



Exploring Gender Development Inequality in Indonesia through Binary Logistic Regression

Sri Sulastri^{1*} , Nurhalisa¹ , and Ajriansyah¹ 

¹Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya (60111), Indonesia

Article Info

Article history:

Received: October 20th, 2025

Revised: November 25th, 2025

Accepted: December 28th, 2025

Published: January 18th, 2026

Published by:



Copyright ©2025 by Author(s)



Under the licence CC BY-NC-SA 4.0

ABSTRACT

The categorical data analysis method used to examine the influence of independent variables on the dependent variable is binary logistic regression. This analysis is applied when the dependent variable consists of two categories (dichotomous) with one or more independent variables of either categorical or continuous types. This study aims to identify the factors that have a significant influence on the Gender Development Index (GDI) at the district/city level in Indonesia in 2023, using data obtained from Statistics Indonesia (Badan Pusat Statistik, BPS). In addition, binary logistic regression analysis is employed to determine the model for Gender Development Index (GDI) cases. The results show that out of five independent variables, only four have a significant effect, namely Women as Professionals with an odds ratio (OR) of 1.031, Average Years of Schooling with an OR of 1.952, Expected Years of Schooling with an OR of 1.452, and Female Labor Force Participation Rate with an OR of 1.072. The independent variable with the most significant influence is average years of schooling, as it has the highest odds ratio value. This study is limited to the use of district/city-level data in Indonesia for the year 2023. The findings are particularly important because binary logistic regression has not previously been applied to analyze the Gender Development Index at the district/city level in Indonesia.

Keywords: Binary Logistic Regression Analysis, Gender Development Index, Odds Ratio.

*Corresponding Author: Email: srisulastri2580@gmail.com

1. Introduction

Binary logistic regression is a method used to describe the relationship between one or more independent variables and a dependent variable. The dependent variable in binary logistic regression is discrete, producing only two possible outcomes success and failure which are commonly represented as success ($Y = 1$) and failure ($Y = 0$) [1], [2]. One technique for modeling the relationship between a dichotomous response a nominal or ordinal variable with two categories and one or more predictor variables is binary logistic regression [3]. Logistic regression is conceptually similar to discriminant analysis, as both aim to test the probability of the occurrence of a dependent variable predicted by its independent variables. Binary logistic regression can be applied to various social research contexts [4]. In this study, it is employed to analyze the Gender Development Index (GDI).

The success of societal welfare is not solely measured by a high Human Development Index (HDI) but also by the quality of human resources. Gender equality is one of the key goals of the Sustainable Development Goals (SDGs) and plays a crucial role in advancing human development [5], [6]. This goal emphasizes the importance of ensuring that men and women have equal rights, opportunities, and access across various aspects of life. Several indicators can be used to evaluate the extent of gender equality or disparity in human development within a country or region. One

such indicator is the Gender Development Index (GDI)[7]. In Indonesia, the district/municipality with the highest GDI in 2023 was Tomohon City, North Sulawesi Province, with a GDI score of 99.32, while the lowest was Asmat District, Papua Province, with a GDI score of 55.99 [8], [9]. The substantial gap between the highest and lowest GDI values indicates that gender development across Indonesia remains uneven. Finding the elements that have a major impact on Indonesia's Gender Development Index is one way to address this problem.

