d_U < d < 4 - d_U$ lies within the acceptance region, accept H_0, and no conclusion can be drawn if $d_L \leq d \leq d_U$ or $4 - d_U \leq d \leq 4 - d_L$.

When:

d = Calculated Durbin–Watson value

d_u = Upper bound value of the Durbin–Watson table

d_L = Lower bound value of the Durbin–Watson table

The decision criteria for determining the presence of autocorrelation using the Durbin–Watson test can be seen in the following table:

Table 1. Durbin–Watson Test Criteria

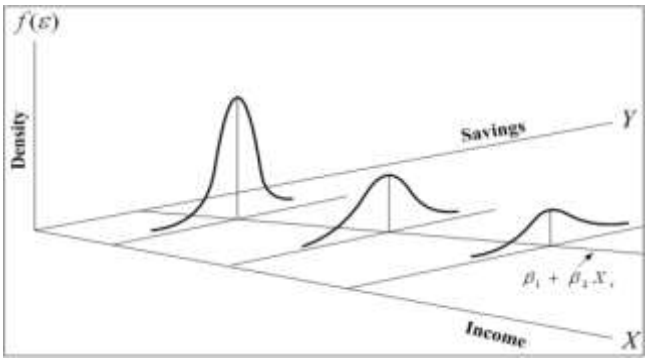
If	Null Hypothesis	Decision
$0 < d < d_i$	Positive autocorrelation occurs	Reject
$d_L \leq d \leq d_u$	No conclusion	None
$4 - d_u < d < 4 - d_L$	No conclusion	None
$4 - d_L < d < 4$	Negative autocorrelation occurs	Reject
$d_u < d < 4 - d_u$	No positive or negative auto-correlation	Accept

The steps of the autocorrelation test using the Durbin–Watson test are as follows:

1. Determining the hypotheses
   H_0 = no autocorrelation occurs
   H_1 = autocorrelation occurs
2. Determining the multiple linear regression equation using the formula:
$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_k X_K$$
3. Determining the residual value using the formula:
$$e_t = Y_i - \hat{Y}_i$$
4. Calculating the Durbin–Watson value
5. Determining the critical Durbin–Watson values, where d_L is the lower bound and d_u is the upper bound, based on the number of observations (n), and the number of parameters (k).
6. Drawing conclusions:
   If $d < d_L$ then H_0 is rejected, indicating positive autocorrelation. If $d > 4 - d_L$ then H_0 is rejected, indicating negative autocorrelation. If $d_u < d < 4 - d_u$ then H_0 is accepted, indicating no autocorrelation. If $d_L < d < d_u$ or $4 - d_u < d < 4 - d_L$ then no conclusion can be made.

2.4.3 Heteroscedasticity Assumption

Heteroscedasticity is a condition in which there is inequality of the variance of the residuals across all observations in a regression model [13]. The heteroscedasticity test is used to determine the inequality of the variance of the residuals from each observation in the regression model. If there are differences in the variance of the residuals in the regression model or if they are not homogeneous, this condition is called heteroscedasticity. Heteroscedasticity is one of the factors that causes a regression model to be inefficient and inaccurate, and it results in the maximum likelihood estimation method for estimating the parameters of the regression model becoming inaccurate. Therefore, a regression model is considered good if it does not contain heteroscedasticity or exhibits homoscedasticity.



Homoscedasticity is one of the classical assumption requirements in a regression model. The homoscedasticity test is used to examine whether the error term in a linear regression model has constant variance (homogeneous) across all observations. The meaning of homogeneous variance is that the error term has the same variance value. Mathematically, this can be written as $\sigma_{\epsilon_i}^2 = \sigma_{\epsilon_j}^2 = \sigma_{\epsilon}^2$. The error term is also a form of data, but its value is difficult to observe directly. Therefore, residual data are used as an estimator of the error term. Thus, if the variance of the residuals is homogeneous, then the error term also has homogeneous variance.

