BAREKENG: Journal of Mathematics and Its Applications
March 2026 Volume 20 Issue 1 Page 0397–0412
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol20iss1pp0397-0412

MODELING AND SEGMENTATION OF FACTORS AFFECTING HUMAN DEVELOPMENT IN ISLANDS OF JAVA USING FIMIX PLS METHOD WITH MEDIATION EFFECT

Muhammad Rosyid Ridho Az Zuhro[1], Ardi Kurniawan[2*], Dita Amelia[3], Idrus Syahzaqi[4]

[1,2,3,4]*Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga*
Jln. Dr. Ir. H. Soekarno, Mulyorejo, Surabaya, Jawa Timur, 60115, Indonesia

*Corresponding author's e-mail: *ardi-k@fst.unair.ac.id*

Article Info

Article History:
Received: 29th March 2025
Revised: 3rd May 2025
Accepted:17th June 2025
Available online: 24th November 2025

Keywords:
Finite mixture;
Gender equality;
Human development;
Inequality;
Mediation;
Segmentation.

ABSTRACT

Human development is a key indicator used to assess the quality of a country's human resources. Although Indonesia's HDI has experienced a significant increase of 75.02 in 2024, inequality is still a pressing issue, especially in terms of gender representation in the workforce. This study aims to identify the influence of poverty, economic, health, employment and education factors on human development in Java Island by considering gender equality as a mediating variable. The data used in the study is limited to 119 districts/cities in Java Island and sourced from BPS publications, the Health Office and the Education Office. The novelty of this study lies in the use of the Finite Mixture Partial Least Square (FIMIX-PLS) approach with mediation effects which is rarely applied in human development research in Indonesia, as well as allowing the identification of latent population heterogeneity and region-based segmentation. The results of this method reveal two distinct district/city segments in Java, with Segment 1 dominated by the variables in this study that have significant direct and indirect effects through the mediation of gender equality on human development, while Segment 2 has characteristics that emphasize the effect of gender equality. Given these differences in characteristics, it is important that contextual and regional segmentation-based development policies are designed by local and central governments. Statistical segmentation approaches such as FIMIX-PLS make a significant contribution to more targeted policy making. By changing the type of intervention according to specific problems, the government can allocate resources more effectively. This supports the achievement of SDG-10 in reducing inequality.

How to cite this article:

M. R. R. A. Zuhro, A. Kurniawan, D. Amelia and I. Syahzaqi., "MODELING AND SEGMENTATION OF FACTORS AFFECTING HUMAN DEVELOPMENT IN ISLANDS OF JAVA USING FIMIX PLS METHOD WITH MEDIATION EFFECT", *BAREKENG: J. Math. & App.*, vol. 20, iss. 1, pp. 0397-0412, Mar, 2026.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Human development is the main indicator used to assess the quality of a country's human resources. The Human Development Index (HDI) developed by the United Nations Development Program (UNDP) in 1990 became the main measure to assess human development achievements based on three main dimensions, namely health, education, and decent living standards [1]. Human development emphasizes the importance of community participation in the development process and everyone should have equal opportunities in achieving prosperity [2]. Globally, Indonesia experienced an increase in the HDI ranking, which previously ranked 114^{th} to 112^{th} out of 193 countries and regions with an HDI figure of 0.713 [3]. In addition, Indonesia also continues to experience an increase in HDI, where in 2024 it reached 75.02 with an increase of 0.63 points or 0.85 percent compared to the previous year. This trend shows a positive development with an average increase of 0.75 points per year from 2020 to 2024. This increase contributes to improved people's welfare, better access to education and health services, and a more decent standard of living. It also reflects Indonesia's commitment to achieving sustainable development targets that focus on improving people's quality of life.

Despite the progress in human development, the challenge of inequality is still a major issue to be faced. The UNDP's 2023 report shows that there are still disparities in human development in different parts of the world, with developed countries experiencing the highest levels of human development, while half of developing countries are regressing [4]. Indonesia itself faces challenges in creating equitable development, where the HDI value corrects to 0.588 when taking into account existing inequality, indicating that development progress is still not being enjoyed equally by all levels of society [4]. According to data collected by the Central Statistics Agency (BPS), the percentage of Indonesia's poor population will reach 9.36% by 2023, with access to education and health services still limited. This inequality is also evident in the expected years of schooling which hovers around 13.12 years, while the average years of schooling only reaches 8.54 years. Therefore, further efforts are needed to ensure that HDI growth is not just a number, but truly reflects inclusive and equitable welfare improvements [3]. This is also in line with the Sustainable Development Goals (SDGs) target point 10, which is to reduce inequality within and between countries by promoting inclusive economic growth and ensuring equal access to education, health and employment opportunities for all people [5].

Gender equality is an important aspect of social and economic development that focuses on providing fair rights and access for men and women in various aspects of life, such as education, employment, and participation in decision-making. The Indonesian government has committed to improving gender equality through various policies, such as Presidential Instruction No. 9 of 2000 on Gender Mainstreaming (PUG) in National Development, which regulates the increased role of women in all areas of social and state life and efforts to achieve gender equality, and Labor Law No. 13 of 2003 on employment, which regulates the rights of female workers. Data from the Central Statistics Agency (BPS) shows that Indonesia's Gender Development Index (GDI) continues to increase, from 90.82 in 2020 to 91.06 in 2023, indicating improvements in women's welfare compared to men. However, this increase in HDI has yet to address structural inequalities in the world of work and education, especially for women with low educational backgrounds. According to [6], strong gender equality can increase productivity, drive economic growth, and improve overall development effectiveness. Gender inequality can hinder the realization of full development potential. Gender inequality is considered an obstacle to economic growth, thus negatively affecting income at the national, regional and individual levels [7]. In the scope of education, it is known that the percentage of female workers with the highest education of junior high school reached 32.1%, while those with high school education amounted to 28.4%, indicating the need to increase access to higher education for women to increase their opportunities in the formal sector. Therefore, strengthening policies that focus on women's empowerment, access to higher education, and employment protection are important steps in realizing more optimal gender equality in Indonesia.

Java Island plays a crucial role in human development and gender equality in Indonesia, given its status as the center of economy, social, and government. With the largest population in Indonesia, 150.65 million people in 2023 according to data from the Central Statistics Agency (BPS), Java Island reflects complex development dynamics, including in the distribution of welfare and access to basic services. As the center of national growth, Java Island has a significant contribution in improving the quality of human resources through the education, health, and economic sectors. However, despite the general improvement in human development, inequality remains a major challenge, particularly in terms of gender equality. Gender inequality in Java remains an issue that needs attention. Although women's participation in various sectors continues to increase, disparities are still evident in the world of work. Based on BPS data in 2023, the number of female workers with labor status in Java Island reached 11.20 million people, while male workers with labor status reached 19.15 million people. This figure shows that women only account for 36.90% of total labor workers, signifying the dominance of male workers in the labor sector. This reflects the challenges in achieving gender equality, especially in employment opportunities and access to more strategic positions.

Based on previous research on human development, [8] conducted an Analysis of Factors Affecting Poverty Percentage and Human Development Index using multiple linear regression and simultaneous regression methods. The study shows that the human development index is significantly influenced by the percentage of poverty. Meanwhile, the percentage of poverty is influenced by the inflation rate and the percentage of per capita expenditure on food. Research by [9] used multiple linear regression analysis method to determine the influence of Gross Regional Domestic Product (GRDP), population, and poverty level on human development index. The results showed that the GRDP and population variables had a significant effect both partially and simultaneously on the human development index. Meanwhile, the poverty rate variable only has a significant effect simultaneously on the human development index. In addition, the results of research conducted by [10] using the ordinal logistic regression method showed a significant influence between the Percentage of Households with Defecation Facilities and the Labor Force Participation Rate on the Human Development Index.