Several studies have been conducted regarding the Gender Development Index (GDI). For instance, examined the factors influencing the GDI in West Sumatra Province using panel data regression analysis [10]. The findings demonstrated that the province's GDI was highly impacted by the variables Average Years of Schooling, Expected Years of Schooling, and Labor Force Participation Rate. Another study by [11] modeled the GDI in East Java Province using nonparametric spline regression. The findings indicated that the factors influencing the GDI in East Java included Female Primary School Participation Rate, Female Junior High School Participation Rate, Female Labor Force Participation Rate, Female Unemployment Rate, Female Morbidity Rate, and the Sex Ratio. Furthermore, conducted a study titled "Classification of the Gender Development Index in Indonesia in 2020 Using Supervised Machine Learning Algorithms" [12]. The results revealed that the variables significantly influencing the GDI were Senior High School Participation Rate, Percentage of Population with Health Complaints, Percentage of Female Civil Servants, Female Income Contribution, and the Sex Ratio. According to the literature evaluation, no research has yet been done in Indonesia that uses binary logistic regression to examine the Gender Development Index (GDI) at the district or municipality level. Thus, the purpose of this study is to create a binary logistic regression model for the GDI and determine the variables that have a major impact on the Gender Development Index in Indonesian districts and municipalities in 2023.

2. Method Details

2.1. Dataset

The dataset used in this study consists of one dependent variable and five independent variables obtained from secondary sources through official publications of Statistics Indonesia (Badan Pusat Statistik, BPS). The structure of the dataset is presented in Table 1.

Table 1. Research Variables

Variable	Measurements Scale (Unit)
Gender Development Index (Y)	Nominal (0 = Low GDI) (1 = High GDI)
Women's Income Contribution (X_1)	Ratio (%)
Women as Professional Workers (X_2)	Ratio (%)
Average Years of Schooling (X_3)	Ratio (Years)
Expected Years of Schooling (X_4)	Ratio (Years)
Female Labor Force Participation Rate (X_5)	Ratio (%)

The classification of a region's Gender Development Index (GDI) into high and low categories is based on the study by [12], which states that a GDI value of less than 90 is classified as low, while a GDI value of 90 or higher is classified as high. Subsequently, modeling was conducted to examine the effects of the independent variables—women's income contribution, women as professional workers, average years of schooling, expected years of schooling, and female labor force participation rate—on the dependent variable, namely the Gender Development Index (GDI).

2.2. Research Flow

2.2.1. Data Preparation and Tabulation

The initial step involved data tabulation to select relevant data and remove unnecessary observations. Subsequently, data type adjustments were performed, in which categorical variables were converted into factor variables to ensure their compatibility with statistical modeling.

2.2.2. Assumption Testing (Multicollinearity)

Prior to model estimation, an assumption test was conducted to examine multicollinearity using the Variance Inflation Factor (VIF). This test aimed to ensure that no significant linear correlation existed among the independent variables. In this study, all variables exhibited VIF values less than 10, indicating that the multicollinearity assumption was satisfied.

2.2.3. First-Stage Binary Logistic Regression Modeling

An initial model was constructed by including all independent variables (five variables) to identify the basic structure of the binary logistic regression model.

2.2.4. First-Stage Hypothesis Testing

At this stage, two types of hypothesis testing were conducted. The simultaneous test (G-test) was used to determine whether at least one independent variable had a significant effect on the model as a whole. The partial test (Wald test) was applied to assess the statistical significance of each independent variable individually.

2.2.5. Model Re-estimation and Second-Stage Testing

The model was re-estimated by including only the independent variables that were found to be significant in the previous stage. This stage again involved a second simultaneous test to validate the revised model and a second partial test to confirm that all remaining variables were statistically significant.

2.2.6. Model Goodness-of-Fit Test

After the final model was established, a goodness-of-fit test using the Hosmer–Lemeshow method was conducted to evaluate whether the model adequately explained the observed data. The model was well fitted if no significant difference was found between the observed values and the model predictions.

2.2.7. Odds Ratio Interpretation and Conclusion

The final step involved interpreting the model using odds ratios to assess the direction and magnitude of the effect of each independent variable on the likelihood of an increase in the Gender Development Index (GDI). The study concluded by identifying the most dominant influencing factors.