The following are the causes of heteroscedasticity and homoscedasticity:

1. The effect of the learning curve, where increasing experience leads to smaller behavioral errors, resulting in decreasing variance. This applies to time series data.
2. Economic growth, where increasing economic conditions lead to more diverse income, increasing alternative expenditures and thus increasing variance. This applies to time series data.
3. Improvement in data collection techniques, where better data techniques can lead to decreasing variance.
4. Model specification errors:
    a. Model specification errors due to omitting important variables in the regression model.
    b. Data transformation errors, for example ratio/first difference transformations, and functional form errors such as linear vs log-linear models.

The method that can be used is the Glejser test. This method is carried out by regressing the independent variables on the absolute value of the residuals. If there is a significant effect between the independent variables and the absolute residuals, it can be concluded that heteroscedasticity occurs. The equation used in this method is as follows.

$$|u_i| = a + \beta X_1 + v_i$$

Where:
$|u_i|$= absolute residual value
a = significance level
β = regression coefficient
X_1 = independent variable
v_i = error value

Steps of heteroscedasticity test using the Breusch–Pagan test:

1. Formulating hypotheses:
   H_0 = no heteroscedasticity occurs
   H_1 = heteroscedasticity occurs
2. Estimating the regression model using the Ordinary Least Squares (OLS) method:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \varepsilon_i$$

   Where:
   Y_i = dependent variable at observation (i)
   β_0 = intercept of the regression line
   $\beta_{1,2,k}$ = regression coefficients
   $X_{1,2,ki}$ = independent variables at observation (i)
   ε_i = residual at observation (i)
3. Estimating σ^2 using residuals from the regression model:

$$\sigma^2 = \frac{\sum \varepsilon_i^2}{n}$$

   Where:
   σ^2 = Variance
   $\sum \varepsilon_i^2$ = sum of squared residuals
4. Constructing a new variable W_i, which will be used in the auxiliary regression:

$$W_i = \frac{\varepsilon_i^2}{\sigma^2}$$

   Where:
   W_i = auxiliary regression variable
   ε_i^2 = squared residual at observation (i)
   σ^2 = Variance
5. Running the auxiliary regression, where W_i is regressed on the independent variables:

$$W_i = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \ldots + \alpha_k X_{ki} + v_i$$

   Where:
   W_i = auxiliary regression variable
   α_0 = intercept of auxiliary regression
   $\alpha_{1,2,k}$ = coefficients of auxiliary regression
   $X_{1,2,ki}$ = independent variables in auxiliary regression
   v_i = residual of auxiliary regression
6. Calculating the coefficient of determination R^2 from the auxiliary regression by determining SSR and SST, then multiplying the result by the sample size.

2.4.4 Multicollinearity Assumption

Multicollinearity is a condition in which there is a perfect or near-perfect linear relationship among the independent variables in a regression model. A regression model is said to experience multicollinearity if there is a perfect linear function among some or all independent variables in the linear function [13].

The detection of the presence or absence of multicollinearity in a regression model is as follows:

a. The R-squared value produced by an empirical regression model estimation is very high, but individually many independent variables are not statistically significant in influencing the dependent variable.
b. Analyzing the correlation matrix of the independent variables. If the Tolerance value and its reciprocal (VIF) indicate multicollinearity. In this case, a good tolerance value is < 0, 10.

Variance Inflation Factor (VIF). In this case, a good VIF value is > 10.

Tolerance measures the variability of a selected independent variable that is not explained by other independent variables. A low tolerance value is equivalent to a high VIF value. The VIF value can be obtained using the following formula:

$$VIF = \frac{1}{TOL}$$

The tolerance formula is as follows:

$$TOL = 1 - r^2 =$$

The VIF value is then used to test the following hypotheses:

H_0: $\beta_i = \beta_j$, Multicollinearity occurs

H_0: $\beta_i \neq \beta_j$, No multicollinearity occurs

The decision criteria are as follows:

H_0 diterima jika nilai VIF > 10

H_0 ditolak jika nilai VIF < 10

Steps of the multicollinearity test using Variance Inflation Factor (VIF):

1. Formulating hypotheses
   H_0 = multicollinearity occurs
   H_1 = no multicollinearity occurs
2. Performing multiple linear regression analysis with the dependent and independent variables to be tested for multicollinearity
3. Setting the VIF threshold value, which is 10
4. Calculating the VIF value for each independent variable
5. Drawing conclusions:
   a. If the VIF value of an independent variable > 10, then H_0 is accepted, indicating multicollinearity occurs
   b. If the VIF value of an independent variable < 10, then H_0 is rejected, indicating no multicollinearity occurs

2.5. Parameter Significance Test

2.5.1. Partial Test

One of the tools in parametric statistics is the partial test, developed by William Sealy Gosset and classified under inferential statistical analysis. The partial test is a statistical test used to examine the validity or falsity of a hypothesis stating that between two sample means randomly drawn from the same population, there is no significant difference.