One method that can be used to see the influence of many factors is Structural Equation Modeling (SEM). In multivariate analysis, SEM can be used to test and estimate the correlation between one or more dependent variables and various factors simultaneously [11]. In the early model of SEM, covariance-based SEM (CB-SEM) was developed which still considered the fulfillment of parametric distribution assumptions. Therefore, to overcome the limitations of fulfilling the assumptions in covariance-based SEM, a component or variant-based SEM called Partial Least Squares (PLS) was developed. The SEM-PLS or PLS Path Modeling (PLS-PM) approach is a more flexible and powerful method with a number of advantages, including the ability to test without a strong theoretical basis, the absence of distribution assumptions, the ability to use various types of indicator scales - nominal (categorical), ordinal, interval, ratio, and so on - and the requirement that the number of samples used must be small [12]. To determine the position of intervening constructs in the model, testing the effect of mediation is required. Mediator constructs have the power to strengthen or weaken the relationship between exogenous and endogenous constructs [13]. In addition, the SEM-PLS method can be developed into the Finite Mixture Partial Least Square (FIMIX-PLS) method as an approach that is able to overcome the problem of heterogeneity in the model by producing the best model segmentation through several procedures [14]. In addition, the FIMIX-PLS method is used to address models containing variables that cannot be measured directly. Based on previous research by [15], research was conducted on identifying factors that affect education in Indonesia. The use of FIMIX-PLS helps in grouping regions based on the similarity of educational factors that affect them, such as regional poverty, school outcomes, and educational facilities and infrastructure. In addition, research [16] related to the structure of poverty during the Covid-19 pandemic with the SEM-PLS and FIMIX-PLS methods which are divided into 2 segments based on the best results. The first segment shows that economic variables have a greater influence on poverty variables than other variables. While in the second segment, health variables have a greater influence on education.

The novelty used by researchers in this study is the use of the Finite Mixture Partial Least Square (FIMIX-PLS) method with mediation effects. This method is very suitable for looking at influences involving many factors and identifying population groups with different characteristics from the segmentation results. The FIMIX-PLS approach is more flexible than conventional regression methods, as it can capture heterogeneity in the relationship between variables across different population segments. By using the FIMIX-PLS approach with mediation effects, this study aims to provide deeper insights into the pattern of human development in Java Island and how poverty, economic, health, employment, education, and gender equality factors interact in the development process. In addition, by considering the mediating effect of gender equality, this study can provide a clearer picture of how gender plays a role in determining the quality of human development. Therefore, this research is expected to contribute to a better understanding of the dynamics of human development in Indonesia, and more specific and targeted policies can be formulated to improve HDI equally in Java.

2. RESEARCH METHODS

2.1 Structural Equation Modelling Partial Least Square

Structural Equation Modeling (SEM) is a multivariate analysis procedure that reveals the linear correlation between latent variables and indicators (observed variables), which cannot be measured simultaneously. [17]. SEM modeling is a further development of path analysis and this modeling determines the causal relationship between variables completely and simultaneously [18]. According to [19], some of the advantages of SEM are the ability to explain the relationship between complex variables and the direct and indirect effects of certain variables on other variables. Broadly speaking, there are two types of SEM, namely Covariance Based SEM (CB-SEM) and Variance Based SEM (VB-SEM) which includes Partial Least Square (PLS) [20].

Partial Least Square (PLS) is an alternative method in SEM that is used to test the relationship between latent constructs and a number of indicators simultaneously. PLS eliminates assumptions such as multivariate normal distribution in the data, does not require certain scale measurements in the data, and does not require a large number of samples so it is also referred to as soft modeling [21]. The PLS was created to overcome the limitations of Ordinary Least Squares (OLS) regression, which calculates a regression line by finding the minimum value of the sum of the squares of the errors between the actual and predicted values, in data that has specific difficulties such as small size, anomalous data, multicollinearity, or missing values. The existence of these problems can be solved by using the bootstrapping method [22].

2.2 PLS Model Specification

According to [12], the overall PLS path model specification consists of three sets of relationships, namely the inner model, outer model, and weight relation with latent variable values that can be estimated as follows.

2.2.1 Measurement Model (Outer Model)

The relationship between latent variables and their indicators is revealed by the measurement model. The measurement model includes two types of models: formative models and reflexive models. The reflexive model, also known as the principal factor model, occurs when latent variables influence indicators. Eqs. (1) and (2) for the reflective indicator model are as follows:

$$\boldsymbol{x} = \boldsymbol{\lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta} \tag{1}$$

$$\boldsymbol{y} = \boldsymbol{\lambda}_y \boldsymbol{\eta} + \boldsymbol{\varepsilon} \tag{2}$$

With $\boldsymbol{\xi}$ is a vector of exogenous latent variables, while η is a vector of endogenous latent variables. $\boldsymbol{x}$ and $\boldsymbol{y}$ are exogenous and endogenous indicators with sizes $q \times 1$ and $p \times 1$, respectively. $\boldsymbol{\lambda}_x$ and $\boldsymbol{\lambda}_y$ are factor loading matrices that describe the relationship between indicators and each latent variable. The $\boldsymbol{\delta}$ and $\boldsymbol{\varepsilon}$ are measurement model error vectors. Meanwhile, the formative model assumes that indicators affect latent variables. The direction of causality flows from the indicator to the latent variable. Eqs. (3) and (4) for the formative model are as follows:

$$\boldsymbol{\xi} = \boldsymbol{\Pi}_\xi \boldsymbol{x} + \boldsymbol{\delta}_\xi \tag{3}$$

$$\boldsymbol{\eta} = \boldsymbol{\Pi}_\eta \boldsymbol{y} + \boldsymbol{\delta}_\eta \tag{4}$$

$\boldsymbol{\Pi}_\xi$ and $\boldsymbol{\Pi}_\eta$ are loading matrices that represent multiple regression coefficients for exogenous and endogenous latent variables. Meanwhile, $\boldsymbol{\delta}_\xi$ and $\boldsymbol{\delta}_\eta$ are residuals from regression for exogenous and endogenous latent variables. The assumptions in the structural model are $E(\boldsymbol{\varepsilon}) = \mathbf{0}, E(\boldsymbol{\delta}) = \mathbf{0}$, $\boldsymbol{\varepsilon}$ is not correlated with $\boldsymbol{\eta}, \boldsymbol{\xi}$, and $\boldsymbol{\delta}$, and $\boldsymbol{\delta}$ is not correlated with $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$.

2.2.2 Structural Model (Inner Model)

The structural model or inner model describes the relationship between dependent latent variables (endogenous) and independent latent variables (exogenous). The structural model shows the relationship between latent variables which can also be referred to as a casual chain system relationship with Eq. (5) as follows:

$$\eta_j = \sum_{i=1, i\neq j}^{J} \beta_{ji}\eta_i + \sum_{i=1}^{J} \gamma_{ji}\xi_i + \zeta_j \tag{5}$$

with β_{ji} is the coefficient of influence of endogenous latent variables, γ_{ji} is the coefficient of influence of exogenous latent variables, and ζ is the residual error variable. The i and j are the range indexes of the number of latent variables. Based on Eq. (6), the structural model equation can be written as follows:

$$\boldsymbol{\eta} = \boldsymbol{B\eta} + \boldsymbol{\Gamma\xi} + \boldsymbol{\zeta} \tag{6}$$

$$\boldsymbol{\eta} = (\boldsymbol{I} - \boldsymbol{B})^{-1}(\boldsymbol{\Gamma\xi} + \boldsymbol{\zeta}) \tag{7}$$

The assumptions in the structural model are $E(\boldsymbol{\eta}) = \mathbf{0}, E(\boldsymbol{\xi}) = \mathbf{0}, E(\boldsymbol{\zeta}) = \mathbf{0}$ and $\boldsymbol{\zeta}$ is uncorrelated with $\boldsymbol{\xi}$ and $(\boldsymbol{I} - \boldsymbol{B})$ is a nonsingular matrix [23].

