Modeling Youth Development Index in Indonesia Using Panel Data Regression for Binary Response with Random Effect

Pressylia Aluisina Putri Widyangga[1], Suliyanto[1*], M. Fariz Fadillah Mardianto[1], and Sediono[1]
[1]Author Affiliation: Faculty of Science and Technology, Airlangga University, Surabaya
*Corresponding author: suliyanto@fst.unair.ac.id

Received: 21 September 2024
Revised: 17 April 2025
Accepted: 6 May 2025

ABSTRACT – Indonesia has the largest youth population in Southeast Asia, yet its Youth Development Index (YDI) ranks only fifth in the region. This study aims to fill the gap in empirical research by modeling the YDI in Indonesia using binary logit and binary probit regressions with random effects, based on panel data from 34 provinces during 2020–2022. The YDI categories are defined according to the national target of 57.67 set by the Ministry of Youth and Sports Affairs. The analysis reveals that the binary probit model performs better than the binary logit model, with a classification accuracy of 93.14% and a McFadden R-squared of 0.4064. Gender Inequality Index (GII) and Expected Years of Schooling (EYS) significantly affect the likelihood of achieving the YDI target. These results highlight the critical role of gender equality and education in advancing youth development in Indonesia. The binary probit model provides a practical tool for policymakers to predict and evaluate the effectiveness of development programs targeting youth outcomes. This research not only contributes methodologically to the study of youth development using advanced econometric models but also offers policy-relevant insights that support the strategic goals of Indonesia Emas 2045. By identifying key leverage points such as gender equity and education access, the findings reinforce the importance of inclusive and evidence-based planning to nurture a generation of resilient, empowered, and high-performing youth who can lead Indonesia toward a prosperous future.

Keywords– Youth Development Index, Panel Data Regression, Random Effect, Indonesia.

I.INTRODUCTION

Amidst the rapid progression of technology and the ever-evolving dynamics of social transformation, the role of young people as the primary drivers of development is gaining heightened significance, particularly in the context of Indonesia. As the country navigates complex changes brought about by globalization and modernization, the youth are positioned at the forefront of efforts to innovate, lead, and shape the future. Their contributions are not only pivotal to advancing economic growth but also to fostering social cohesion and addressing critical challenges faced by the nation. In this era of change, empowering and investing in youth development becomes imperative for Indonesia's long-term progress and sustainability. Indonesia occupies the first position with the largest number of youth in Southeast Asia, reaching 273.52 million people in 2023 (BPS, 2023). In today's modern era, youth are expected to be agents of change who are able to face the challenges of the times with innovation and creativity. The importance of the role of youth in Indonesia's development is reflected in the government's efforts to map and measure the progress of youth in Indonesia. In 2017, this mapping effort was proven through the launch of the Association of South-East Asian Nations (ASEAN) Youth Development Index (YDI) in Jakarta with support from the United Nations Population Fund (UNFPA) and the ASEAN Ministerial Meeting on Youth (AMMY) (Kemenpora, 2021). The ASEAN YDI identifies several important domains, including education, health and well-being, employment and opportunities, participation and leadership, and gender and discrimination (Kemenpora, 2023).

YDI has a close relationship with the achievement of the Sustainable Development Goals (SDGs). The YDI includes aspects of youth education that are in line with SDGs point 4 related to increasing inclusive access and equal quality of education. The health domain of the YDI includes access to reproductive health and mental health services in line with SDGs target point 3 to improve general health and well-being. The economic aspect of YDI includes youth employment and entrepreneurship opportunities emphasized in SDGs point 8 to achieve inclusive economic growth and productive employment development for youth. The participation and engagement domain of the YDI reflects youth involvement in development and decision-making processes in line with SDGs point 16 on the importance of effective, inclusive and participatory governance. In an effort to improve YDI in Indonesia, the government has taken strategic steps, such as organizing youth entrepreneurship lectures, providing policy interventions related to youth leadership education, strengthening reproductive health, and efforts to prevent stunting and protect children in order to increase youth participation in development, improve the quality of life of youth, and explore the maximum potential of Indonesian youth in facing global challenges. However, despite various efforts to improve the YDI in Indonesia, it still ranks fifth in Southeast Asia, lower than Singapore, Malaysia, Brunei Darussalam, and Vietnam (Rashed, 2019).

The challenges of youth development in Indonesia are highlighted by the fact that the YDI target has not been achieved evenly in all provinces. According to the Ministry of Youth and Sports (Kemenpora) of the Republic of

Indonesia, Indonesia's YDI is targeted to reach 57.67 in 2024 (Kemenpora, 2023). However, according to the Central Statistics Agency (BPS) publication related to Indonesian Youth Statistics, there are still 10 provinces in Indonesia with YDI below the Ministry of Youth and Sports target in 2022 (BPS, 2023). The YDI is particularly important for Indonesia, not only because the country has the largest youth population in Southeast Asia, but also because this demographic group plays a vital role in shaping the nation's future. As Indonesia moves toward its long-term national vision of "Indonesia Emas 2045," youth development becomes essential. The YDI provides a comprehensive framework to assess the well-being and potential of young people across key areas such as education, health, employment, and civic participation. Understanding the factors that influence YDI is therefore critical to formulating policies that effectively support and empower youth to contribute meaningfully to national progress. This study responds to a gap in the existing literature, where the modeling of YDI in the Indonesian context remains limited. Most previous research has relied on descriptive approaches. In an effort to improve YDI, it is necessary to conduct an in-depth analysis to identify the factors that influence YDI in Indonesia using panel data regression analysis covering all provinces in Indonesia in the time period 2020 to 2022. Panel data regression is a statistical technique that integrates both cross-sectional and time series data, making it particularly well-suited for analyzing datasets that exhibit significant variability over time and across entities. This method allows researchers to account for individual heterogeneity, observe dynamic changes, and better understand the relationships between variables in scenarios where fluctuations are common. By leveraging the strengths of both cross-sectional and time series analyses, panel data regression offers a more robust framework for capturing the complexities inherent in volatile data, providing deeper insights into trends and patterns (Chairunnisa, 2023). The response in panel data regression is divided into two, namely continuous response and categorical response. In panel data regression with continuous response, there are three general approaches, including common effect model, fixed effect model, and random effect model (Stiglitz, 2021). Panel data regression with categorical responses is divided into two types, namely responses with nominal and ordinal scales. Panel data regression with a response of two categories on a nominal scale with two categories can be approached with two types of link functions, including binary logit and binary probit (Timoneda, 2021). This approach allows for a more detailed understanding of the probability that a province meets the national YDI target set by the Ministry of Youth and Sports Affairs. These models are specifically designed to handle dichotomous outcomes and offer appropriate estimation techniques for predicting probabilities constrained between zero and one. Compared to linear probability models, which may predict values outside the logical bounds and suffer from heteroscedasticity, the logit and probit approaches offer more statistically robust estimates. Additionally, these models allow for meaningful interpretation of marginal effects, which is crucial for understanding the impact of explanatory variables such as gender inequality and expected years of schooling on the probability of achieving youth development goals. The findings reveal that variables such as gender inequality and expected years of schooling significantly affect YDI outcomes. As a result, this research not only enhances academic understanding of youth development in Indonesia but also provides valuable insights for policymakers. These insights support more targeted and data-driven strategies to cultivate a generation of capable, resilient, and high-performing youth who will play a central role in achieving Indonesia's future aspirations.