2.3. Binary Logistic Regression

A statistical technique for describing the relationship between one or more independent variables and a dependent variable is binary logistic regression. The dependent variable in binary logistic regression is discrete, producing only two possible outcomes success and failure which are commonly denoted as success ($Y = 1$) and failure ($Y = 0$) [1]

The logistic regression model with k explanatory variables is formulated in terms of the probability value $\pi(x) = E(Y = 1|x)$, $\pi(x)$, and is expressed as follows: [1]

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_k + \varepsilon)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_k + \varepsilon)} \quad (1)$$

where $\pi(x)$ represents the probability of a success event $0 \leq \pi(x) \leq 1$, β_0 denotes the intercept (a constant term), $\beta_1, \beta_2, \dots, \beta_p$ is the logistic regression coefficient, and x_1, \dots, x_k is the value of the independent variable.

To facilitate the estimation of the regression parameters, the probability function $\pi(x)$ in Equation (1) is transformed using the logit transformation, as follows [13], [14]:

$$\begin{aligned} \text{logit}[\pi(x)] &= g(x) \\ g(x) &= \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] \\ &= \ln \left[\frac{\frac{\exp(\beta_0 + \beta_1 x_{i2} + \beta_2 x_{i2} + \dots + \beta_r x_{ir})}{1 + \exp(\beta_0 + \beta_1 x_{i2} + \beta_2 x_{i2} + \dots + \beta_r x_{ir})}}{1 - \frac{\exp(\beta_0 + \beta_1 x_{i2} + \beta_2 x_{i2} + \dots + \beta_r x_{ir})}{1 + \exp(\beta_0 + \beta_1 x_{i2} + \beta_2 x_{i2} + \dots + \beta_r x_{ir})}} \right] \\ &= \ln [\exp(\beta_0 + \beta_1 x_{i2} + \beta_2 x_{i2} + \dots + \beta_r x_{ir})] \end{aligned}$$

$$\begin{aligned}
 g(x) &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_r x_{ir} \\
 &= [1 \ x_{i1} \ x_{i2} \ \dots \ x_{ir}]^T [\beta_0 \ \beta_1 \ \beta_2 \ \dots \ \beta_r] \\
 \text{logit}[\pi(x)] &= x_i^T \boldsymbol{\beta}
 \end{aligned} \tag{2}$$

where $g(x)$ denotes the link function of the logistic regression model, commonly referred to as the logit link; x represents the independent variable(s) of the model; and $\boldsymbol{\beta}$ denotes the vector of regression coefficient parameters.

After obtaining the expression for $\pi(x)$ via the logit transformation, the logistic regression model used in this study is given by:

$$\pi(x) = \frac{\exp[g(x)]}{1 + \exp[g(x)]} \tag{3}$$

The multicollinearity test is the assumption test that needs to be met in binary logistic regression. The Variance Inflation Factor (VIF) is a frequently employed technique for identifying the existence of multicollinearity among independent variables [15].

$$VIF_j = \frac{1}{1 - R_j^2}, \quad j = 1, 2, \dots, k \tag{4}$$

2.4. Simultaneous Test

The purpose of this test is to ascertain the statistical significance of the logistic regression model [14].

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1 : \text{at least one } \beta_p \neq 0, \quad p = 1, 2, \dots, p$$

The test statistic used is as follows:

$$G^2 = -2 \ln \left(\frac{L_1}{L_0} \right) \tag{5}$$

The test statistic G^2 follow a *chi-square* distribution with degrees of freedom (*df*). Reject H_0 at the significance level α if $G > \chi_{(\alpha, df)}^2$ or equivalently if the $p_{value} < \alpha$.

2.5. Partial Test

This test is used to determine whether a particular variable is appropriate to be included in the model or not [16], [17]. The hypotheses for this test are formulated as follows:

$$H_0 : \beta_p = 0$$

$$H_1 : \beta_p \neq 0, \text{ where } p = 1, 2, \dots, p$$

The Wald test statistic in Equation (5) is calculated using the following formula

$$W_i = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)}; \quad i = 1, 2, \dots, p \tag{6}$$

The goodness-of-fit test is used to assess whether the logistic regression model fits the observed data adequately. A model is considered a good fit if there is no significant difference between the observed outcomes and the predicted probabilities generated by the model [1].