2.2.3 Weight Relation

Connecting weights are weights that connect the inner model with the outer model to estimate latent variables. One of the main characteristics of the PLS method is its ability to project latent variable score values through estimation which can be expressed as follows:

$$\hat{\eta}_j = \textstyle\sum_{m=1}^{M} \widetilde{w}_{jm} y_{jm} \tag{8}$$

$$\hat{\xi}_i = \textstyle\sum_{k=1}^{K} \widetilde{w}_{ik} x_{ik} \tag{9}$$

with $\hat{\eta}_j$ and $\hat{\xi}_i$ are latent variable scores. $\widetilde{w}_{jm}$ and $\widetilde{w}_{ik}$ are weights to estimate the exogenous and endogenous latent variables as linear combinations of the observed indicators.

2.3 PLS Parameter Estimation

In the Partial Least Square (PLS) parameter estimation stage, the least squares method and the three-stage PLS algorithm are used to obtain the model coefficients [23].

Stage 1: the weights required to provide the final estimate for each latent variable $\hat{\xi}_i = \sum_{k=1}^{K} \widetilde{w}_{ik} x_{ik}$ are calculated by the iteration process. Iteration starts initializing each latent variable by generating a linear combination of indicators in the measurement model or outside approximation to obtain weights that will be used in estimating latent variable scores, with the intention of increasing the variance as much as possible in indicators and latent variables [23].

$$y_i = \textstyle\sum_{k=1}^{K} \widetilde{w}_{ik} x_{ik} \tag{10}$$

With y_i is the score of the exogenous latent variable, and x_{ik} is the score on the k^{th} indicator of the i^{th} exogenous latent variable. Furthermore, the estimation of structural model parameters (inside approximation) assesses the relationship between latent variables to obtain initialization for each latent variable.

$$Z_j = \textstyle\sum_{i=1, i\neq j}^{J} e_{ji} y_i \tag{11}$$

With Z_j being the number of latent variables for $j = 1,2,\ldots,J$, e_{ji} is the structural model weight. After obtaining the estimated value of the structural model parameters, the internal estimate of Z_{ji} must be considered with respect to the indicator. Consideration of the value is done by updating the weights in the measurement model in the reflective indicator model for each weight w_{jk} is the regression coefficient of $\boldsymbol{Z_j}$ in the simple regression $\boldsymbol{x_{jk}}$, namely the simple regression $\boldsymbol{x_{jk}} = \boldsymbol{w_{jk}} Z_j$ with the following equation:

$$\boldsymbol{w_{jk}} = \left(\boldsymbol{Z_j' Z_j}\right)^{-1} \boldsymbol{Z_j' x_{jk}} \tag{12}$$

As for each iteration procedure, for example $S = 1,2,\ldots$, a convergence check is performed by comparing the outer weight at each step S against the outer weight $S-1$ with the convergence criterion $\left|\widetilde{w}_{jk}^{S-1} - \widetilde{w}_{jk}^{S}\right| < 10^{-5}$ **[23]**.

Stage 2: calculation of path coefficient estimation and loading. The path coefficient value of the structural model, $\boldsymbol{\hat{\beta}_{ji}}$ is estimated by OLS like multiple linear regression analysis.

$$y_j = \textstyle\sum_{i=1}^{I} \hat{\beta}_{ji} y_i \tag{13}$$

$$\hat{\beta}_{ji} = (\boldsymbol{y_i' y_i})^{-1} \boldsymbol{y_i' y_j} \tag{14}$$

Then the path coefficient value of the measurement model $\hat{\boldsymbol{\lambda}}_{jk}$ is estimated using the simple linear regression method.

$$\boldsymbol{x}_k = \hat{\boldsymbol{\lambda}}_{jk}\boldsymbol{y}_j \tag{15}$$

$$\hat{\lambda}_{jk} = \left(\boldsymbol{y}_j'\boldsymbol{y}_j\right)^{-1}\boldsymbol{y}_j'\boldsymbol{x}_k \tag{16}$$

Stage 3: calculate parameter estimates to produce means and constants (location parameters) for indicators and latent variables using a reflective model [24].

2.4 PLS Model Evaluation

2.4.1 Measurement Model Evaluation

To determine the validity and reliability of the model, an evaluation of its measurement is performed. To analyze the reflective measurement model, the factor loading value that was applied to evaluate convergent validity can be used. According to [12], confirmatory research, the loading factor value must be more than 0.7 while explanatory research, the loading factor value between 0.6 - 0.7 is still acceptable.

The AVE value has a value criterion that must be greater than 0.5, which means that 50% or more of the variance of the indicator can be explained [12]. The AVE calculation formula is:

$$AVE_i = \frac{\sum_{k=1}^{n} \lambda_{ik}^2}{\sum_{k=1}^{n} \lambda_{ik}^2 + \sum_{i=1}^{I} var(\varepsilon_i)} \tag{17}$$

The principle that tools used to measure different constructs should not have a high correlation is known as discriminant validity. Within a reflexive measurement model, it is possible to estimate the value of discriminant validity by taking into account the cross-loading presented by the indicators in each latent variable. It is important to achieve a cross-loading value higher than 0.7 for each variable.

In PLS, one of the reliability tests that can be carried out is using the composite reliability method or also known as composite reliability refers to a set of indicators used to measure a latent variable by evaluating internal consistency measures. The composite reliability process is carried out to ensure the accuracy, consistency, and reliability of the instrument in measuring the latent variable, with the reliability value expected to exceed 0.7, although the value of 0.6 is still acceptable [12]. The composite reliability value can be calculated using the following formula:

$$\rho_{c_i} = \frac{\left(\sum_{k=1}^{n} \lambda_{ik}\right)^2}{\left(\sum_{k=1}^{n} \lambda_{ik}\right)^2 + \sum_{i=1}^{I} var(\varepsilon_i)} \tag{18}$$

Cronbach's Alpha is used to calculate the lower limit of the reliability value of a latent variable. This value shows how reliable all the indicators in the model are. The Cronbach's Alpha value must be more than 0.7 for confirmatory research and Cronbach's Alpha more than 0.6 for explanatory research.

2.5 Hypothesis Testing

Hypothesis testing in PLS involves evaluating the λ parameter resulting from the measurement model, as well as the β and γ parameters obtained from the structural model. The hypotheses proposed are as follows:

2.5.1 Statistical Hypotheses for the Measurement Model

$H_0 : \lambda_{jk} = 0$ (loading factor is not significant in measuring latent variables)
$H_1 : \lambda_{jk} \neq 0$ (loading factor is significant in measuring latent variables)

with $j = 1,2,\ldots,J$ and j as the number of indicators, and $k = 1,2,\ldots,K$ and k as the number of exogenous latent variables.

2.5.2 Statistical Hypotheses for the Structural Model

$H_0 : \beta_{mn} = 0$ (endogenous latent variable is not significant in measuring other latent variables)
$H_1 : \beta_{mn} \neq 0$ (endogenous latent variables are significant in measuring other latent variables)

with n and m are the number of endogenous latent variables, $m = 1,2,\ldots,M$, $m \neq n$.

2.5.3 Statistical Hypotheses for Structural Models (Exogenous to Endogenous Latent Variables)

$H_0 : \gamma_{mn} = 0$ (exogenous latent variables are not significant in measuring other endogenous latent variables)
$H_1 : \gamma_{mn} \neq 0$ (exogenous latent variables are significant in measuring other endogenous latent variables)

Hypothesis testing is carried out using the t value where how to calculate t is the coefficient $\hat{\lambda}, \hat{\beta}, \hat{\gamma}$ obtained from the path coefficient of the endogenous latent variable and the exogenous latent variable divided by the standard error of the coefficient $\hat{\lambda}, \hat{\beta}, \hat{\gamma}$.

$$t_{\hat{\lambda}} = \frac{\hat{\lambda}}{SE(\hat{\lambda})} \tag{19}$$

$$t_{\hat{\beta}} = \frac{\hat{\beta}}{SE(\hat{\beta})} \tag{20}$$

$$t_{\hat{\gamma}} = \frac{\hat{\gamma}}{SE(\hat{\gamma})} \tag{21}$$

The decision-making criteria are based on the resulting t value. If the value of $t < t(\frac{\alpha}{2}, df)$ with df (degree of freedom) which is $(n-1)$ and n is many observations, then the decision taken is to accept H_0 [22].