Research on the correlation of YDI with the Human Development Index (HDI) has been conducted by Efendi (2020) and the results show that there is a correlation between the YDI variable and the HDI with a correlation coefficient of 0.485 so it can be concluded that YDI and HDI have a unidirectional relationship. Previous research has the disadvantage that it only analyzes the correlation between YDI and HDI and does not analyze the dimensions and indicators of YDI as a whole. Yolanda et al. (2022) conducted research related to gender inequality. Based on the research that has been conducted, it is concluded that the gender gap in youth development can affect the achievement of youth development that is lost due to gender disparities in the dimensions of health, empowerment, and the labor market so that the decreasing Gender Inequality Index (GII) in Indonesia from year to year can provide great hope for the future of development in Indonesia (Hupkau & Petrongolo, 2020). Research conducted by Rahminawati (2023) emphasizes the importance of education, especially for Indonesian youth because the higher the expected years of schooling, the higher the level of education undertaken and the higher the quality of a person, especially youth in their mindset and pattern of action.

Therefore, this study is important because there are limited studies that use binary logistic regression approach to analyze the factors that influence YDI in Indonesia based on the targets set by the Ministry of Youth and Sports Affairs. This study will divide the YDI response into two categories, namely provinces with YDI above the Kemenpora target and provinces with YDI below the Kemenpora target. This study aims to compare binary logit and binary probit panel data models with random effects using McFadden R-Squared and classification accuracy values as evaluation parameters to obtain the best model in modeling YDI in Indonesia and find out the factors that have a significant effect on YDI. The results of this study are expected to help the government, especially the Ministry of Youth and Sports Affairs, in achieving the YDI target in all provinces in Indonesia in the following years.

II.LITERATURE REVIEW

A. Youth Development Index (YDI)

The Youth Development Index (YDI) is an index used to measure the quality of youth in Indonesia. The importance of the YDI can be seen from its broad impact. Youth are important assets in a country's development as they are potential agents of change. Therefore, understanding the conditions and needs of youth is an important step in formulating effective policies for sustainable development. The YDI is a measurement tool for youth development across five basic domains namely education, health and well-being, employment and opportunities, participation and leadership, and gender and discrimination. As such, the YDI is not only a measurement tool, but also a foundation for the development of youth-oriented policies that aim to create an enabling environment for their growth and well-being, and enable youth to actualize their full potential in society.

B. Factors that Affect YDI in Indonesia

The representation of women in managerial positions has a direct influence on youth development, as workplace gender equality not only embodies the principles of social justice but also plays a vital role in fostering economic growth and advancing social progress. The proportion of women in leadership roles is a key metric in youth development analysis, as it reflects the degree of gender parity within the labor market, the empowerment of women in economic spheres, and the extent of opportunities available for women to attain leadership and decision-making roles. This indicator not only underscores the progress toward achieving gender equality but also highlights the broader implications for inclusive economic development, providing young people with role models and shaping a more equitable future workforce (Larashati, 2022). In addition, in terms of gender and health, the Gender Inequality Index (GII) reflects human development potential that is hampered by gender disparities in reproductive health, empowerment, and labor market dimensions (Bertocchi *et al.*, 2021). Gender disparities in youth development can affect youth development achievements that are lost due to gender disparities in the dimensions of health, empowerment, and the labor market. The declining GII in Indonesia from year to year will provide great hope for the future of youth development in Indonesia. (Yolanda *et al.*, 2022).

In terms of education, one of the parameters that can be used to ensure equitable access and progress of education in Indonesia is the Expected Years of Schooling (EYS). EYS also takes into account the number of years that children of a certain age are expected to experience education in the future. The greater the expected years of schooling, the higher the level of education achieved, which in turn enhances the overall quality of an individual, particularly young people, in terms of their mindset and behavioral patterns. Longer educational exposure equips youth with critical thinking skills, broader knowledge, and the ability to approach problems with more informed perspectives. As a result, they are more likely to engage in constructive actions, contribute meaningfully to society, and drive positive change in various spheres. Therefore, education plays a pivotal role in shaping not only intellectual development but also the attitudes and actions of young people (Rahminawati, 2023).

C. Probability Distributions

A probability distribution is a set of all possible outcomes of an experiment, each accompanied by its chance or probability. In binary logit and binary probit panel data regression models with random effects, several types of distributions are used including normal, binomial, and logistic distributions. The normal distribution is a distribution function shown by a symmetrical graph called a bell curve. If the random variable X is normally distributed with mean μ and variance σ^2 denoted $X \sim N(\mu, \sigma^2)$, then the probability density function X is as follows.

$$g(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \quad , \quad -\infty < y < \infty \tag{1}$$

Based on the probability density function (pdf) above, the cumulative density function (cdf) of the normal distribution is obtained as follows.

$$P(X \le x) = G(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right) dt \tag{2}$$

If the random variable Z has mean 0 and variance 1, then Z is standard normal distributed and is denoted $Z \sim N(0,1)$. The probability density function of the random variable Z is as follows.

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) \quad , \quad -\infty < z < \infty \tag{3}$$

Based on the probability density function above, we get the cumulative probability density function of the random variable Z is as follows.

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) dt \tag{4}$$

Suppose that X is the random variable where number of successful events in n mutually independent trials with probability of success on each trial equal to π and the probability of failure is $(1-\pi)$ then X is binomial distributed with the following probability density function.

$$f(x) = \binom{n}{x} \pi^x (1-\pi)^{n-x} \ ; \quad x = 0, 1, 2, \dots, n \tag{5}$$

In the binomial distribution, if each experiment is only performed once and the events are independent of each other, the Bernoulli Distribution is obtained with the following probability density function.

$$f(x) = \pi^x (1-\pi)^{1-x} \ ; \quad x = 0, 1 \tag{6}$$

The logistic distribution is a distribution that has a continuous probability density function with a symmetrical curve shape and is similar to the normal distribution. The primary distinction between the normal distribution and the logistic distribution is found in their tail behavior. Specifically, the logistic distribution features a slightly longer tail than the normal distribution. This characteristic has important implications for statistical modeling and inference. The extended tails of the logistic distribution allow it to better accommodate extreme values and outliers, making it particularly useful in contexts where such observations may significantly influence the analysis. A random variable X is said to be logistic distributed or $X \sim Logistic(\gamma, \delta)$ if it has the following probability density function.

$$h(x) = \frac{\exp\left[-\frac{(x-\gamma)}{\delta}\right]}{\delta\left[1+\exp\left[-\frac{(x-\gamma)}{\delta}\right]\right]^2} ;\ -\infty < \gamma < +\infty, \delta > 0, -\infty < x < +\infty \tag{7}$$

Based on Equation (7), the cumulative probability density function of X can be obtained by the following equation.

$$H(x) = \frac{1}{1+\exp\left[-\frac{(x-\gamma)}{\delta}\right]} \tag{8}$$

D. Panel Data

Panel data consists of observations of cross-section units over several time periods so that it is more suitable in studying the dynamics of change. Suppose there are T time period with $t = 1, 2, \ldots, T$ and N the number of individuals or subjects with $i = 1, 2, \ldots, N$, then the total panel data observations to be used are as many as NT (Gujarati & Damodar, 2009). When each individual has the same number of time periods, the dataset is referred to as a balanced panel. Conversely, if the number of time periods varies across individuals, it is known as an unbalanced panel (Baltagi, 2021). The advantages of using panel data are numerous and include the ability to provide a richer dataset, which results in greater degrees of freedom and more efficient estimates. Panel data allows researchers to control for individual heterogeneity by accounting for variations between entities that may remain constant over time, offering more precise and reliable insights. Additionally, it enables the identification and measurement of effects that may go undetected in purely cross-sectional or time series data. Panel data can also help mitigate the problem of omitted variable bias by including both time and individual dimensions, thus capturing more relevant factors. Furthermore, it is well-suited for studying dynamic changes, as it tracks the evolution of variables over time, providing a deeper understanding of temporal processes and trends.