The partial test (*t*-test) is conducted to test research hypotheses regarding the effect of each independent variable individually on the dependent variable. The t-statistic is a value used to determine the level of significance in hypothesis testing by obtaining the t-statistic value through a bootstrapping procedure (a statistical

estimation method). Based on the sample size, the partial test (*t*-test) is divided into two procedures, namely:

1. One-sample *t*-test (one sample *t*-test)

The one-sample *t*-test is a statistical method used to test whether the mean of a single sample differs significantly from a known or hypothesized population mean. This test is generally used when you have one sample of data and want to compare the sample mean with a specific value or theoretical expectation. The specific value is generally a parameter value used to measure a population.

2. Paired-sample *t*-test (paired *t*-test)

The paired t-test is a statistical method used to compare the means of two related samples taken from the same subjects. The paired sample *t*-test is an analysis involving two measurements on the same subject under a certain influence or treatment. If the treatment has no effect, then the mean difference is zero.

The application of the partial test is the process of testing the significance of each independent variable separately on the dependent variable in a statistical model, such as linear regression. The purpose of the partial test is to determine whether each independent variable has a significant effect on the dependent variable when the other variables are held constant. The general steps in applying the partial test include:

1. Determining an appropriate statistical model for the analysis, such as simple or multiple linear regression.
2. Estimating the regression coefficients for each independent variable in the model.
3. Conducting hypothesis testing on each regression coefficient to determine whether the variable significantly contributes to the dependent variable.
4. Interpreting the results of the partial test to understand how strong the effect of each independent variable is on the dependent variable.

The use of the partial test is important to understand the relative contribution of each independent variable in explaining the variability of the dependent variable in a statistical model. Decision-making is carried out by examining the significance value in the Coefficients table. Generally, regression results are tested using a confidence level of 95% or a significance level of 5% (α = 0.05). The steps of the partial test are as follows:

1. Formulating Hypotheses

   A hypothesis is a statement that is temporarily accepted and still needs to be tested. A hypothesis is stated as a provisional truth and serves as the basis and guideline in data analysis. The hypotheses in the partial test are as follows:

   $H_0: \beta_i = 0$

   $H_1: \beta_i \neq 0$

   $(i = 1, 2)$

   If $H_0: \beta_i = 0$ is accepted, then there is no effect between the independent variable and the dependent variable. If $H_1: \beta_i \neq 0$ is rejected, it means there is an effect between the independent variable and the dependent variable.

2. Table Determining the Significance Level and t-table Value

   The significance level is the threshold of tolerance in accepting the error of hypothesis testing against the population parameter value. The significance level can be determined by the researcher; commonly used levels are 1%, 5%, and 10%.

The formula for the t-table used is as follows:

$$t_{tabel\left(\frac{\alpha}{2}\right)(n-(k+1))}$$

Where :
a = significance level alpha
n = number of observations
k = number of independent variables

3. Calculating the *t*-statistic

The formula for calculating the t-statistic is as follows:

$$t_{hitung} = \frac{\beta_1}{s \times \sqrt{C_{11}}}$$

Where:
β_1 = regression coefficient of the i-th independent variable $(i = 1, 2)$
S = standard error, which can be calculated using the formula:

$$S = \sqrt{\frac{JKG}{(n-(k+1))}}$$

Where :
JKG = Sum of Squares of Error, which can be calculated using the formula:

$$JKG = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2$$

Further Where:
n = number of observations
k = number of independent variables
C_{ii} = diagonal element of covariance matrix C_{ii} from $(X'X)^{-1}$ $(i = 1, 2)$
Where:

$$C_{ii} = \begin{bmatrix} C_{00} & C_{01} & C_{02} \\ C_{10} & C_{11} & C_{12} \\ C_{20} & C_{21} & C_{22} \end{bmatrix}$$

4. Making a Decision

The following are the decision-making criteria in the partial *t*-test. There are two approaches that can be used as the basis for decision-making: first, by observing the significance value (Sig), and second, by comparing the t-statistic value with the *t*-table value.