2.6 Finite Mixture Partial Least Square

Segmentation is necessary because the structural model presents heterogeneity due to the fact that the sample comes from different populations; therefore, the FIMIX-PLS estimation is performed based on this hypothesis. [14]. FIMIX-PLS uses the results of latent variable estimation and modification of the inner model relationship:

$$\boldsymbol{B\eta_i} + \boldsymbol{\Gamma\xi_i} = \boldsymbol{\zeta_i} \tag{22}$$

The segment-specific heterogeneity of the model is concentrated on the estimation of relationships between latent variables [25]. Based on [26], segmentation is performed by assuming η_i is a finite mixture distribution of the conditional multivariate normal density function $f_{(i|k)}(\cdot)$ with $(K < \infty)$ segments which can be described as follows:

$$\boldsymbol{\eta_i} \sim \sum_{k=1}^{K} \rho_k \left[\frac{|\boldsymbol{B}_k|}{\sqrt[M]{2\pi}\sqrt{|\boldsymbol{\Psi}_k|}} e^{-\frac{1}{2}(\boldsymbol{B_k\eta_i} + \Gamma\xi_i)'\Psi_k^{-1}(\boldsymbol{B_k\eta_i} + \Gamma\xi_i)} \right] \tag{23}$$

with ρ_k is the mixing proportion of latent class k, where $\rho_k > 0$ and$\sum_{k=1}^{K} \rho_k = 1$. Then, $\boldsymbol{B_k}$ is the path coefficient matrix between endogenous latent variables, $\boldsymbol{\Gamma_k}$ is the path coefficient matrix between exogenous latent variables, and $\boldsymbol{\Psi_k}$ is the variance matrix of each regression of the inner model for latent class k.

Researchers can compare CAIC (Consistent AIC) and BIC (Bayesian Information Criterion) values to choose segments [27]. The lowest value in the comparison of segment classes indicates the selection of the best-fitting class, taking CAIC and BIC values as a reference [28]. Based on [29], the formula used to calculate the CAIC and BIC criteria is as follows:

$$BIC_k = -2\ln L + \ln N_k \tag{24}$$

$$CAIC_k = -2\ln L + 2N_k + \frac{2N_k(N_k+1)}{n - N_k - 1} \tag{25}$$

where L is the likelihood function of η, N_k is the total number of parameters for the k^{th} segment for $k = 1,2,\ldots,K$. While n is the number of observations.

Likewise, [30] proposed the use of class selection criteria based on normalized entropy (EN). EN is used to examine the results of FIMIX-PLS class specifications, which range from 0 to 1. The EN value tends to approach 1 when it is higher, indicating that the model has better separation quality and can be interpreted. The EN calculation formula is as follows:

$$EN_k = 1 - \frac{\left[\sum_{i=1}^{I} \sum_{k=1}^{K} -P_{ik} \ln(P_{ik})\right]}{I \ln(K)} \tag{26}$$

where P_{ik} is the probability of the i^{th} observation.

2.7 Research Data Sources

The data used in this study are secondary data obtained from publications by the Central Statistics Agency which can be accessed at bps.go.id, the publication of health profiles for each province by the Health Office, and the Education Data Portal website which can be accessed through data.dikdasmen.go.id in 2023. Data were taken from 119 districts/cities from 6 provinces in Java Island, namely East Java, Central Java, West Java, Banten, DKI Jakarta, and DI Yogyakarta. The latent variables used consist of an endogenous variable Human Development (η_1), a mediating variable Gender Equality (η_2) with 5 exogenous variables Poverty (ξ_1), Economy (ξ_2), Health (ξ_3), Employment (ξ_4), and Education (ξ_5). Details of the indicators of each latent variable and the research path diagram used are as follows:

Table 1. Latent Variables and Indicators

Latent Variables	Indicator	Label
Human Development (η_1)	Expenditure per Capita	$y_{1.1}$
	Average Years of Schooling	$y_{1.2}$
	Expected Years of Schooling	$y_{1.3}$
	Life Expectancy	$y_{1.4}$
	Population Growth Rate	$y_{1.5}$
Gender Equality (η_2)	Sex Ratio (Soul)	$y_{2.1}$
	Gender Development Index	$y_{2.2}$
	Percentage of female workers with labor status (%)	$y_{2.3}$
	Percentage of Female Workers with the Highest Education at Junior High School (%)	$y_{2.4}$
	Percentage of Female Workers with Highest Education (%)	$y_{2.5}$
Poverty (ξ_1)	Percentage of Poor Population (%)	$x_{1.1}$
	Poverty Depth Index	$x_{1.2}$
	Poverty Severity Index	$x_{1.3}$
	Poverty Line (Thousand Rupiah)	$x_{1.4}$
	Minimum Wage (Thousand Rupiah)	$x_{1.5}$
Economy (ξ_2)	Average Wage/Net Salary of Workers/Employees (Thousand Rupiah)	$x_{2.1}$
	Average Net Wage/Wage of Informal Workers by Main Industry Employment (Thousand Rupiah)	$x_{2.2}$
	Average Net Wage/Wage of Informal Workers by Main Service Employment (Thousand Rupiah)	$x_{2.3}$
	Average Monthly Expenditure per Capita on Food (Thousand Rupiah)	$x_{2.4}$
	Average Monthly Non-Food Expenditure per Capita (Thousand Rupiah)	$x_{2.5}$
	GRDP at Constant Price (ADHK) (Million Thousand Rupiah)	$x_{2.6}$
Health (ξ_3)	Percentage of Households with Access to Proper Sanitation (%)	$x_{3.1}$
	Percentage of Households with Access to Adequate Drinking Water (%)	$x_{3.2}$
	Ratio of Medical Personnel per 1,000 Population	$x_{3.3}$
	Ratio of Health Center Availability per 100,000 Population	$x_{3.4}$
	Ratio of Hospital Beds per 1,000 Population	$x_{3.5}$
Employment (ξ_4)	Labor Force Participation Rate - TPAK (%)	$x_{4.1}$
	Open Unemployment Rate - TPT (%)	$x_{4.2}$
	Percentage of Population Aged 15 Years & Over Employed in the Agricultural Sector (%)	$x_{4.3}$
	Percentage of Population Aged 15 Years & Over Employed in the Industrial Sector (%)	$x_{4.4}$
	Percentage of Population Aged 15 Years & Over Employed in the Services Sector (%)	$x_{4.5}$
	Percentage of Population Aged 15 Years & Over Employed with Primary Employment Status Laborer (%)	$x_{4.6}$
Education (ξ_5)	Literacy Rate (%)	$x_{5.1}$
	Junior High School Teacher-Student Ratio	$x_{5.2}$
	Teacher-Student Ratio of Senior High School	$x_{5.3}$
	Junior High School Pupil-School Ratio	$x_{5.4}$
	High school student-to-school ratio	$x_{5.5}$

Data Sources: Central Bureau of Statistics (BPS), Health Office, Education Office

The latent variables and indicators used in the research are depicted in the path diagram in Fig. 1. as follows:

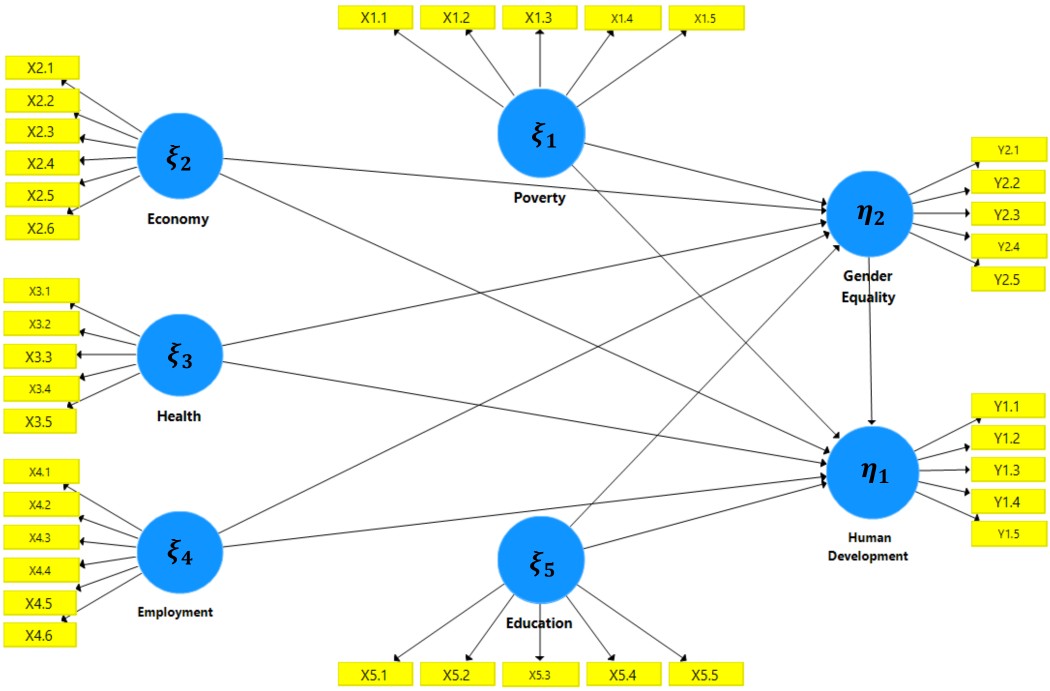

Figure 1. Research Model Path Diagram

2.8 Data Analysis Steps

The analysis steps in this study are described as follows:

1. Collecting research data from all provinces on the island of Java totaling 119 districts/cities within the 6 observed provinces, namely East Java, Central Java, West Java, Banten, DKI Jakarta, and DI Yogyakarta.
2. Analyze descriptive statistics of the collected data.
3. Modeling the effect of poverty, economy, health, employment, and education on human development mediated by gender equality in each district/city in Java Island based on the Finite Mixture Partial Least Square method with mediation effects with the following steps:
   a. Develop a research path diagram to illustrate the relationship pattern between endogenous and exogenous latent variables, as well as between latent variables and their respective indicators.
   b. Evaluate the measurement model and structural model.
   c. Determine the measurement model equation system and structural model.
   d. Verification of model validity using the Goodness of Fit (GoF) Index to assess whether the measurement and structural models are suitable.
   e. Conduct hypothesis testing for the measurement model and structural model based on Eqs. (21), (22), and (23).
4. Performing data segmentation with FIMIX-PLS needs to be done through several stages as follows:
   a. Determine the best number of segments by comparing the BIC, CAIC, and EN values. The best number of segments is the segment division that has the smallest BIC and CAIC values, and the largest EN value.
   b. Determine the segment members of the segment divisions formed.
   c. Evaluate the segmentation results with FIMIX-PLS including the influence of variables through path coefficients.
5. Draw conclusions and formulate suggestions from the results of the data analysis that has been carried out.

3. RESULTS AND DISCUSSION

3.1 Evaluation of Measurement Model

The purpose of evaluating the measurement model is to characterize the connection between latent variables and their indicators through reliability and validity assessments. In this research, the discriminant validity of the model was the basis for evaluating the validity tests. This test can be measured based on the outer loading and AVE values. Table 2. shows the results which indicate that several indicators, namely indicators $y_{1.5}, y_{2.1}, y_{2.4}, x_{1.4}, x_{1.5}, x_{3.1}, x_{3.2}, x_{3.4}, x_{4.1}, x_{4.2}, x_{4.3}, x_{4.4}, x_{5.3}$ must be eliminated according to the criteria because they are worth less than 0.6 which is considered not to have a sufficient impact on the model.

Table 2. Outer Loadings Value of Latent Variables

Variables	Indicator	Outer Loading	Modification
Human Development (η_1)	$y_{1.1}$	0.832	0.841
	$y_{1.2}$	0.945	0.945
	$y_{1.3}$	0.805	0.788
	$y_{1.4}$	0.680	0.703
	$y_{1.5}$	0.278	eliminated
Gender Equality (η_2)	$y_{2.1}$	-0.234	eliminated
	$y_{2.2}$	0.780	0.802
	$y_{2.3}$	0.896	0.931
	$y_{2.4}$	-0.483	eliminated
	$y_{2.5}$	0.935	0.942
Poverty (ξ_1)	$x_{1.1}$	0.919	0.938
	$x_{1.2}$	0.896	0.987
	$x_{1.3}$	0.816	0.941
	$x_{1.4}$	-0.757	eliminated
	$x_{1.5}$	-0.691	eliminated
Economy (ξ_2)	$x_{2.1}$	0.918	0.919
	$x_{2.2}$	0.739	0.741
	$x_{2.3}$	0.935	0.937
	$x_{2.4}$	0.921	0.919
	$x_{2.5}$	0.899	0.896
	$x_{2.6}$	0.761	0.765
Health (ξ_3)	$x_{3.1}$	0.590	eliminated
	$x_{3.2}$	0.246	eliminated
	$x_{3.3}$	0.920	0.965
	$x_{3.4}$	0.247	eliminated
	$x_{3.5}$	0.891	0.952
Employment (ξ_4)	$x_{4.1}$	-0.627	eliminated
	$x_{4.2}$	0.569	eliminated
	$x_{4.3}$	-0.974	eliminated
	$x_{4.4}$	0.104	eliminated
	$x_{4.5}$	0.921	0.928
	$x_{4.6}$	0.897	0.922
Education (ξ_5)	$x_{5.1}$	0.711	0.706
	$x_{5.2}$	0.797	0.810
	$x_{5.3}$	0.450	eliminated
	$x_{5.4}$	0.902	0.901
	$x_{5.5}$	0.717	0.716

Furthermore, the measurement of convergence validity can also be seen from the AVE value. The following in Table 3 presents the AVE value of each latent variable as follows.

Table 3. Average Variance Extracted Value

Variables	Average Variance Extracted	Description
η_1	0.679	Valid
η_2	0.799	Valid
ξ_1	0.913	Valid
ξ_2	0.751	Valid
ξ_3	0.918	Valid

Variables	Average Variance Extracted	Description
ξ_4	0.855	Valid
ξ_5	0.619	Valid

According to the results of the study in Table 3, it can be seen that the AVE value of each latent variable is greater than 0.5. This indicates that all the variables investigated have adequate convergent validity.

In addition, testing the validity of the model can also be tested with discriminant validity, which is shown in the measurement of certain latent variables by their indicators must have a higher correlation value than other latent variables. These measurements are evaluated through cross loading which can be seen in Table 4 below.

Table 4. Cross Loading Value

Variables	Indicator	η_1	η_2	ξ_1	ξ_2	ξ_3	ξ_4	ξ_5
(η_1)	y_1	**0.841**	0.745	-0.474	0.765	0.556	0.698	0.411
	y_2	**0.945**	0.905	-0.684	0.719	0.744	0.897	0.642
	y_3	**0.788**	0.600	-0.361	0.306	0.646	0.571	0.267
	y_4	**0.703**	0.649	-0.466	0.304	0.463	0.475	0.556
(η_2)	z_2	0.698	**0.802**	-0.400	0.305	0.489	0.515	0.524
	z_3	0.816	**0.931**	-0.636	0.745	0.557	0.921	0.653
	z_5	0.869	**0.942**	-0.626	0.627	0.726	0.865	0.653
(ξ_1)	$x_{1.1}$	-0.690	-0.699	**0.938**	-0.617	-0.478	-0.736	-0.587
	$x_{1.2}$	-0.554	-0.577	**0.987**	-0.439	-0.380	-0.574	-0.524
	$x_{1.3}$	-0.486	-0.500	**0.941**	-0.317	-0.340	-0.473	-0.464
(ξ_2)	$x_{2.1}$	0.574	0.581	-0.504	**0.919**	0.312	0.683	0.312
	$x_{2.2}$	0.283	0.308	-0.262	**0.741**	0.092	0.420	0.130
	$x_{2.3}$	0.501	0.510	-0.401	**0.937**	0.240	0.598	0.237
	$x_{2.4}$	0.608	0.584	-0.414	**0.919**	0.431	0.717	0.310
	$x_{2.5}$	0.824	0.767	-0.541	**0.896**	0.583	0.817	0.469
	$x_{2.6}$	0.447	0.421	-0.354	**0.765**	0.285	0.385	0.259
(ξ_3)	$x_{3.3}$	0.760	0.686	-0.435	0.504	**0.965**	0.700	0.437
	$x_{3.5}$	0.646	0.587	-0.378	0.277	**0.952**	0.605	0.464
(ξ_4)	$x_{4.5}$	0.810	0.797	-0.583	0.775	0.745	**0.928**	0.575
	$x_{4.6}$	0.712	0.830	-0.598	0.578	0.516	**0.922**	0.617
(ξ_5)	$x_{5.1}$	0.576	0.606	-0.661	0.537	0.452	0.690	**0.706**
	$x_{5.2}$	0.287	0.440	-0.412	0.329	0.151	0.414	**0.810**
	$x_{5.4}$	0.538	0.639	-0.386	0.182	0.498	0.523	**0.901**
	$x_{5.5}$	0.290	0.370	-0.174	-0.042	0.237	0.264	**0.716**