E. Binary Logit Regression for Panel Data with Random Effect

Binary logit regression on panel data is a regression analysis with a panel data structure that has binary response and influenced by predictor variables that are categorical or continuous with random effects. Suppose there is panel data y_{it} and X_{it} with $i = 1, 2, \ldots, N; t = 1, 2, \ldots, T$ fulfill the variance component model as follows:

$$y_{it}^* = \boldsymbol{X_{it}\beta} + u_i + v_{it} \tag{9}$$

where:
y_{it}^* is the latent response variable.
$\boldsymbol{X_{it}} = (1 \quad X_{1it} \quad X_{2it} \quad \cdots \quad X_{pit})$ is the vector of predictor variables at i-th unit cross-section and t-th unit time series.
$\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)'$ is the parameter vector.
u_i is the random effect of the i-th unit cross-section which is assumed to be identically independent (iid) distributed with $N(0, \sigma_u^2)$.
v_{it} is the random error of the i-th unit cross-section and t-th unit time series which is assumed to be logistic distributed with mean equal to 0 and variance equal to 1 independent of u_i. Suppose y_{it} is the observed response variable of the i-th unit cross-section and t-th unit time series (Greene, 2023).

$$y_{it} = \begin{cases} 0, & y_{it}^* \leq 0 \\ 1, & y_{it}^* > 0 \end{cases} \tag{10}$$

Parameter estimation of binary logit regression model on panel data with random effects is done using the maximum likelihood method. Suppose $H(\cdot)$ is the cumulative probability density function of the logistic distribution, the probability of success is obtained as follows:

$$\pi_{it} = \Pr(y_{it} = 1 | \boldsymbol{X_{it}}) = \Pr(\boldsymbol{X_{it}\beta} + u_i | v_{it})$$
$$\pi_{it} = \Pr(v_{it} > -\boldsymbol{X_{it}\beta} - u_i)$$
$$\pi_{it} = 1 - P(v_{it} \leq -\boldsymbol{X_{it}\beta} - u_i)$$
$$\pi_{it} = 1 - H(-\boldsymbol{X_{it}\beta} - u_i)$$
$$\pi_{it} = 1 - \frac{1}{1+\exp(\boldsymbol{X_{it}\beta} + u_i)}$$
$$\pi_{it} = \frac{\exp(\boldsymbol{X_{it}\beta} + u_i)}{1+\exp(\boldsymbol{X_{it}\beta} + u_i)} \tag{11}$$

where π_{it} is the probability of success corresponding to the predictor variable $\boldsymbol{X_{it}}$.

F. Binary Probit Regression for Panel Data with Random Effect

Binary probit regression for panel data with random effect model is same as in Equation (9) and the response categorization is based on Equation (10) where u_i is the random effect of the i-th unit cross-section which is assumed to be identically independent (iid) distributed with $N(0, \sigma_u^2)$ and v_{it} is the random error of the i-th unit cross-section and t-th unit time series which is assumed to be distributed $N(0,1)$ which is independent of u_i. Parameter estimation of binary probit regression model on panel data with random effects is done using the maximum likelihood method.

Suppose $\Phi(\cdot)$ is the standard normal distribution function. Based on Equation (10), the success probability of the response variable is obtained y_{it} is as follows:

$p(\boldsymbol{X_{it}}) = \Pr(y_{it} = 1|u_i, \boldsymbol{X_{it}})$
$p(\boldsymbol{X_{it}}) = \Pr(y_{it}^* > 0)$
$p(\boldsymbol{X_{it}}) = \Pr(v_{it} > -\boldsymbol{X_{it}\beta} - u_i)$
$p(\boldsymbol{X_{it}}) = \Pr(v_{it} < \boldsymbol{X_{it}\beta} + u_i)$
$p(\boldsymbol{X_{it}}) = \Phi(\boldsymbol{X_{it}\beta} + u_i)$
$p(\boldsymbol{X_{it}}) = \Phi(z_{it})$ (12)

with $z_{it} = \boldsymbol{X_{it}\beta} + u_i$.

Based on Equation (12), the failure probability of the response variable is obtained y_{it} is as follows.

$q(\boldsymbol{X_{it}}) = 1 - p(\boldsymbol{X_{it}})$
$q(\boldsymbol{X_{it}}) = 1 - \Phi(z_{it})$

The random effects model assumes homoskedasticity in the unit variance (Butler and Moffitt, 1982). u_i and v_{it} are independent random variables with $E[v_{it}|\boldsymbol{X}] = 0; Cov[v_{it}, v_{js}|\boldsymbol{X}] = Var[v_{it}|\boldsymbol{X}] = 1$ if $i = j$ and $t = s$ and 0 for all others. $E[u_i|\boldsymbol{X}] = 0$ and $Cov[u_i, u_j|\boldsymbol{X}] = Var[u_i|\boldsymbol{X}] = \sigma_u^2$ if $i = j$ and 0 for all others and $Cov[v_{it}, u_j|\boldsymbol{X}] = 0$ for all i, t, j. Suppose $\varepsilon_{it} = u_i + v_{it}$then $E[\varepsilon_{it}|\boldsymbol{X}] = 0, Var[\varepsilon_{it}|\boldsymbol{X}] = 1 + \sigma_u^2$. Thus, the covariance between errors is obtained as follows.

$Cov[\varepsilon_{it}, \varepsilon_{is}|\boldsymbol{X}] = E[(\varepsilon_{it} - 0)(\varepsilon_{is} - 0)]$
$Cov[\varepsilon_{it}, \varepsilon_{is}|\boldsymbol{X}] = E(\varepsilon_{it} \cdot \varepsilon_{is}) = E[(u_i + v_{it})(u_i + v_{is})]$
$Cov[\varepsilon_{it}, \varepsilon_{is}|\boldsymbol{X}] = E(u_i^2) + E(v_{it}v_{is}) = \sigma_u^2$

Based on the calculation of covariance between errors, the correlation between individual errors is obtained, namely $Corr[\varepsilon_{it}, \varepsilon_{is}|\boldsymbol{X}] = \rho = \frac{\sigma_u^2}{1+\sigma_u^2}$ so that the new free parameters are obtained, namely $\sigma_u^2 = \frac{\rho}{1-\rho}$ with ρ is the correlation between errors on the same individual (Greene, 2012).