1) If the significance value of t > 0.05 or the *t*-statistic value < *t*-table value, then H_0 is accepted. This means that the independent variable does not have a significant effect on the dependent variable.
2) If the significance value of t < 0.05 or the *t*-statistic value > *t*-table value, then H_0 is rejected. This means that the independent variable has a significant effect on the dependent variable.

Steps of the partial test are as follows:

1. Formulating hypotheses:
   H_0 = the independent variable has no significant effect on the dependent variable

H_1 = the independent variable has a significant effect on the dependent variable
2. Setting the significance level, where $\alpha = 0.05$
3. Calculating the t-test statistic value
4. Determining the critical value from the t-distribution table with $df = n - k - 1$.
5. Drawing conclusions:
   If $t_{hitung} > t_{tabel}$ then H_0 is rejected, meaning the independent variable has a significant effect on the dependent variable. Conversely, if $t_{hitung} < t_{tabel}$ then H_0 is accepted, meaning the independent variable does not have a significant effect on the dependent variable.

2.5.2. Simultaneous Test

The *F*-test (simultaneous test) is used to determine whether the independent variables, jointly (simultaneously), have a significant effect on the dependent variable. In the simultaneous test, the effect of the independent variables is tested together on the dependent variable. The test statistic used in the simultaneous test is the *F*-test, with the following formula:

$$f - value = \frac{JKR/\,k}{JKG/(n-k-1)}$$

Where:
JKR : Regression Sum of Squares
JKG : Error Sum of Squares
n : Number of samples
k : Number of independent variables

For testing the simultaneous effect, the hypotheses are as follows:

H_0: $\beta = 0$, the explanatory variables jointly do not have a significant effect on the response variable, versus

H_1: $\beta \neq 0$, the explanatory variables jointly have a significant effect on the response variable.

The decision criteria for the F-test are:

1. H_0 : is rejected if $F_{hitung} > F_{tabel}$
2. H_1: is accepted if $F_{hitung} < F_{tabel}$

The rejection of H_0 indicates that the independent variables jointly have a significant effect on the dependent variable. Conversely, the acceptance of H_0 indicates that the independent variables do not jointly have a significant effect on the dependent variable [25].

Steps of the simultaneous test are as follows:

1. Formulating hypotheses:
   H_0 = the independent variables jointly affect the dependent variable
   H_1 = the independent variables jointly do not affect the dependent variable
2. Setting the significance level, where $\alpha = 0.05$
3. Calculating the *F*-test statistic
4. Determining the critical value from the F distribution table with degrees of freedom: numerator $= k - 1$ and denominator $= n - k$, where n is the number of observations
5. Drawing conclusions:
   If the *F*-statistic < critical value, then H_0 is accepted, meaning the independent variables do not jointly have a significant effect on the dependent variable.

Conversely, if the *F*-statistic > critical value, then H_0 is rejected, meaning the independent variables jointly have a significant effect on the dependent variable.

2.6. Model Reliability Test

The coefficient of determination test (R^2) is used to measure how well the model is able to explain the variation of the dependent variable (Y). The value of the coefficient of determination ranges between zero and one. If the value of R^2 is small, it indicates that the ability of the independent variables (X) to explain the variation of the dependent variable (Y) is very limited. Conversely, if the value of R^2 is close to one, it indicates that the independent variables (X) provide almost all the information needed to predict the variation of the dependent variable (Y) [27].