Meanwhile, reliability testing can be determined through the composite reliability and Cronbach's alpha values. According to the findings in Table 5, each of the latent variables has a composite reliability greater than 0.7 and a Cronbach's alpha exceeding 0.6. Therefore, it can be stated that each indicator is reliable and demonstrates accuracy, stability, and precision when measuring latent variables.

Table 5. Reliability Value

Variables	Composite Reliability	Cronbach's Alpha	Description
η_1	0.893	0.838	Reliable
η_2	0.922	0.874	Reliable
ξ_1	0.969	0.953	Reliable
ξ_2	0.947	0.934	Reliable
ξ_3	0.957	0.912	Reliable
ξ_4	0.922	0.830	Reliable
ξ_5	0.866	0.796	Reliable

3.2 Structural Model Evaluation

The coefficient of determination, also known as R-squared (R^2), is the tool used to examine the relationship between latent variables in the structural model. The results of the analysis related to the coefficient of determination are presented in Table 6 below.

Table 6. Coefficient of Determination

Variables	R-Square
Human Development	0,854
Gender Equality	0,811

Based on the results of the analysis, it is found that the R^2 value for the human development variable is 0.854, which means that the diversity of human development variables can be explained by the variable factors of gender equality (η_2), poverty (ξ_1), economy (ξ_2), health (ξ_3), employment (ξ_4), and education (ξ_5) in this study by 85.4% while the remaining 14.6% is explained by other factors outside this study. Then, the gender equality variable obtained an R^2 value of 0.811, which means that 81.1% of the gender equality variable can be explained by the variable factors in this study and the remaining 18.9% is explained by other factors outside the study. According to [12], R-Square results above 0.67 are categorized as a substantial or good model.

Based on the R^2 value in the coefficient of determination and AVE in the previous measurement, this value can be used to measure Goodness of Fit (Gof) which is used to evaluate the structural and measurement models as a whole. The calculation formula is as follows.

$$GoF = \sqrt{\overline{com} \times \overline{R^2}}$$

$$GoF = \sqrt{\left(\frac{0,679 + 0,799 + 0,913 + 0,751 + 0,918 + 0,855 + 0,619}{7}\right) \times \left(\frac{0,854 + 0,811}{2}\right)}$$

$$GoF = 0,808$$

The model has a high capacity to explain the empirical data, as the calculated Gof value is 0.808, which is in the high category. In general terms, it can be argued that the model developed is suitable for the research.

In this investigation, the path coefficient between latent variables was also used to evaluate the structural model. The measurements are direct and indirect measurements through the gender equality variable as a mediating variable. The measurement results are presented in Table 7 as follows.

Table 7. Path Coefficient between Latent Variables

Variables	**Path Coefficient**	**Estimated Path Coefficient**	**T Statistics (|O/STDEV|)**	**P Values**	**Description**
$\eta_2 \rightarrow \eta_1$	β_{21}	0.663	9.580	0.000	significant
$\xi_1 \rightarrow \eta_1$	γ_{11}	-0.084	1.844	0.065	not significant
$\xi_1 \rightarrow \eta_2$	γ_{21}	-0.069	1.336	0.182	not significant
$\xi_2 \rightarrow \eta_1$	γ_{12}	0.164	2.341	0.019	significant
$\xi_2 \rightarrow \eta_2$	γ_{22}	0.075	1.235	0.217	not significant
$\xi_3 \rightarrow \eta_1$	γ_{13}	0.276	3.656	0.000	significant
$\xi_3 \rightarrow \eta_2$	γ_{23}	0.127	2.778	0.006	significant
$\xi_4 \rightarrow \eta_1$	γ_{14}	-0.087	0.716	0.474	not significant
$\xi_4 \rightarrow \eta_2$	γ_{24}	0.567	6.907	0.000	significant
$\xi_5 \rightarrow \eta_1$	γ_{15}	-0.056	1.117	0.264	not significant
$\xi_5 \rightarrow \eta_2$	γ_{25}	0.197	4.136	0.000	significant

Based on the results of direct measurement in Table 7, 6 paths are obtained that have a significant effect. The value obtained is greater than $t_{\left(\frac{\alpha}{2};n-1\right)} = t_{(0.025;118)}$ which is 2.270. Thus, it can be decided to reject H_0, which means that the variables of gender equality (η_2), economy (ξ_2), health (ξ_3) significantly affect human development (η_1). Then, the variables of health (ξ_3), employment (ξ_4) and education (ξ_5) have a significant effect on human development (η_1). gender equality (η_2). The indirect measurement in the model is presented in Table 8 as follows.

Table 8. Path Coefficient between Latent Variables with Mediation Effect

Variables	Estimated Path Coefficient	T Statistics |O/STDEV|)	P Values	Description
$\xi_1 \to \eta_2 \to \eta_1$	0.050	1.341	0.181	not significant
$\xi_2 \to \eta_2 \to \eta_1$	-0.046	1.268	0.205	not significant
$\xi_3 \to \eta_2 \to \eta_1$	0.084	2.629	0.009	significant
$\xi_4 \to \eta_2 \to \eta_1$	0.376	5.663	0.000	significant
$\xi_5 \to \eta_2 \to \eta_1$	0.130	3.342	0.001	significant

Based on the analysis shown in Table 8, the 3 paths obtained through the mediation variable have a t_{count} value greater than $t_{\left(\frac{\alpha}{2};n-1\right)} = t_{(0.025;118)}$ which is 2.270. So it can be decided to reject H_0. Thus, the variables of health (ξ_3), employment (ξ_4) and education (ξ_5) have a significant effect on human development (η_1) with gender equality (η_2) as a mediating variable can be accepted. The overall structural equation model can be written as follows.

$$\eta_1 = 0{,}663\eta_2 + 0164\xi_2 + 0{,}36\xi_3 + 0{,}376\xi_4 + 0{,}13\xi_5$$

$$\eta_2 = 0{,}127\xi_3 + 0{,}567\xi_4 + 0{,}197\xi_5$$

3.3 Finite Mixture Partial Least Square

Segmentation was conducted based on the assumption of heterogeneity that might occur considering that each district and city has its own characteristics, government, and policies. In the segmentation process, what needs to be done first is to determine the best number of segments through the comparison of BIC, CAIC, and EN values. The results obtained are presented in Table 9 with K as the number of segment divisions as follows:

Table 9. Segment Criteria Comparison

K	lnL	BIC	CAIC	EN
2	-99.095	327.225	354.225	0.697
3	-70.661	337.266	378.266	0.781
4	-55.714	374.279	429.279	0.789
5	-24.805	379.37	448.37	0.846
6	4.940	386.786	469.786	0.888

Based on Table 9, the results of the comparison on $k = 2, 3, 4, 5$, and 6. The results of the comparison obtained by the division of segments as many as two segments ($k = 2$) have the smallest BIC and CAIC values, which means that the division of segments into two segments is the best segment division for the model, although the EN value shows a value that is contrary to BIC and CAIC. The effect of each variable on other variables is presented in Table 10 below.