Suppose $\boldsymbol{\varepsilon_i} = (\varepsilon_{i1} \quad \varepsilon_{i2} \quad \cdots \quad \varepsilon_{iT})'$, then obtained

$$\boldsymbol{\Sigma_i} = cov(\boldsymbol{\varepsilon_i}) = E(\boldsymbol{\varepsilon_i} \cdot \boldsymbol{\varepsilon_i'}) = E\left(\begin{bmatrix}\varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{iT}\end{bmatrix}[\varepsilon_{i1} \quad \varepsilon_{i2} \quad \cdots \quad \varepsilon_{iT}]\right) \tag{13}$$

$$\boldsymbol{\Sigma_i} = cov(\boldsymbol{\varepsilon_i}) = E(\boldsymbol{\varepsilon_i} \cdot \boldsymbol{\varepsilon_i'}) = \begin{bmatrix}1+\sigma_u^2 & \sigma_u^2 & \cdots & \sigma_u^2 \\ \sigma_u^2 & 1+\sigma_u^2 & \cdots & \sigma_u^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_u^2 & \sigma_u^2 & \cdots & 1+\sigma_u^2\end{bmatrix} \tag{14}$$

Due to the following observations$i$ and j-th are independent, then the covariance matrix for the NT observations is as follows:

$$\boldsymbol{\Omega} = \begin{pmatrix}\boldsymbol{\Sigma}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_2 & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{\Sigma}_\mathrm{N}\end{pmatrix} \tag{15}$$

G. Parameter Significance Test

Simultaneous parameter testing is conducted to determine whether the predictor variables, when considered together, have a statistically significant impact on the response variable. This type of testing evaluates the collective influence of all the explanatory variables in a model, as opposed to testing them individually. By performing this test, researchers can assess whether the combined effect of the predictors contributes meaningfully to explaining the variation in the dependent variable, which is essential for understanding the overall model's validity and predictive power. The hypothesis formulation used to test simultaneously is as follows.

$\mathrm{H}_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$
H_1 : there is at least one $\beta_j \neq 0 \; ; \; j = 1, 2, \dots, p$

The *Wald* test is a type of test used to test parameters simultaneously. Wald Test statistics are presented through the following equation:

$$W = (\boldsymbol{\hat{\beta}} - \boldsymbol{\beta})'\boldsymbol{V^{-1}}(\boldsymbol{\hat{\beta}} - \boldsymbol{\beta}) = (\boldsymbol{\hat{\beta}} - \boldsymbol{0})'\boldsymbol{V^{-1}}(\boldsymbol{\hat{\beta}} - \boldsymbol{0}) \tag{16}$$

with $\boldsymbol{V} = [\boldsymbol{I}(\boldsymbol{\hat{\beta}})]^{-1} = \left[-E\left(\boldsymbol{H}(\boldsymbol{\hat{\beta}})\right)\right]^{-1} = \left[-\frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\right]^{-1}$ where $\boldsymbol{H}(\boldsymbol{\hat{\beta}})$ is the Hessian matrix.

The Wald test is Chi-Square distribution which has free degrees p where p is the number of predictor variables. The null hypothesis is rejected if the $W > \chi^2_{(p;\alpha)}$ or p-value is smaller than α which means that the predictor variables simultaneously affect the response variable.

The normal standard test is also used to test the significance of predictor variables individually on the response variable. The hypothesis formulation used to test individually or partially is as follows.

$\mathrm{H}_0 : \beta_j = 0$
$\mathrm{H}_1 : \beta_j \neq 0 \; ; \; j = 1, 2, \dots, p$

The normal standard test statistic used to test the parameters individually is presented in the following equation:

$$Z_j = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \tag{17}$$

with $SE(\hat{\beta}_j)$ is the standard error *of* the estimator $\hat{\beta}_j$ which is obtained from the main diagonal of the $j+1$ of the matrix $\boldsymbol{V}$. The test criteria is to reject the null hypothesis if the value of $|Z_j| > Z_{\frac{\alpha}{2}}$ or p-value is smaller than α which means that the predictor variable affects the response variable.

H. Goodness of Fit Test

Modeling is undertaken with the goal of establishing a relationship that accurately represents the connection between the response and predictor variables. The primary objective is to develop a model that captures the underlying patterns and dependencies within the data. Once a model is constructed, a model fit test is performed to evaluate how well the predicted model aligns with the actual data. The hypothesis formulation used to test the suitability of the model is as follows.

H_0 : Binary logit or binary probit regression models on panel data with random effects are not appropriate.

H_1 : Binary logit or binary probit regression models on panel data with random effects fit.

The test used to test the hypothesis formulation above is the Likelihood Ratio (LR) Test with the following test statistics:

$$\Lambda = -2\ln\left(\frac{L_{H_0}}{L_{H_1}}\right) \tag{18}$$

with

L_{H_0} : maximum likelihood of a standard binary logit or binary probit model under the hypothesis of H_0

L_{H_1} : maximum likelihood of a standard binary logit or binary probit model under the hypothesis of H_1

The rejection criterion is to reject the null hypothesis if the value $\Lambda > \chi^2_{(1;\alpha)}$ or *p-value* is smaller than α which means that the binary logit or binary probit regression model on panel data with random effects is appropriate.

I. Classification Accuracy

Classification accuracy is a commonly used metric for evaluating a model's performance, specifically in identifying the likelihood of errors when classifying objects. One key measure of these errors is the Apparent Error Rate (APPER), which represents the proportion of samples that are misclassified by the classification function. Essentially, APER quantifies the fraction of incorrect predictions made by the model out of the total number of observations. By assessing the APER, researchers can gauge the model's reliability and its potential for misclassification, providing valuable insights into its effectiveness in distinguishing between different classes (Agresti, 2018). Measurement of classification accuracy can be done based on Table 1.

Table 1 Calculation of Classification Accuracy Value

Observation Results	**Prediction Result**		**Total**
	Y = 0	Y = 1	
Y = 0	n_{00}	n_{01}	n_{0+}
Y = 1	n_{10}	n_{11}	n_{1+}
Total	n_{+0}	n_{+1}	n_{++}

The APPER value can be calculated through the formula presented in the following equation:

$$\text{APPER} = \frac{n_{10} + n_{01}}{n_{11} + n_{10} + n_{01} + n_{00}} \tag{19}$$

with n_{11} is the number of successful observations that are classified as successful based on the prediction result, n_{10} is the number of successful observations that are classified as failed based on the prediction results, n_{01} is the number of failed observation events that are classified as successful based on the prediction results, and n_{00} is the number of failed observations that are classified as failed based on the prediction results. The accuracy value of the model classification is obtained through the formula presented in the following equation.

$$\text{Accuracy Value} = (1 - \text{APPER}) \times 100\% \tag{20}$$

J. R-Squared McFadden

R-squared McFadden is a statistical coefficient used to assess how much influence the independent variables have on the dependent variable, particularly in logistic regression and other generalized linear models. It serves as an indicator of the model's explanatory power, measuring the extent to which the model accounts for the variation in the dependent variable. Unlike the traditional R-squared used in linear regression, R-squared McFadden typically takes values between 0 and 1 (Johnston, 2009). If the R-squared McFadden value is closer to 1, then the model is considered better in explaining the diversity of response variables based on predictor variables and vice versa. R-squared McFadden is designed for nonlinear models such as logistic regression because the calculation uses the likelihood function. McFadden's R-squared can be formulated through the following equation.

$$R^2_{McF} = 1 - \frac{\ln(L_{H_1})}{\ln(L_{H_0})} \tag{21}$$

K. Marginal Effect

In the analysis of binary logit and binary probit regression for panel data with random effect, not only the coefficient value depend on the model, but also the marginal effects. The concept of marginal effects is employed to assess the degree of influence that predictor variables exert on the response variable, moving beyond reliance solely on the values of the model coefficients. Marginal effects quantify how a small change in a predictor variable impacts the expected value of the response variable, providing a more intuitive understanding of relationships within the model. This approach is particularly useful in nonlinear models, such as logistic regression, where the effect of predictor variables may vary across different levels of those variables. By calculating marginal effects, researchers can gain deeper insights into the practical significance of predictors, enabling them to interpret results in a more meaningful way that reflects real-world implications. The marginal effects in the binary logit regression for panel data with random effect are presented in Equation (22).