H_0 : The observed and anticipated values do not differ, indicating that the model fits the data.

H_1 : The actual and anticipated values diverge, indicating that the model does not adequately describe the facts.

The test statistic:

$$\hat{C} = \sum_{k=1}^g \frac{(o_k - n_k \bar{\pi}_k)^2}{n_k \bar{\pi}_k (1 - \bar{\pi}_k)} \tag{7}$$

Based on the test statistic in Equation (6), the H_0 rejection criterion is if $\hat{C} > \chi^2_{\alpha(g-2)}$ with a significance level of (α) and degrees of freedom (db), where $db = g - 2$ represents the number of groups. The Odds Ratio is a measure used to determine the risk tendency of experiencing a particular event between one category and another. Thus, the Odds Ratio value is given by:

$$\varphi = \frac{\exp(\beta_0 + \beta_1)}{\exp(\beta_0)} = \exp(\beta_1) \tag{8}$$

3. Results and Discussion

In this section, the research results are presented and discussed comprehensively. The results may be displayed in the form of diagrams, graphs, tables, or other visual aids to facilitate readers understanding. The discussion may be divided into several sub-sections [18].

3.1. Descriptive Statistics

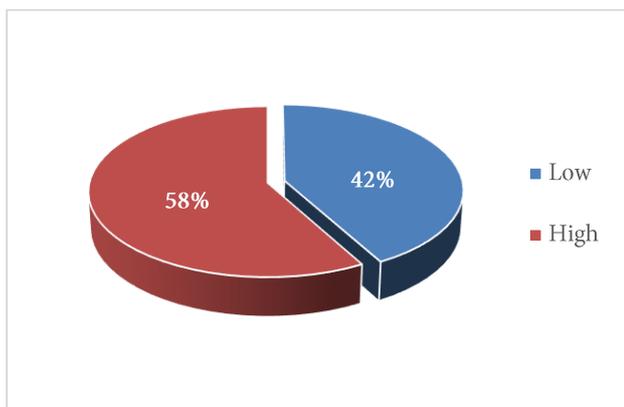


Fig. 1. Percentage of Gender Development Index

Based on Figure 1, it can be observed that 42% of regencies/municipalities in Indonesia, or a total of 214 regions, fall into the Low category. Meanwhile, 58%, or 300 regencies/municipalities, fall into the High category.

3.2. Assumption Testing

In this study, the multicollinearity assumption test was applied in the modeling process. The results of the multicollinearity test are presented [19] in Table 1.

Table 2. Variance Inflation Factor (VIF) Values of Independent Variables

Variable	VIF	Note
X_1	1,315982	No Multicollinearity Detected
X_2	1,254300	No Multicollinearity Detected
X_3	2,695079	No Multicollinearity Detected
X_4	2,766508	No Multicollinearity Detected
X_5	1,426279	No Multicollinearity Detected

The five independent variables in Table 1 all have VIF values below 10, which suggests that there is neither multicollinearity nor a significant linear correlation between them. This finding implies that the binary logistic regression assumption has been met, allowing the investigation to move on to the next phase, which is model creation.

3.3. Binary Logistic Regression Modelling

The initial Binary Logistic Regression model was used to identify the model structure obtained from the analysis before determining the best model with variables that significantly influence the Gender Development Index.