Table 10. Segment Grouping Based on FIMIX-PLS Results

Path	Segment	
	1	2
$\eta_2 \to \eta_1$	0.613	0.281
$\xi_2 \to \eta_1$	0.410	-0.466
$\xi_3 \to \eta_1$	0.389	0.346
$\xi_3 \to \eta_2$	0.148	0.169
$\xi_4 \to \eta_2$	0.489	0.734
$\xi_5 \to \eta_2$	0.180	0.279
$\xi_3 \to \eta_2 \to \eta_1$	0.091	0.048
$\xi_4 \to \eta_2 \to \eta_1$	0.300	0.206
$\xi_5 \to \eta_2 \to \eta_1$	0.110	0.078

Based on Table 10, it shows that segment 1 has characteristics dominated by the variables in this study that have a significant effect both directly and indirectly through the mediation of gender equality on human development with an estimated value in segment 1 greater than in segment 2. Meanwhile, segment 2 has characteristics that emphasize the influence of gender equality with an estimated value in segment 2 greater than in segment 1. The results of district/city segmentation based on FIMIX-PLS can be seen in Table 11 below.

Table 11. District/City Grouping for Each Segment

Segment	Number of Members	Districts/cities
1	99	Pacitan, Ponorogo, Trenggalek, Tulungagung, Blitar, Kediri, Malang, Jember, Banyuwangi, Bondowoso, Situbondo, Probolinggo, Pasuruan, Mojokerto, Jombang, Nganjuk, Madiun, Magetan, Ngawi, Bojonegoro, Tuban, Lamongan, Gresik, Sampang, Pamekasan, Sumenep, Kediri City, Blitar City, Probolinggo City, Pasuruan City, Mojokerto City, Madiun City, Surabaya City, Batu City, Cilacap, Banyumas, Purbalingga, Banjarnegara, Purworejo, Wonosobo, Magelang, Boyolali, Wonogiri, Sragen, Grobogan, Blora, Rembang, Pati, Kudus, Jepara, Demak, Semarang, Temanggung, Kendal, Batang, Pekalongan, Pemalang, Tegal, Brebes, Magelang City, Surakarta City, Pekalongan City, Tegal City, Kulonprogo, Gunungkidul, Bogor, Cianjur, Garut, Bandung, Tasikmalaya, Cirebon, Ciamis, Sumedang, Kuningan, Majalengka, Indramayu, Purwakarta, Bandung, West, Pangandaran, Bogor City, Cirebon City, Sukabumi City, Bandung City, Bekasi City, Cimahi City, Banjar City, Depok City, Tasikmalaya City, Thousand Islands, South Jakarta City, East Jakarta City, North Jakarta City, West Jakarta City, Pandeglang, Lebak, Cilegon City, Tangerang, Serang City, Tangerang City
2	20	Lumajang, Sidoarjo, Bangkalan, Malang City, Kebumen, Klaten, Sukoharjo, Karanganyar, Salatiga City, Semarang City, Bantul, Sleman, Yogyakarta City, Sukabumi, Subang, Karawang, Bekasi, Central Jakarta City, Serang, South Tangerang City

Based on the interpretation in Table 10 and the FIMIX-PLS segmentation results in Table 11, segment 1 consists of 99 districts/cities with characteristics dominated by the variables in this study that have a significant effect both directly and indirectly through the mediation of gender equality on human development. This indicates that the regions in this segment have a better level of development in terms of economy, health, employment, education, and gender equality. The same research results were also found by [15] that the more advanced human development segments are in provinces with high education levels and low poverty rates. Therefore, policies that need to be implemented in this area are in the form of economic improvement, access to health services, strengthening educational programs, sufficient employment capacity, and equal opportunities and access for women.

Meanwhile, segment 2 is characterized by factors in this study that significantly affect gender equality. These regions have greater inequality problems in terms of opportunities and access for women. As in [31], a similar pattern is obtained in grouping regions in Indonesia based on the empowerment index and gender inequality. Regions with high inequality scores fall into certain clusters, which means that affirmative policies are needed to truly improve gender inclusion. In addition, [32] conducted a cluster analysis of the human development index and gender inequality and found that women in areas with low HDI consistently experience inequality in access to basic services. Efforts can be made by prioritizing policies in strengthening women's empowerment programs, eliminating gender discrimination in the workplace, and increasing women's participation in the public sector and government. Striving for this will accelerate the achievement of gender equality and encourage the formation of a more just and inclusive society.

4. CONCLUSION

Based on the results of research using the SEM-PLS method, the results of the measurement model evaluation obtained, namely there are 26 indicators that meet the modeling criteria. The model formed is a valid model as evidenced by the GoF value which is categorized as large, namely 0.808. In addition, the R^2 value for the human development variable is 85.4%, which shows that the diversity of human development variables can be explained, while the remaining 14.6% is explained by other factors outside this study. Then, in the gender equality variable, the R^2 value of 81.1% of the gender equality variable can be explained by the factor variables in this study and the remaining 18.9% is explained by other factors outside the study. Based on the results of structural evaluation, it is found that the variables that have a significant direct effect on human development (η_1) are gender equality (η_2), economy (ξ_2), and health (ξ_3), while the employment (ξ_4) and education (ξ_5) variables have an indirect significant effect on human development through the mediation of the gender equality variable (η_2). In addition, the variables of security (ξ_3), employment (ξ_4) and education (ξ_5) have a significant effect on gender equality (η_2). Segmentation using the FIMIX-PLS method resulted in the best segmentation using the BIC and CAIC value criteria, namely two segments of districts/cities in Java. The first segment has characteristics dominated by the variables in this study that have a significant effect both directly and indirectly through the mediation of gender equality on human development. Meanwhile, segment 2 has characteristics that emphasize the influence of gender equality. However, this study only uses cross-sectional secondary data without the use of spatial models. This is a limitation in this study and can be a consideration for future research in the use of spatial or longitudinal data. So that the research can provide more optimal results and can be used as the main consideration for the government in implementing its policies.

Author Contributions

Muhammad Rosyid Ridho Az Zuhro: Conceptualization, Data Curation, Formal Analysis, Methodology, Investigation, Visualization, Writing – Original Draft, Writing – Review and Editing. Ardi Kurniawan: Project Administration, Resources, Software, Supervision, Validation. Dita Amelia: Project Administration, Resources, Supervision, Validation, Idrus Syahzaqi: Resources, Supervision, Validation. All authors discussed the results and contributed to the final manuscript.

Funding Statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgment

The authors would like to thank the Statistics Study Program, Faculty of Science and Technology, Universitas Airlangga, the Central Bureau of Statistics (BPS), the Health Office, and the Education Office as data providers and all parties involved in supporting this research and publication.

Declarations

The authors declare no competing interest.

REFERENCES

[1] Badan Pusat Statistik, “INDEKS PEMBANGUNAN MANUSIA 2020,” Jakarta, 2021.