$$\frac{\partial \Pr(y_{it}=1|u_i,\boldsymbol{x_{it}})}{\partial x_{jit}} = \hat{\beta}_j \frac{\exp(\boldsymbol{X_{it}\beta}+u_i)}{(1+\exp(\boldsymbol{X_{it}\beta}+u_i))^2}\ ; j=1,2,\dots,p \tag{22}$$

Different from binary logit regression for panel data with random effect, the marginal effects in the binary probit regression for panel data with random effect are presented in Equation (23).

$$\frac{\partial \Pr(y_{it}=1|u_i,\boldsymbol{x_{it}})}{\partial x_{jit}} = \hat{\beta}_j \emptyset(\boldsymbol{x_{it}\beta}+u_i)\ ; j=1,2,\dots,p \tag{23}$$

III. METHODOLOGY

The data used in this study are secondary data regarding data on factors affecting YDI in Indonesia in 2020-2022. The data was obtained from the publication of the Central Bureau of Statistics and the Ministry of Youth and Sports Affairs with the unit of observation in this study being 34 provinces in Indonesia. The binary categorization of the YDI variable is grounded in several previous studies that have explored factors influencing youth development. The variables used to define YDI are selected based on their theoretical and empirical relevance, as suggested by the literature. Therefore, this binary classification can be justified and deemed appropriate for analysis. The specifics of the response and predictor variables utilized in this study are comprehensively outlined in Table 2.

Table 2 Research Variables

Variables	Description	Operational Definition	Scale	Unit
Y	Youth Development Index (YDI)	Index that provides an overview of the three layers of youth development domains, including individual development, livelihood and welfare development, and participation in various spheres of life.	Category	0 = YDI < 57.67 1 = YDI ≥ 57.67
$\mathbf{X_1}$	Percentage of Women in Managerial Positions	The proportion of women holding management positions within a company or organization refers to the percentage of female employees who occupy leadership roles compared to the total number of management positions available.	Ratio	Percent
$\mathbf{X_2}$	Gender Inequality Index (GII)	This measurement evaluates the extent to which human development is hindered by gender inequality across three critical dimensions: reproductive health, empowerment, and economic participation. By examining disparities between men and women in these areas, the index sheds light on the significant impact of gender inequality on overall development outcomes. It highlights not only the challenges faced by women in achieving equitable access to health services and economic opportunities but also the broader societal implications of such disparities.	Ratio	Index
$\mathbf{X_3}$	Expected Years of Schooling (EYS)	The term for the number of years of schooling a child is expected to receive at a certain age in the future. This indicator projects the total years of formal education a child entering the education system is likely to complete, based on current enrollment rates and trends.	Ratio	Year

Based on the research objectives that have been formulated in the introduction, the data analysis stages are as follows.

1. Describing the characteristics of YDI variables in Indonesia in 2020 to 2022.
2. Describing the factors that influence YDI in Indonesia in 2020 to 2022 based on previous research.
3. Grouping the YDI response variable into two categories, namely Y=0 for provinces with YDI values that have not reached the Ministry of Youth and Sports Affairs target of 57.67 and Y=1 for provinces with YDI values above the Ministry of Youth and Sports Affairs target of 57.67 in 2020 to 2022.

4. Regressing response variables and predictor variables to obtain binary logit and binary probit regression models for panel data with random effects.
5. Comparing McFadden R-squared and classification accuracy values to find the best model.
6. Testing the best model of panel data binary logistic regression with random effects simultaneously and partially.
7. Analyzing and interpret the best model obtained.
8. Obtaining conclusions from the data analysis results that has been carried out.

The dataset encompasses multiple regions across Indonesia over a specified time period, providing a panel structure that strengthens the analytical framework. However, as with most secondary data analyses, certain limitations are present. The aggregation of indicators at the regional level may obscure intra-regional disparities, and the use of cross-sectional snapshots for some variables may limit the ability to fully capture dynamic trends. Furthermore, potential biases may arise from differences in data collection standards across regions or from unobserved variables that were not included in the model but may influence youth development outcomes. These limitations are acknowledged as part of the analytical context and are considered when interpreting the results.

IV.RESULTS AND DISCUSSIONS

A. *Descriptive Statistics*

The descriptive statistics of the three predictor variables including the percentage of women in managerial positions, the Gender Inequality Index (GII), and the Expected Years of Schooling (EYS) that are thought to affect YDI are outlined in Table 3.

Table 3 Descriptive Statistics

Variables	Minimum		Maximum	
	Province	Year	Province	Year
Percentage of Women in Managerial Positions	Southeast Sulawesi	2020	Gorontalo	2020
Gender Inequality Index (GII)	Special Region of Yogyakarta	2020	West Nusa Tenggara	2020
Expected Years of Schooling (EYS)	Papua	2020	Special Region of Yogyakarta	2022

Based on the data presented in Table 3, variations can be observed across several key indicators that are believed to influence the YDI in Indonesia. The first variable is the percentage of women in managerial positions where BPS noted that Southeast Sulawesi Province in 2020 had the lowest percentage of women in managerial positions. Conversely, in the same year, Gorontalo Province exhibited a more favorable percentage, achieving the highest proportion of women in managerial positions. This indicates a notable advancement in gender representation within leadership roles in the region. In contrast, GII reveals significant disparities. These disparities highlight the ongoing challenges faced by women in other areas, underscoring the complexities of gender equality in Gorontalo. While the increase in female managerial representation is commendable, the contrasting GII suggests that further efforts are needed to address broader issues of gender inequality in various dimensions, including reproductive health, empowerment, and economic participation. In 2020, Yogyakarta Special Region Province recorded the lowest GII value, indicating better gender equality compared to other provinces. In contrast, West Nusa Tenggara Province in the same year had the highest GII value, indicating greater gender inequality. The third variable is Expected Years of Schooling (EYS). BPS noted that in 2020, Papua Province showed the lowest score, reflecting the major challenges in education in the province. Meanwhile, in 2022, the Special Region of Yogyakarta Province managed to achieve the highest EYS value, indicating a significant achievement in extending the schooling period for its population.

The dataset utilized in this study comprised 102 observations for each variable analyzed. The YDI response variable used in this study uses categorical data. The classification of YDI variable categories is based on the Ministry of Youth and Sports Affairs target of 57.67 in each year presented in Table 4.

Table 4 Percentage of YDI Target Achievement

Year	YDI Target Achieved		YDI Target Not Achieved	
	Frequency	Percentage	Frequency	Percentage
2020	2	1.96%	32	31.37%
2021	3	2.94%	31	30.39%
2022	5	4.90%	29	28.43%
Total	**10**	**9.80%**	**92**	**90.20%**

Based on Table 4, it can be seen that the achievement of the YDI target set by the Ministry of Youth and Sports Affairs of 57.67 is still far from expectations during the period 2020 to 2022. Overall, in these three years, only 9.80% of all provinces in Indonesia in 2020 to 2022 managed to achieve the YDI target, while the remaining 90.20% of provinces did not reach the target. In 2020, only 1.96% of Indonesian provinces achieved the expected YDI target and 31.37% failed to meet the target. The figure increased slightly in 2021, where there were 2.94% of provinces that had reached the YDI target and the remaining 30.39% that did not succeed in reaching the target. In 2022, there was a more significant increase with 4.90% of provinces having achieved the YDI target, but 28.43% of other provinces had not reached the target.