Table 3. Variables in the Equation

Variable	Coefficient (β)
Intercept	-15,649
X_1	0,003
X_2	0,031
X_3	0,670
X_4	0,368
X_5	0,068

From Table 2, the initial model obtained is as follows:

$$\begin{aligned}
 g(x) &= \ln \left[\frac{\pi(x)}{1-\pi(x)} \right] \\
 &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 \\
 &= -15.649 + 0.003x_1 + 0.031x_2 + 0.670x_3 + 0.368x_4 + 0.068x_5 \\
 \frac{\pi(x)}{1-\pi(x)} &= \exp(-15.649 + 0.003x_1 + 0.031x_2 + 0.670x_3 + 0.368x_4 + 0.068x_5) \\
 \hat{\pi}(x) &= \frac{\exp(-15.649 + 0.003x_1 + 0.031x_2 + 0.670x_3 + 0.368x_4 + 0.068x_5)}{1 + \exp(-15.649 + 0.003x_1 + 0.031x_2 + 0.670x_3 + 0.368x_4 + 0.068x_5)}
 \end{aligned}$$

3.4. First Stage Hypothesis Testing

The first stage of hypothesis testing was conducted to model the data and determine the effect of each variable that is statistically significant. Insignificant variables were excluded from the model.

3.4.1. Simultaneous Test

Table 3 displays the simultaneous test results based on the analysis's findings.

Table 4. Simultaneous Parameter Significance Test

	G	df	$\chi^2_{(\alpha,df)}$
Model	161,633	5	11,07

This simultaneous test uses a significance level of $\alpha = 0.05$ (5%) with the decision rule that H_0 is rejected if the significance value α and $G > \chi^2_{(\alpha,df)}$. Table 3 shows that the test statistic value obtained in the Logistic Regression model is $166.590 > \chi^2_{(\alpha,df)} = 11.07$. It can be concluded that, based on the simultaneous significance test, H_0 is rejected, which means that at least one independent variable has a significant effect on the Gender Development Index (GDI) at the district/city level in Indonesia. The next step is to conduct a partial test to determine which variables have a significant effect.

3.4.2. Partial Test

Based on the results of the analysis, the partial test results are presented in Table 4.

Table 5. Partial Parameter Significance Test

Parameter	p - value	Note
Intercept (β_0)	< 0,001	Significant
Women's Income Contribution (X_1)	0,827	Not Significant
Women as Professional Workers (X_2)	0,039	Significant
Average Years of Schooling (X_3)	< 0,001	Significant

Parameter	p - value	Note
Expected Years of Schooling (X_4)	0,012	Significant
Female Labor Force Participation Rate (X_5)	< 0,001	Significant

For the hypothesis test H_0 is rejected at a significance level of $\alpha < 0.05$. Based on the output shown in Table 4, the variables with $p\text{-value} < \alpha = 0.05$ are Women as Professional Workers (X_2), Average Years of Schooling (X_3), Expected Years of Schooling (X_4), and Female Labor Force Participation Rate (X_5). There is one independent variable that is not significant, namely Women’s Income Contribution (X_1). Therefore, the next step is to re-test and re-model by excluding the non-significant independent variable.

3.5. Second Binary Logistic Regression Modeling

The variable Women's Income Contribution was eliminated in the second modeling. Only the factors that significantly impacted the Gender Development Index (GDI) were kept when the model was rebuilt using the findings of the earlier investigation. According to the findings of this second modeling, Table 5 displays the variables that were part of the model.

Table 6. Variables in the Equation

Variable	Coefficient (β)
Intercept	-15,648
X_2	0,031
X_3	0,669
X_4	0,373
X_5	0,069

Based on Table 5, after excluding the variable, the second logistic regression model obtained is as follows:

$$\begin{aligned}
 g(x) &= \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] \\
 &= \beta_0 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 \\
 &= -15.648 + 0.031x_2 + 0.669x_3 + 0.373x_4 + 0.069x_5 \\
 \frac{\pi(x)}{1 - \pi(x)} &= \exp(-15.648 + 0.031x_2 + 0.669x_3 + 0.373x_4 + 0.069x_5) \\
 \hat{\pi}(x) &= \frac{\exp(-15.648 + 0.031x_2 + 0.669x_3 + 0.373x_4 + 0.069x_5)}{1 + \exp(-15.648 + 0.031x_2 + 0.669x_3 + 0.373x_4 + 0.069x_5)}
 \end{aligned}$$

3.6. Second Stage Hypothesis Testing

The second stage of hypothesis testing was conducted after the non-significant variable from the first model was removed.