The test statistic used is:

$$R^2 = \frac{JKR}{JKT}$$

Where:

R^2 = coefficient of determination

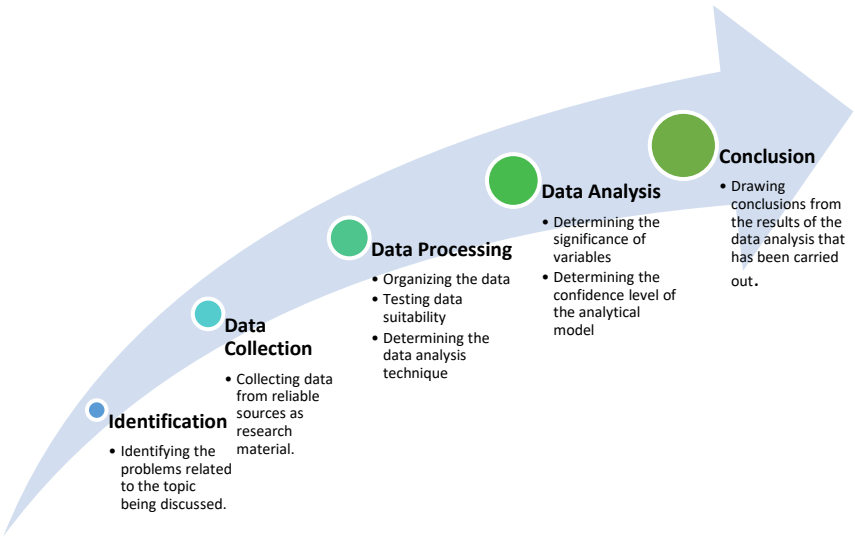
JKR = regression sum of squares, with the formula:

$\text{JKR} = \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2$

JKT = total sum of squares, with the formula:

$$\text{JKT} = \sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

3. Research Methodology



3.1. Research Design

The method used in this study is a secondary data analysis approach. [5] explains that the quantitative secondary data analysis approach is a research method that utilizes data that are already available from various data-providing institutions.

3.2. Research Time

This study was conducted for approximately four months, starting from February 2025 to May 2025.

3.3. Population and Sample of the Study

The stages of the study begin with determining the data that serve as the object of the research from reliable sources. The data were obtained from the Central Bureau of Statistics (BPS) Indonesia and the Ministry of Manpower of the Republic of Indonesia. The data used in this study include Indonesian Migrant Workers (PMI) in Indonesia in 2024, the Open Unemployment Rate in Indonesia in 2024, and the number of poor people in Indonesia in 2024. The sample used consists of 32 provinces in Indonesia, which include Aceh, North Sumatra, West Sumatra, Riau, Jambi, South Sumatra, Bengkulu, Lampung, DKI Jakarta, West Java, Central Java, DI Yogyakarta, East Java, Banten, Bali, West Nusa Tenggara, East Nusa Tenggara, West Kalimantan, Central Kalimantan, South Kalimantan, East Kalimantan, North Kalimantan, North Sulawesi, Central Sulawesi, South Sulawesi, Southeast Sulawesi, Gorontalo, West Sulawesi, Maluku, West Papua, and Papua. This sample is used to draw conclusions, with Indonesia as the population in this study.

3.4. Operational Definition of Variables

In this study, as a basis for analysis, the researcher selects several key indicators that are considered to influence the increase in the number of Indonesian Migrant Workers as the dependent variable.

Table 2. Operational Definition of Variables

Variable Type	Variable Name	Year	Information
Response	PMI (Y)	2024	PMI stands for Indonesian Migrant Workers. The purpose of this study is to identify the factors that influence the growth of PMI.
Description	Open Unemployment Rate (X_1)	2024	Open unemployment is a condition in which a person is unable to work for certain reasons. This variable is believed by the researcher to influence the growth of PMI.

Description	Number of Poor Population (X_2)	2024	Poverty is a condition in which a person has difficulty or is unable to meet the basic needs required for survival. The growth of poverty is considered to have an influence on the growth of PMI.

3.5. Data Analysis Technique

In this study, to examine the relationship between poverty and unemployment on migrant workers in Indonesia, the analysis used is regression analysis. Regression analysis has long been developed to study patterns and measure statistical relationships between two or more variables [18]. In regression analysis, a regression equation is developed, which is defined as a formula used to estimate the value of the dependent variable based on known independent variables. The data analysis in this study was conducted using RStudio software. Before drawing conclusions from a regression model, classical assumption tests are required to ensure that the results do not lead to misleading interpretations. [13] explains that without classical assumption testing, there will be uncertainty regarding whether the regression equation obtained is accurate in estimation, unbiased, and consistent.