B. Modeling YDI Using Binary Logit Regression for Panel Data with Random Effect

Based on the modeling of YDI using the binary logit regression approach for panel data with random effects, the following model equation is obtained:

$$\hat{y}_{it}^{*} = -24.40787 - 0.027465\, x_{1it} - 15.06267\, x_{2it} + 2.169165\, x_{3it} \tag{24}$$

with $\hat{y}_{it}^{*}$ is the latent variable YDI in Indonesia in 2020 to 2022 so that the YDI prediction is obtained as follows.

$$\hat{y}_{it} = \begin{cases} 1 \text{ jika } \hat{y}_{it}^{*} > 0 \\ 0 \text{ jika } \hat{y}_{it}^{*} \leq 0 \end{cases}$$

If $\hat{y}_{it} = 1$, then the YDI value has reached the target of the Ministry of Youth and Sports and vice versa if $\hat{y}_{it} = 0$, then the YDI value has not reached the Ministry of Youth and Sports target of 57.67. Based on the analysis of the simultaneous test using the Wald test, involving the variables of the percentage of women in managerial positions, GII, and EYS within a binary logit regression model for panel data with random effects, the Wald test statistical value was 8.87049 with a p-value of 0.0311 so that the decision to reject the null hypothesis was obtained because the test statistical value was greater than $\chi^2_{(3;0,05)} = 7{,}814728$ and the p-value is smaller than alpha 0.05. Thus, it can be concluded that there is at least one predictor variable that significantly influence YDI in Indonesia. The results of testing the significance of the individual parameters of the three predictor variables on the response variable are presented in Table 5.

Table 5 Partial Test Model Binary logit regression for panel data with random effect

Variables	Z_{score}	P-Value	Decision
Percentage of Women in Managerial Positions	-0.305880	0.7597	Fail to Reject H_0
GII	-2.740520	0.0061	Reject H_0
EYS	2.316222	0.0205	Reject H_0

Therefore, based on the analysis presented in Table 5, which employs binary logit regression for panel data with random effects, it can be concluded that only the GII and EYS variables significantly influence the likelihood of achieving YDI that exceeds the targets set by the Ministry of Youth and Sports Affairs. This finding underscores the importance of addressing gender inequality and enhancing educational opportunities to foster youth development in alignment with governmental objectives. Furthermore, the proportion of women in managerial positions does not have a significant impact on achieving YDI that exceeds the target set by the Ministry of Youth and Sports Affairs.

C. Modeling YDI Using Binary Probit Regression for Panel Data with Random Effect

Based on the modeling of YDI using the binary probit regression approach for panel data with random effects, the following model equation is obtained:

$$\hat{y}_{it}^{*} = -12.53974 - 0.014906\, x_{1it} - 7.930959\, x_{2it} + 1.114637\, x_{3it} \tag{25}$$

with $\hat{y}_{it}^{*}$ is the latent variable YDI in Indonesia in 2020 to 2022 so that the YDI prediction is obtained as follows.

$$\hat{y}_{it} = \begin{cases} 1 \text{ jika } \hat{y}_{it}^{*} > 0 \\ 0 \text{ jika } \hat{y}_{it}^{*} \leq 0 \end{cases}$$

If $\hat{y}_{it} = 1$, then the YDI value has reached the target of the Ministry of Youth and Sports Affairs if $\hat{y}_{it} = 0$, then the YDI value has not reached the Ministry of Youth and Sports Affairs target of 57.67. Based on the analysis of the simultaneous test of the variables of the percentage of women in managerial positions, GII, and EYS using the Wald test with the binary probit regression approach for panel data with random effects, the Wald test statistical value is 9.665288 which is greater than $\chi^2_{(3;0,05)} = 7{,}814728$ and a p-value *of* 0.0216 which is smaller than the significance level or alpha 0.05. As a result, the rejection of the null hypothesis indicates that there is at least one predictor variable that has a significant impact on the YDI in Indonesia. The results of the significance testing for the individual parameters of the three predictor variables in relation to the response variable are detailed in Table 6.

Table 6 Partial Test Model Binary logit regression for panel data with random effect

Variables	Z_{score}	P-Value	Decision
Percentage of Women in Managerial Positions	-0.331780	0.7401	Fail to Reject H_0
GII	-2.755533	0.0059	Reject H_0
EYS	2.375490	0.0175	Reject H_0

Referring to Table 6, which employs the binary probit regression method for panel data with random effects, it can be concluded that only the GII and EYS variables exert a significant influence on the achievement of YDI surpassing the target set by the Ministry of Youth and Sports Affairs. Furthermore, the percentage of women in managerial positions does not show a significant effect on the attainment of YDI above the ministry's target.

D. Model Fit Test and Best Model Selection

The fit test of binary logit and binary probit panel data regression models with random effects based on the Likelihood Ratio Test is presented in Table 7.

Table 7 Model Fit Test

Model	LR-Statistics	P-Value	Description
Binary logit regression for panel data with random effects	26.32832	0.000008	Suitable Model
Binary probit regression for panel data with random effects	26.59272	0.000007	Suitable Model

Based on Table 7, it can be seen that both binary logit and binary probit panel data regression with random effects have test statistic values that are greater than $\chi^2_{(1;0,05)} = 8{,}41459$ and the p-value is smaller than the alpha 0.05. Consequently, the null hypothesis is rejected, leading to the conclusion that both binary logit regression and binary probit regression for panel data with random effect is a fit model. The selection of the optimal model between the two is determined by comparing their classification accuracy and McFadden's R-squared as shown in Table 8.

Table 8 Best Model Selection

Model	Classification Accuracy Value	R-squared McFadden
Binary logit regression for panel data with random effects	92.16	0.402367
Binary probit regression for panel data with random effects	93.14	0.406407

As shown in Table 8, the binary probit regression model for panel data with random effects demonstrates a higher classification accuracy and McFadden R-squared value compared to the binary logit regression model with random effects. The classification accuracy of 93.14% for the binary probit model indicates that it performs well in accurately classifying data according to actual observations. Additionally, the McFadden R-squared value of 0.406407 suggests that the model accounts for 40.6407% of the variation in the observed data, in comparison to a null model without predictors. Therefore, the binary probit regression for panel data with random effects is considered the most suitable model for predicting YDI in Indonesia. The calculation of the classification accuracy for this model is derived from the APPER based on the prediction results presented in Table 9.

Table 9 Classification Accuracy Value of Binary probit regression model for panel data with random effect

Observation Results	Prediction Result		Total
	Y = 0	Y = 1	
Y = 0	92	7	99
Y = 1	0	3	3
Total	92	10	102

Based on Table 9, the APPER calculation and the classification accuracy value of the binary probit regression model for panel data with random effects can be done as follows.

$$\text{APPER} = \frac{n_{10} + n_{01}}{n_{11} + n_{10} + n_{01} + n_{00}} = \frac{0 + 7}{3 + 0 + 7 + 92} = \frac{7}{102} = 0.06863$$

Based on the APPER calculation on binary probit regression for panel data with random effects, the classification accuracy value is obtained as follows.

Classification Accuracy Value $= (1 - \text{APPER}) \times 100\% = (1 - 0.06863) \times 100\%$

Classification Accuracy Value $= 0.93137 \times 100\% = 93.137\% \approx 93.14\%$

E. *Binary Probit Regression for Panel Data with Random Effect Model Interpretation*

Based on the best model, the model used to identify factors affecting YDI in Indonesia between 2020 and 2022 is a panel data binary probit regression with random effects. YDI is a measure that reflects the quality of life and development potential of youth in various aspects, including education, health, employment opportunities, participation, and gender. Meanwhile, the independent variables that allegedly affect YDI include the percentage of women in managerial positions, GII, and EYS.