3.6.1. Simultaneous Test

Based on the results of the analysis, the simultaneous test results are presented in Table 3.

Table 7. Second Simultaneous Parameter Significance Test

	G	df	$\chi^2_{(\alpha,df)}$
Model	161,585	4	9,48

This simultaneous test uses a significance level of $\alpha = 0.05$ or 5%, with the decision rule that H_0 is rejected if α or if $G > \chi^2_{(\alpha,df)}$. Table 6 shows that the resulting test statistic value in the Logistic Regression model is $166.585 > \chi^2_{(0.05,4)} = 9.48$. It can be concluded that for the simultaneous significance test, H_0 is rejected, which means that at least one independent variable has a significant effect on the Gender Development Index (GDI) at the regency/city level in Indonesia. Subsequently, a partial test is conducted to determine which variables have a significant effect.

3.6.2. Partial test

Based on the testing results, the partial test outcomes are presented in Table 7.

Table 8. Second Partial Parameter Significance Test

Parameter	p - value	Note
Intercept (β_0)	< 0,001	Significant
Women as Professional Workers (X_2)	0,039	Significant
Average Years of Schooling (X_3)	< 0,001	Significant
Expected Years of Schooling (X_4)	0,012	Significant
Female Labor Force Participation Rate (X_5)	< 0,001	Significant

The output results in Table 7 show that all variables have a $p - value < \alpha = 0.05$, leading to the conclusion that H_0 is rejected. This means the variables Professional Female Workforce(X_2), Mean Years of Schooling(X_3), Expected Years of Schooling(X_4), and Female Labor Force Participation Rate(X_5) are significant to the model and are deemed suitable for predicting the increase in the Gender Development Index in Indonesia. Subsequently, a model goodness-of-fit test was conducted.

3.7. Model Goodness-of-Fit Test

The goodness-of-fit test is used to assess whether the model is adequate or not. A model is considered adequate if there is no significant discrepancy between the observed outcomes and the model's predicted probabilities [1], [20].

Table 9. Results of the Model Goodness-of-Fit Test

$\chi^2_{(\alpha,df)}$	df	p - value
7,123	8	0,523

H_0 is rejected if the $p - value < \alpha(0.05)$. The obtained Chi-Square χ^2 value is 7.123 with a significance value $p - value = 0,523$. Since the $p - value > \alpha(0.05)$, it can be concluded that we fail to reject H_0 . This indicates that the resulting binary logistic regression model—specifically, the Gender Development Index (GDI) model—is influenced by the collective set of independent variables. In other words, the model fits well, as there is no significant difference between the observed outcomes and the model's predictions. This is evident from the result that the $p - value = 0,523$ is greater than 0.05. With a 95% confidence level, it can be asserted that the logistic regression model used is sufficiently capable of explaining the data (the model is adequate). Subsequently, the model was interpreted using the Odds Ratio.

The model results indicate that four variables are significant in demonstrating the relationship between the response variable and the independent variables. One variable, the X_1 variable, was not significant and was therefore excluded from the discussion. To determine the level of risk (propensity), the odds ratio values of the significant variables were examined at a 95% confidence level or a 5% error rate. The odds ratio results are as follows.