3.5.1 Classical Assumption Tests

The classical assumption tests for multiple linear regression include:

1. Normality Assumption

   The normality test was developed by Samuel Shapiro and Martin Wilk in 1965. This test does not require prior determination of the mean and variance and is valid for testing normality with sample sizes between 3 and 50.

2. Non-Autocorrelation Assumption

   Autocorrelation is a violation of the OLS assumption stating that there is no autocorrelation between error terms in different observations. Autocorrelation may occur in studies where the order of observations is meaningful. The Durbin–Watson test is used to detect first-order autocorrelation in the error term of a regression model.

3. Heteroscedasticity Assumption

   Heteroscedasticity is one of the factors that causes a regression model to become inefficient and inaccurate, and it leads to the maximum likelihood estimation method producing unreliable parameter estimates. Therefore, a regression model is considered good if it does not contain heteroscedasticity or if it exhibits homoscedasticity. The heteroscedasticity test can be conducted using the Breusch–Pagan test. The model is considered free from heteroscedasticity if the Breusch–Pagan test results show a p-value greater than 0.05.

4. Multicollinearity Assumption

   Multicollinearity can be examined using the Variance Inflation Factor (VIF). This measure shows which independent variables can be explained by other independent

variables, meaning that each independent variable becomes a dependent variable and is regressed on the other independent variables. If the VIF value is not greater than 10, then multicollinearity does not occur.

3.5.2 Parameter Significance Test

After the Classical Assumption Tests are satisfied, the next step is to conduct the Parameter Significance Test.

1. Partial Test (*t*-test)

   One of the tools in parametric statistics is the partial test, developed by William Sealy Gosset and included in inferential statistical analysis. The partial test is a statistical test used to examine the validity or falsity of a hypothesis stating that between two sample means randomly drawn from the same population, there is no significant difference.

2. Simultaneous Test (*F*-test)

   The *F*-test (simultaneous test) is used to determine whether the independent variables, jointly (simultaneously), have a significant effect on the dependent variable. In the simultaneous test, the effect of the independent variables is tested together on the dependent variable.

3.5.3 Model Reliability Test

To measure how well the model is able to explain the variation of the dependent variable (Y). If the value of R^2 is small, it indicates that the ability of the independent variables (X) to explain the variation of the dependent variable (Y) is very limited. Conversely, if the value of R^2 is close to one, it indicates that the independent variables provide almost all the information needed to predict the variation of the dependent variable (Y).

4. Results and Discussion

4.1 Classical Assumption Tests

4.1.1 Normality Assumption

Using the R programming language, the following is the output of the Shapiro–Wilk test for normality:

Table 3. Normality Test Results

Shapiro-Wilk	***p-value***
0.9487	0.1324

Since the *p-value* $= 0.1324 > 0.05$, we fail to reject H_0, meaning that the residuals are normally distributed. Therefore, the normality assumption is satisfied. Although the normality assumption is met, other assumptions are still required to ensure reliable regression inference.

4.1.2 Non-Autocorrelation Assumption

Using the R programming language, the following is the output of the Durbin–Watson test for autocorrelation:

Table 4. Non-Autocorrelation Test Results

DW	***p-value***
2.343568	0.318

Since the *p-value* = 0.318, we fail to reject H_0, indicating that there is no autocorrelation in the model. Thus, the non-autocorrelation assumption is satisfied. With normality and non-autocorrelation assumptions fulfilled, the next test is heteroscedasticity.

4.1.3 Heteroscedasticity Assumption

Using the R programming language, the following is the output of the Breusch–Pagan test:

Table 5. Heteroscedasticity Test Results

BP	***p-value***
4.7162	0.0946

Since the *p-value* = 0.0946 > 0.05, we fail to reject H_0, indicating that the variance of the error terms is homoscedastic (no heteroscedasticity). Therefore, the heteroscedasticity assumption is satisfied. With normality, non-autocorrelation, and homoscedasticity assumptions fulfilled, the next step is multicollinearity testing.