According to Equation (25), the coefficient for the variable representing the percentage of women in managerial positions is -0.014906. The negative sign of this coefficient indicates that, assuming all other variables remain constant, an increase in the percentage of women in managerial roles is associated with a slight decrease in the likelihood of an increase in the YDI. This finding suggests that, contrary to expectations, greater representation of women in leadership positions may not necessarily correlate with improvements in youth development outcomes, warranting further investigation into the underlying factors influencing this relationship. While this result may seem counterintuitive, several factors could explain it. First, the percentage of women in managerial positions may not be high enough to generate a significant positive impact on overall YDI. Additionally, the uneven influence of women in leadership positions implies that, although more women occupy managerial roles, this alone may not be sufficient to drive substantial structural changes in youth development. Nevertheless, the percentage of women in managerial roles does not have a significant effect on YDI. While increasing the representation of women in leadership positions is commonly viewed as a positive move toward achieving gender equality and fostering workforce inclusion, the policies or initiatives that promote this increase may not be directly aligned with efforts to enhance youth development. This disconnect suggests that merely increasing the number of women in managerial roles does not automatically translate into improved outcomes for youth, highlighting the need for targeted strategies that explicitly link leadership diversity with youth development initiatives (Romero, 2023). For instance, greater representation of women in management may stem from company or sector-specific policies focused on gender equality, while policies that directly impact YDI, such as those related to education, health, and social programs, may not be significantly influenced by changes in managerial representation.

GII is a composite indicator that captures the extent of gender disparities across three critical dimensions: reproductive health, empowerment, and labor market participation. It provides a comprehensive reflection of the structural inequalities that limit opportunities for women and girls in many societies. A higher GII value indicates greater gender inequality, while a lower value signifies a more equitable distribution of resources and opportunities between men and women. In this study, the GII coefficient of -7.930959 demonstrates a statistically significant and substantial negative relationship with the YDI, meaning that a one-unit decrease in GII is associated with an increase of 7.930959 in the YDI. This result suggests that reducing gender inequality can lead to significant improvements in youth development outcomes. Gender inequality is often embedded in deep-rooted social, cultural, and institutional norms that restrict women's access to essential resources, such as education, health services, and employment opportunities. These restrictions not only hinder the development of women themselves but also affect the broader societal fabric by limiting the collective potential for growth and innovation. When women face barriers to full participation in economic, political, and social life, the conditions necessary for youth development are weakened. The marginalization of women also means fewer role models and mentors for younger generations, which can hinder aspirations and reduce motivation among youth, particularly girls. As emphasized by Idrus (2023), gender inequality is not an isolated issue affecting only one group, but a systemic problem that undermines the development prospects of entire communities. In contexts where women are excluded from decision-making, leadership, and equal access to education or employment, the opportunities for youth to benefit from inclusive and equitable development are diminished. The lack of gender inclusivity in key sectors leads to a developmental environment that is neither representative nor responsive to the needs of the younger population. Therefore, addressing gender inequality is not only a matter of justice, but also a critical factor in achieving sustainable youth development. Policymakers should consider the integration of gender-sensitive frameworks into education, labor, and health policies, while simultaneously promoting female representation in political and institutional leadership roles.

EYS, on the other hand, reflects the number of years a child entering the education system is expected to spend in formal education, based on current patterns of enrollment and attendance. This indicator not only measures access to education, but also serves as a proxy for the overall quality and inclusivity of the education system. In the present analysis, the positive and statistically significant EYS coefficient of 1.114637 indicates that a one-unit increase in expected years of schooling can increase YDI by 1.114637. This result reinforces the view that education is a cornerstone of youth development, as it enhances knowledge, skills, critical thinking, and life preparedness. Greater access to prolonged and quality education improves a young person's ability to make informed decisions, compete in the labor market, and engage actively in civic life. Higher educational attainment is closely linked to greater productivity, higher income potential, and improved health outcomes. In addition, education increases awareness of social, environmental, and political issues, enabling youth to act as informed citizens and potential agents of change in their communities. Better-educated youth are more likely to challenge injustice, engage in political and social discourse, and contribute to the shaping of inclusive societies. Moreover, a longer duration of schooling increases the chances of acquiring specialized knowledge and technical skills that are aligned with the demands of a modern and evolving economy. Education also fosters entrepreneurial thinking and innovation, both of which are vital to long-term economic resilience. As such, expanding access to education and ensuring its quality and relevance should be a central focus of youth development strategies. Interventions that reduce dropout rates, address disparities in access, and improve teaching effectiveness are essential to realize the transformative potential of education in promoting human capital development and social equity.

The findings of this study, which reveal a significant negative impact of GII and a positive influence of EYS on YDI, are consistent with broader empirical patterns observed in development research. Gender inequality continues to serve as a structural barrier to inclusive social and economic progress. When gender disparities persist, particularly in access to education, healthcare, and economic opportunities, the potential for youth to thrive within society is substantially diminished. High levels of gender inequality often reflect systemic exclusion that hinders not only women’s advancement but also the collective well-being of younger generations. This study’s result, showing that a reduction in GII significantly increases the probability of achieving youth development targets, underscores the importance of gender-inclusive policies in creating enabling environments for youth. In parallel, the positive relationship between EYS and YDI reaffirms the fundamental role of education in shaping youth capabilities and opportunities. Higher expected years of schooling suggest greater exposure to formal education, which contributes to better skill development, civic participation, and long-term employability. Prolonged and equitable access to education equips young people with the tools they need to participate in modern economies, navigate complex social challenges, and contribute meaningfully to their communities. The observed association in this study supports the view that investing in educational systems is not only a pathway to individual success but also a strategic lever for national development. These results, therefore, strengthen the argument for integrated policy frameworks that simultaneously promote gender equality and educational advancement as twin pillars for sustainable youth development.

In this discussion, the YDI model in Indonesia will be analyzed and interpreted by taking the example of one of the provinces in Indonesia, namely Capital Region of Jakarta in 2022. It is known that Capital Region of Jakarta has a percentage of women in managerial positions of 37.21, GII of 0.32, and EYS of 13.06. Thus, the latent response variable model obtained with the best model approach, namely panel data binary probit regression with random effects, is as follows.

$$\hat{y}^*_{6,3} = -12.53974 - 0.014906\ (37.21) - 7.930959\ (0.32) + 1.114637\ (13.06)$$
$$\hat{y}^*_{6,3} = -1.07513992$$

The result is obtained $\hat{y}^*_{6,3}$ of -1.07513992 where the value $\hat{y}^*_{6,3} < 0$. Thus, it can be concluded that based on the panel data binary probit model with random effects, Capital Region of Jakarta Province in 2022 is included in the category of provinces that have not reached the YDI target according to the Ministry of Youth and Sports target.