Table 10. Odds Ratio

Parameter	Odds Ratio
Women as Professional Workers (X_2)	1,031
Average Years of Schooling (X_3)	1,952

Expected Years of Schooling (X_4)	1,452
Female Labor Force Participation Rate (X_5)	1,072

Based on the output results from Table 4.14, the odds ratio for the Professional Female Workforce (X_2) is 1.031. This indicates that for each one-percentage point increase in the proportion of women in professional occupations in a regency/city in Indonesia, the propensity for the Gender Development Index in that region to be higher increases by a factor of 1.031 compared to regions with no increase. For the Mean Years of Schooling (X_3), the obtained odds ratio is 1.952. This means that for each additional year in the mean years of schooling in a regency/city, the propensity for the Gender Development Index in that region to be higher increases by a factor of 1.952 compared to regions with no increase in mean years of schooling.

Regarding the Expected Years of Schooling (X_4), the obtained odds ratio is 1.452. This signifies that for each additional year in the expected years of schooling in a regency/city, the propensity for the Gender Development Index in that region to be higher increases by a factor of 1.452 compared to regions with no increase in expected years of schooling. For the Female Labor Force Participation Rate (X_5), the obtained odds ratio is 1.072. This indicates that for each one-percentage point increase in the female labor force participation rate in a regency/city, the propensity for the Gender Development Index in that region to be higher increases by a factor of 1.072 compared to regions with no increase in the female labor force participation rate.

4. Conclusion

Based on the data analysis and discussion conducted in the modeling of the Gender Development Index (GDI) case at the regency/city level in Indonesia for the year 2023, which consisted of 514 data points, 1 dependent variable, and 5 independent variables using binary logistic regression analysis, the following conclusions are obtained:

1. The best model formed for the Gender Development Index case at the regency/city level in Indonesia for the year 2022 is :

$$\hat{\pi}(x) = \frac{e^{(-15,648+0,031X_2+0,669X_3+0,373X_4+0,069X_5)}}{1 + e^{(-15,648+0,031X_2+0,669X_3+0,373X_4+0,069X_5)}}$$

2. The factors that have a significant effect on the dependent variable at a significance level of $\alpha = 0.05$ or 5% are: Professional Female Workforce, Mean Years of Schooling, Expected Years of Schooling, and Female Labor Force Participation Rate. The obtained Odds Ratio values are as follows: Professional Female Workforce with an odds ratio of 1.031, Mean Years of Schooling with an odds ratio of 1.952, Expected Years of Schooling with an odds ratio of 1.452, and Female Labor Force Participation Rate with an odds ratio of 1.072. Mean Years of Schooling is the variable with the most significant influence on the Gender Development Index (GDI), as it possesses the largest odds ratio value.

Limitations

There are no specific limitations to report from this study.

Author Contributions Statement

All authors have made significant contributions to this research and fulfill the authorship criteria based on the CRediT (Contributor Roles Taxonomy) guidelines. The specific contributions of each author are as follows :

- Sri Sulastris
Conceptualization, Methodology Development, Formal Analysis, Writing – Original Draft
- Nurhalisa
Data Curation, Investigation, Visualization, Writing – Review & Editing
- Ajriansyah
Supervision, Validation, Writing – Review & Editing

Conflict of Interest Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

- The authors declare the following financial interests/personal relationships which may be considered as potential competing interests.