4.1.4 Multicollinearity Assumption

Using the R programming language, the following is the output of the VIF test:

Table 6. Multicollinearity Test Results

X_1	X_2
1.053562	1.053562

Since all VIF values are < 10, we fail to reject H_0 (VIF > 10 indicates multicollinearity). Therefore, there is no multicollinearity in the model, and the multicollinearity assumption is satisfied.

4.2 Significance Test

4.2.1 Partial Test

Table 7. Partial Test Results

	Estimate	***t value***	**Pr(>|t|)**
(*intercept*)	1037,0895	0,624	0,5372
X_1	-661,5751	-1,780	0,0856
X_2	7,3563	15,188	2,44e-15

The regression results indicate that only X_2 (Poverty Rate / Number of Poor People) has a statistically significant effect in the model, since its *p-value* < 0.05, meaning that H_0 is rejected.

4.2.2 Simultaneous Test

Table 8. Simultaneous Test Results

F-statistic	**DF**	***p-value***
116,8	29	1.34E-14

The *p-value* < 0.05 indicates that H_0 is rejected. This means that the parameters in the model jointly have a significant effect on the dependent variable.

4.3 Model Reliability Test

Table 9. Model Reliability Results

Multiple R-Squared	*Adjusted R-Squared*
0.8895	0.8819

The multiple linear regression model has a Multiple R-Squared value of 0.8895, indicating that approximately 88.95% of the variation in the dependent variable can be explained by the independent variables in the model. The Adjusted R-Squared value of 0.8819 indicates that after adjusting for the number of predictors used, the model still explains 88.19% of the data variation. Since the difference between R-Squared and Adjusted R-Squared is small, it indicates that the variables in the model are relevant and the model does not suffer from overfitting.

The relatively high coefficient of determination ($R^2 = 0.8895$) suggests that the selected explanatory variables capture a substantial proportion of the provincial variation in the number of Indonesian Migrant Workers. This may be explained by the fact that poverty and labor market conditions are among the primary socioeconomic factors influencing individuals' decisions to seek employment abroad. Provinces with larger poor populations generally experience greater economic pressure, increasing the likelihood of labor migration as a strategy to improve household income. Furthermore, the use of official provincial-level data from the same observation period (2024) reduces temporal inconsistencies and contributes to the strong explanatory performance of the regression model.

4.4 Multiple Linear Regression Model

From Table 7, the multiple linear regression model obtained is as follows:

$$\mathrm{Y} = 1037{,}0895 - 661{,}5751(\mathrm{X}_1) + 7{,}3563\ (\mathrm{X}_2)$$

1. Constant (1037.0895): If X_1 and X_2 are equal to 0, then the predicted value of Y is 1037.0895.
2. Coefficient of X_1 (-661.5751): Every one-unit increase in X_1 will decrease Y by 661.5751, assuming X_2 is constant. This indicates a negative relationship between X_1 and Y.
3. Coefficient of X_2 (7.3563): Every one-unit increase in X_2 will increase Y by 7.3563, assuming X_1 is constant. This indicates a positive relationship between X_2 and Y.

5. Conclusion

Based on the results of the multiple linear regression analysis conducted, it can be concluded that the regression model satisfies all classical assumption tests, namely normality, non-autocorrelation, heteroscedasticity, and multicollinearity assumptions. The Shapiro–Wilk test shows that the residuals are normally distributed, the Durbin–Watson test indicates no autocorrelation, the Breusch–Pagan test confirms the absence of heteroscedasticity, and the VIF values < 10 indicate that there is no multicollinearity among the independent variables. n the significance test, it is found that variable X_2 (number of poor people) has a significant partial effect on the dependent variable, while X_1 is not statistically significant. However, simultaneously, both variables have a significant effect on the model, as indicated by the *F-statistic p-value* being less than 0.05. The model also shows high reliability with a Multiple R^2 value of 0.8895 and an Adjusted R^2 value of 0.8819, meaning that approximately 88.95% of the variation

in the dependent variable can be explained by the model. The small difference between these values indicates that the model does not suffer from overfitting. Thus, the regression model $Y = 1037{,}0895 - 661{,}5751X_1 + 7{,}3563\,X_2$ can be used effectively to explain the relationship between the independent and dependent variables in this study.