In addition, the marginal effects of binary probit panel data regression with random effects will be calculated for two significant predictor variables, namely GII and EYS. The calculation of the marginal effects for the GII variable is as follows.

$$\frac{\partial \Pr\left(y_{6,3} = 1 \middle| u_6, \boldsymbol{x_{6,3}}\right)}{\partial x_{2,6,3}} = (-7.930959)\ \emptyset\left(-12.53974 - 0.014906\ (37.21) - 7.930959\ (0.32) + 1.114637\ (13.06)\right)$$
$$\frac{\partial \Pr\left(y_{6,3} = 1 \middle| u_6, \boldsymbol{x_{6,3}}\right)}{\partial x_{2,6,3}} = -1.775128$$

The marginal effect value of -1.775128 indicates that a one-unit increase in the GII variable will reduce Capital Region of Jakarta Province to enter the category of provinces with YDI values that have reached the Kemenpora target of 1.775128. In addition, the calculation of the marginal effect for the EYS variable using the binary probit regression approach for panel data with random effects is as follows.

$$\frac{\partial \Pr\left(y_{6,3} = 1 \middle| u_6, \boldsymbol{x_{6,3}}\right)}{\partial x_{3,6,3}} = (1.114637\ )\ \emptyset\left(-12.53974 - 0.014906\ (37.21) - 7.930959\ (0.32) + 1.114637\ (13.06)\right)$$
$$\frac{\partial \Pr\left(y_{6,3} = 1 \middle| u_6, \boldsymbol{x_{6,3}}\right)}{\partial x_{3,6,3}} = 0.249481$$

The marginal effect value of 0.249481 suggests that a one-unit increase in the EYS variable will elevate the Capital Region of Jakarta Province to the category of provinces that meet the YDI target set by the Ministry of Youth and Sports, with a corresponding value of 0.249481.

V.CONCLUSIONS AND SUGGESTIONS

According to the best model analysis, binary probit regression with random effects for panel data proves to be superior in analyzing the factors influencing YDI in Indonesia. This model outperforms binary logit regression with random effects due to its higher classification accuracy and McFadden R-squared values. The simultaneous test indicates that at least one predictor variable significantly impacts the response variable. The significant predictors of YDI in Indonesia, as identified by binary probit regression with random effects, are the GII and EYS. In regions with a high GII, women have fewer opportunities to engage in economic and political development, fostering an environment that may hinder youth development. Additionally, as expected years of schooling increase, youth are afforded more time to acquire the knowledge and skills necessary to enhance their quality of life.

REFERENCES

[1] A. Agresti, Statistical methods for the social sciences, 5th ed. Pearson, 2018.

[2] A. Dewi Anggraeni Chairunnisa and A. Fauzan, “Implementation of Panel Data Regression in the Analysis of Factors Affecting Poverty Levels in Bengkulu Province in 2017-2020: Implementation of Panel Data Regression”, EKSAKTA: J. Sci. Data Anal., vol. 4, no. 1, pp. 40–45, Jan. 2023.

[3] A. Efendi, "Studi Korelasi Indeks Pembangunan Pemuda (IPP) dengan Indeks Pembangunan Manusia (IPM) di Indonesia," Jurnal Paradigma, vol. 9, no. 1, 2020.

[4] A. M. Rashed, M. A. Shalaby, and M. E. Khalil, "Youth Development Index: A holistic model of measuring youth development in Egypt," International Journal of Social Economics, vol. 48, no. 1, pp. 118-137, 2021. DOI: 10.1108/IJSE-07-2019-0428.

[5] A. M. Yolanda, A. Adnan, and A. A. Dwiputri, "Pemodelan Klasifikasi Pada Indeks Ketimpangan Gender (GII) Tahun 2020 dengan Metode Naïve Bayes," Jurnal Keluarga Berencana, vol. 7, no. 1, pp. 22–31, 2022. [Online]. Available: https://doi.org/10.37306/kkb.v7i1.118. [Accessed: July. 7, 2024].

[6] A. Stiglitz, "Panel data analysis: An application to the study of individual retirement decisions," Journal of Econometrics, vol. 223, no. 1, pp. 1-12, 2021. DOI: 10.1016/j.jeconom.2021.01.003.

[7] B. Baltagi, Econometric Analysis of Panel Data. Springer, 2021.

[8] Badan Pusat Statistik, Statistik Pemuda Indonesia 2023, vol. 21. Badan Pusat Statistik (BPS), 2023. [Online]. Available: https://www.bps.go.id/id/publication/2023/12/29/18781f394974f2cae5241318/statistik-pemuda-indonesia-2023.html. [Accessed: June. 23, 2024].

[9] D. N. Gujarati and D. C. Porter, Basic econometrics. McGraw-Hill/Irwin, 2009.

[10] I. A. Idrus, S. Anurlia, and D. Fadiyah, "Analysis of the impact of patriarchal culture on the role of women in politics and governance," JSIP: Jurnal Studi Ilmu Pemerintahan, vol. 4, no. 1, pp. 1–10, 2023.

[11] J. C. Timoneda, "Estimating group fixed effects in panel data with a binary dependent variable: How the LPM outperforms logistic regression in rare events data," Social Science Research, vol. 93, p. 102486, Jan. 2021. DOI: 10.1016/j.ssresearch.2020.102486.

[12] J. Johnston and J. E. Dinardo, Econometric methods. New York, Ny: Macgraw-Hill, 2009.

[13] J. S. Butler and R. Moffitt, "A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model," The Econometric Society, vol. 50, no. 3, pp. 761–764, 1982.

[14] Kemenpora, "Indeks Pembangunan Pemuda Naik Kemenpora: Modal Luar Biasa Untuk Optimalkan Bonus Demografi dengan Pemuda Unggul Menuju Indonesia Emas 2045," Kemenpora, 2023. [Online]. Available: https://www.kemenpora.go.id/detail/4183/indeks-pembangunan-pemuda-naik-kemenpora-modal-luar-biasa-untuk-optimalkan-bonus-demografi-dengan-pemuda-unggul-menuju-indonesia-emas-2045. [Accessed: June. 13, 2024].

[15] L. de Castro Romero, V. Martín Barroso, and R. Santero-Sánchez, "Does gender equality in managerial positions improve the gender wage gap? Comparative evidence from Europe," Economies, vol. 11, no. 12, p. 301, 2023. DOI: 10.3390/economies11120301.

[16] Larashati, "Ketimpangan Dan Peningkatan Kesetaraan Gender Dalam Sdgs (Sustainable Development Goals)," Jurnal Sains Edukatika Indonesia, vol. 4, no. 2, pp. 55–61, 2022.

[17] M. Bertocchi, G. Brunetti, and C. Torricelli, "Marriage and divorce: The role of income," Review of Economics of the Household, vol. 19, pp. 321-347, 2021. DOI: 10.1007/s11150-019-09477-x.

[18] N. Rahminawati, "Strategi Peningkatan Harapan Lama Sekolah (HLS) dan Rata-Rata Lama Sekolah (RLS)," TA'DIB: Jurnal Pendidikan Islam, vol. 12, no. 2, pp. 367–382, 2023. [Online]. Available: https://doi.org/10.29313/tjpi.v12i2.12852. [Accessed: Aug. 2, 2024].

[19] S. Hupkau and B. Petrongolo, "Work, care and gender during the COVID-19 crisis," Fiscal Studies, vol. 41, no. 3, pp. 623-651, 2020. DOI: 10.1111/1475-5890.12245.

[20] S. Akbari and J. Imani, "Application of Naive Bayes Classifier in Gender Classification of Speech Signals," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 9, pp. 3879-3888, 2021. DOI: 10.1109/TNNLS.2020.3048946.

[21] W. H. Greene, Econometric analysis, 5th ed., in Statistical Papers, vol. 4, pp. 983–984, 2023. [Online]. Available: https://doi.org/10.1007/s00362-010-0315-8. [Accessed: July. 3, 2024].