References

- [1] S. Hosmer, D. W., Lemeshow, "Hosmer_2000.pdf," 2000.
- [2] Z. Bursac, C. H. Gauss, D. K. Williams, and D. W. Hosmer, "Purposeful selection of variables in logistic regression," *Source Code Biol Med*, vol. 3, 2008, doi: 10.1186/1751-0473-3-17.
- [3] A. I. Sofiyat, A. Tjalla, and Mahdiyah, "Pemodelan Regresi Logistik Biner Terhadap Penerimaan Pegawai Di PT XYZ Jakarta," *Matematika Sains*, vol. 1, no. 1, 2023.
- [4] H. H. Dukalang, "Perbandingan Regresi Logistik Biner Dan Probit Biner dalam Pemodelan Tingkat Partisipasi Angkatan Kerja," *Euler: Jurnal Ilmiah Matematika, Sains dan Teknologi*, vol. 7, no. 2, 2019, doi: 10.34312/euler.v7i2.10355.
- [5] UNDP, "Human Development Report 2021/2022: uncertain times, unsettled lives: shaping our future in a transforming world.," 2001.
- [6] A. Karimah and H. Susanti, "Gender Inequality in Education and Regional Economic Growth in Indonesia," *Jurnal Ekonomi Pembangunan*, vol. 20, no. 1, 2022, doi: 10.29259/jep.v20i1.17841.
- [7] S. Suharnanik, S. Yuliani, and E. E. P. Laksminto, "Analysis of Gender Development Index to Prove Its Correlation with the Decrease of Poverty Line in Kediri City, Indonesia," *Woman in Russian Society*, no. 2, 2023, doi: 10.21064/WinRS.2023.2.3.
- [8] C. D. Anggraini and S. R. Faridatussalam, "The Effect of the Determinants of Gender Empowerment Index on Economic Growth in West Nusa Tenggara Period 2018–2021," 2024. doi: 10.2991/978-94-6463-204-0_70.
- [9] Y. Nuraeni and I. L. Suryono, "Analysis of Gender Equality on Employment in Indonesia," *Nakhoda: Jurnal Ilmu Pemerintahan*, vol. 20, no. 1, 2021.
- [10] I. Elisa, "Faktor-Faktor yang Mempengaruhi Indeks Pembangunan Gender (IPG) Provinsi Sumatera Barat Menggunakan Analisis Regresi Data Panel," *Journal of Mathematics UNP*, vol. 7, no. 2, 2022, doi: 10.24036/unpjomath.v7i2.12666.
- [11] I. N. Budiantara and N. Fajriyyah, "Pemodelan Indeks Pembangunan Gender dengan Pendekatan Regresi Nonparametrik Spline di Indonesia," *Jurnal Sains dan Seni ITS*, vol. 4, no. 2, 2015.
- [12] A. Indrasetyaningih, F. Fitriani, and P. J. Kusuma, "Klasifikasi Indeks Pembangunan Gender di Indonesia Tahun 2020 Menggunakan Supervised Machine Learning Algorithms," *Inferensi*, vol. 4, no. 2, 2021, doi: 10.12962/j27213862.v4i2.10940.
- [13] Risdayanti and M. Kasim Aidid, "Analisis Regresi Logistik Biner Untuk Menentukan Model Pengguna Kb Di Kelurahan Langnga Kabupaten Pinrang," vol. 1, no. 2, pp. 1–8, 2020.
- [14] S. Hosmer, D. W., Lemeshow, "Applied logistic regression (2nd ed.)," 2000, Wiley.
- [15] R. M. O'Brien, "A caution regarding rules of thumb for variance inflation factors," *Qual Quant*, vol. 41, no. 5, 2007, doi: 10.1007/s11135-006-9018-6.
- [16] Agresti, *An introduction to categorical data analysis (2nd ed.)*, vol. 28, no. 11. Wiley, 2007. doi: 10.1002/sim.3564.
- [17] A. Agresti and M. Kateri, "Categorical data analysis. In M. Lovric (Ed.)," in *International encyclopedia of statistical science*, Springer, 2011, pp. 206–208. doi: https://doi.org/10.1007/978-3-642-04898-2_175.
- [18] J. W. Creswell, "Research design Qualitative quantitative and mixed methods approach," *Research design Qualitative quantitative and mixed methods approach*, 2003, doi: 10.3109/08941939.2012.723954.
- [19] H. Midi, S. K. Sarkar, and S. Rana, "Collinearity diagnostics of binary logistic regression model," *Journal of Interdisciplinary Mathematics*, vol. 13, no. 3, 2010, doi: 10.1080/09720502.2010.10700699.
- [20] S. Menard, *Logistic Regression: From Introductory to Advanced Concepts and Applications*. 2014. doi: 10.4135/9781483348964.