Author Contributions: Conceptualization, Y.S. and S.N. A.; methodology, N.N.; software, M.M.; validation, Y.S., N.N. and M.A.; formal analysis, Y.S.; investigation, S.N.A; resources, N.N.; data curation, M.A.; writing—original draft preparation, S.N.A; writing—review and editing, M.A.; visualization, Y.S.
Funding: This research received no external funding
Data Availability Statement: We encourage all authors of articles published in SCIENOVA journals to share their research data. In this section, please provide details regarding where data supporting reported results can be found, including links to publicly archived datasets analyzed or generated during the study.
Conflicts of Interest: The authors declare no conflicts of interest.

References

[1] Apriyani, W. A., & Sari, A. (n.d.). *Analisis faktor penyebab siswa lulusan SMA tidak melanjutkan ke perguruan tinggi di Sintang*. 1–9.

[2] Bhakti Setiawan, M., & Hidayat, A. (2013). *Indeks pembangunan manusia Indonesia*. 18–26.

[3] Dewi Junita, A. M. (2022). *Pengaruh tingkat pendidikan dan penempatan kerja terhadap...* 96–108.

[4] Farida, N. (2022). *Pengaruh pendidikan dan tingkat pengangguran terbuka terhadap Indeks Pembangunan Manusia (IPM) di Indonesia*.

[5] Fatkhurokhman, F., & Yuliana, M. (n.d.). *Klasifikasi Indeks Pembangunan Manusia kabupaten/kota se-Indonesia dengan pendekatan Smooth Support Vector Machine (SSVM) kernel radial basis function (RBF)*. 88–97.

[6] Febrianto, E. N. (2020). *Hubungan sektor informal dan pertumbuhan ekonomi di Indonesia*. 151–158.

[7] Febrianto, L. S., & Karomah, N. K. (2018). *Perbandingan metode robust least median of square (LMS) dan penduga S untuk menangani outlier pada regresi linier berganda*. 84–95.

[8] Fauzi, F., & Yusup, M. Y. (n.d.). *Klasifikasi Indeks Pembangunan Manusia kabupaten/kota se-Indonesia dengan pendekatan SSVM kernel RBF*. 88–97.

[9] Farraas, A. (2024). *Pengaruh tingkat pengangguran terbuka, upah minimum provinsi dan rata-rata lama sekolah terhadap peningkatan Pekerja Migran Indonesia (PMI) di Pulau Jawa tahun 2011–2021*.

[10] Hartono Widodo, R. J. B. (2019). *Perlindungan pekerja migran Indonesia*. Binamulia Hukum, 8(1).

[11] Mardiatmoko, G. (2020). *Pentingnya uji asumsi klasik pada analisis regresi linier berganda (studi kasus penyusunan persamaan allometrik kenari muda [Canarium indicum L.])*. 333–342.

[12] Nasution, L. M. (2017). *Statistik deskriptif*. 49–55.

[13] Nila Hiliyatur Rabbani, R. A. K., & F. S. (2022). *Analisis faktor-faktor yang menyebabkan penduduk memutuskan menjadi Pekerja Migran Indonesia (PMI) di Dusun Bantir Desa Banyu Urip Kecamatan Gerung Kabupaten Lombok Barat*. Society: Jurnal Jurusan Tadris Ilmu Pengetahuan Sosial, 13(2).

[14] Otok, B. W., & (2007). *Mengatasi penyimpangan asumsi normalitas pada pemodelan persamaan struktural menggunakan bootstrap*. 217–230.

[15] Prasetyo, R. A., & H. (2022). *Analisis regresi linear berganda untuk melihat faktor yang berpengaruh terhadap kemiskinan di Provinsi Sumatera Barat*. 62–68.

[16] Puspitasari, W. I., & S. K. (2017). *Faktor-faktor yang mempengaruhi migrasi tenaga kerja ke luar negeri berdasarkan provinsi di Indonesia*. Jurnal Ilmu Ekonomi Terapan.

[17] Ridwan, M., & S. S. (n.d.). *Mendeteksi dan mengatasi multikolinieritas pada data penelitian diabetes melitus wanita suku Indian tahun 2018*. 64–70.

[18] Setiawan, M. B., & Hidayat, A. (2013). *Indeks pembangunan manusia Indonesia*. 18–26.