[2] M. Ul Haq, *REFLECTIONS ON HUMAN DEVELOPMENT*. Oxford university Press, 1995. doi: https://doi.org/10.1093/oso/9780195101911.001.0001

[3] United Nations Development Programme, “HUMAN DEVELOPMENT INSIGHTS,” Human Development Reports. Accessed: Feb. 18, 2025. [Online]. Available: https://hdr.undp.org/data-center/country-insights#/ranks

[4] United Nations Development Programme, “HASIL TEMUAN PROGRAM PEMBANGUNAN PBB: NEGARA-NEGARA MAJU CAPAI REKOR PEMBANGUNAN MANUSIA, NAMUN SETENGAH DARI NEGARA-NEGARA BERKEMBANG ALAMI KEMUNDURAN.” Accessed: Feb. 06, 2025. [Online]. Available: https://www.undp.org/id/indonesia/press-releases/hasil-temuan-program-pembangunan-pbb-negara-negara-maju-capai-rekor-pembangunan-manusia-namun-setengah-dari-negara-negara

[5] United Nations, “REDUCE INEQUALITY WITHIN AND AMONG COUNTRIES. DEPARTMENT OF ECONOMIC SOCIAL AFFAIRS SUSTAINABLE,” United Nations. Accessed: Feb. 06, 2025. [Online]. Available: https://sdgs.un.org/goals/goal11

[6] E. G. Tampubolon, M. Irvan, and D. Hartono, “DETERMINAN PERTUMBUHAN EKONOMI KABUPATEN/KOTA PROVINSI JAWA TIMUR TAHUN 2014-2020,” *JABE (Journal of Applied Business and Economic)*, vol. 9, no. 1, pp. 68–80, 2022. doi: https://doi.org/10.30998/jabe.v9i1.14711

[7] D. Hartono, “PENTINGNYA PEMBANGUNAN DAN PEMBERDAYAAN GENDER DALAM RANGKA UPAYA MENDUKUNG PEMBANGUNAN MANUSIA PROVINSI DKI JAKARTA,” *Jurnal Mirai Management*, vol. 8, no. 1, pp. 398–404, 2023.

[8] M. D. Wagho, “MODEL PERSAMAAN SIMULTAN UNTUK ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI PERSENTASE KEMISKINAN DAN INDEKS PEMBANGUNAN MANUSIA DI PROVINSI JAWA TIMUR,” Institut Sains & Teknologi AKPRIND Yogyakarta, Yogyakarta, 2021.

[9] T. M. Sapaat, A. L. C. P. Lapian, and S. Y. L. Tumangkeng, “ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI INDEKS PEMBANGUNAN MANUSIA DI PROVINSI SULAWESI UTARA TAHUN (2005-2019),” *Jurnal Berkala Ilmiah Efisiensi*, vol. 20, no. 03, 2020.

[10] M. A. Budiman and N. Cahyani, “PEMODELAN REGRESI LOGISTIK ORDINAL PADA INDEKS PEMBANGUNAN MANUSIA (IPM) DI JAWA TIMUR TAHUN 2020,” *Jurnal Statistika Dan Komputasi*, vol. 1, no. 2, pp. 64–73, 2022.doi: https://doi.org/10.32665/statkom.v1i2.1169

[11] J. Wang and X. Wang, *STRUCTURAL EQUATION MODELING: APPLICATIONS USING MPLUS*. John Wiley & Sons, 2019. doi: https://doi.org/10.1002/9781119422730

[12] I. Ghozali, *STRUCTURAL EQUATION MODELING: METODE ALTERNATIF DENGAN PARTIAL LEAST SQUARES (PLS)*. Semarang: Badan Penerbit Universitas Diponegoro, 2014.
[13] A. Pariesti and U. R. Christa, "PENGARUH KOMPETENSI DAN GAYA KEPEMIMPINAN TRANSFORMASIONAL TERHADAP KINERJA PEGAWAI DENGAN MOTIVASI SEBAGAI VARIABEL INTERVENING PADA KANTOR INSPEKTORAT KABUPATEN KATINGAN," *Journal of Environment and management*, vol. 3, no. 1, pp. 35–45, 2022.doi: https://doi.org/10.37304/jem.v3i1.4284
[14] I. Ghozal, *PARTIAL LEAST SQUARES KONSEP, TEKNIK, DAN APLIKASI MENGGUNAKAN PROGRAM SMARTPLS 3.2.9 UNTUK PENELITIAN EMPIRIS*. Semarang: Badan Penerbit Universitas Diponegoro, 2021.
[15] D. S. Vanisa, T. R. Rahmanita, E. Ana, and A. Kurniawan, "PROVINCIAL SEGMENTATION IN INDONESIA: EXPLORING FACTORS INFLUENCING EDUCATION WITH SEM-PLS METHOD, INCORPORATING MODERATION EFFECTS AND FIMIX-PLS APPROACH," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 18, no. 3, pp. 1955–1962, 2024.doi: https://doi.org/10.30598/barekengvol18iss3pp1955-1962
[16] R. R. Nuryanti and T. Soebagijo, "SEM-PLS UNTUK ANALISIS STRUKTUR KEMISKINAN PADA MASA PANDEMI COVID-19," in *Seminar Nasional Official Statistics*, 2021, pp. 195–203.doi: https://doi.org/10.34123/semnasoffstat.v2021i1.836
[17] E. D. Anggita, A. Hoyyi, and A. Rusgiyono, "ANALISIS STRUCTURAL EQUATION MODELLING PENDEKATAN PARTIAL LEAST SQUARE DAN PENGELOMPOKAN DENGAN FINITE MIXTURE PLS (FIMIX-PLS)(STUDI KASUS: KEMISKINAN RUMAH TANGGA DI INDONESIA 2017)," *Jurnal Gaussian*, vol. 8, no. 1, pp. 35–45, 2019.doi: https://doi.org/10.14710/j.gauss.v8i1.26620
[18] M. Abdullah, "METODE PENELITIAN KUANTITATIF," 2015, *Aswaja pressindo*.
[19] H. A. Sandri, "PEMODELAN NIAT PEMBELIAN KEMBALI PADA PELANGGAN MARKETPLACE BERDASARKAN METODE STRUCTURAL EQUATION MODELING DENGAN PENDEKATAN PARTIAL LEAST SQUARE (STUDI KASUS DI MARKETPLACE SHOPEE)," 2021.
[20] S. Haryono and P. Wardoyo, "STRUCTURAL EQUATION MODELING," *Bekasi: PT Intermedia Personalia Utama*, 2012.
[21] J. F. Hair, G. T. Hult, C. M. Ringle, and M. Sarstedt, *A PRIMER ON PARTIAL LEAST SQUARES STRUCTURAL EQUATION MODELLING (PLS-SEM)*. SAGE Publications, Inc., 2014.
[22] I. Ghozali, *MODEL PERSAMAAN STRUKTURAL KONSEP DAN APLIKASI DENGAN PROGRAM AMOS 22 UPDATE BAYESIAN SEM*. Semarang: Badan Penerbit Universitas Diponegoro, 2011.
[23] G. S. Trujillo, "PATHMOX APPROACH: SEGMENTATION TREES IN PARTIAL LEAST SQUARES PATH MODELING," Universitat Politècnica de Catalunya (UPC), 2009.
[24] S. Nurhalizah, G. Kholijah, and Z. Gusmanely, "ANALISIS STRUCTURAL EQUATION MODELING PARTIAL LEAST SQUARE PADA KINERJA PEGAWAI PT. BANK PEMBANGUNAN DAERAH JAMBI," *Indonesian Journal of Applied Statistics*, vol. 6, no. 2, pp. 125–139, 2024.doi: https://doi.org/10.13057/ijas.v6i2.78921
[25] C. M. Ringle, M. Sarstedt, and E. A. Mooi, "RESPONSE-BASED SEGMENTATION USING FINITE MIXTURE PARTIAL LEAST SQUARES: THEORETICAL FOUNDATIONS AND AN APPLICATION TO AMERICAN CUSTOMER SATISFACTION INDEX DATA," in *Data mining: Special issue in annals of information systems*, Springer, 2009, pp. 19–49. doi: https://doi.org/10.1007/978-1-4419-1280-0_2
[26] I. N. Afifah, "ANALISIS STRUCTURAL EQUATION MODEL (SEM) DENGAN FINITE MIXTURE PARTIAL LEAST SQUARES (FIMIX-PLS)," *Penguatan Peran Matematika Dan Pendidikan Matematika Untuk Indonesia Yang Lebih Baik, November*, pp. 978–979, 2013.
[27] M. Sarstedt, J.-M. Becker, C. M. Ringle, and M. Schwaiger, "UNCOVERING AND TREATING UNOBSERVED HETEROGENEITY WITH FIMIX-PLS: WHICH MODEL SELECTION CRITERION PROVIDES AN APPROPRIATE NUMBER OF SEGMENTS?," *Schmalenbach Business Review*, vol. 63, pp. 34–62, 2011.doi: https://doi.org/10.1007/BF03396886
[28] E. E. Rigdon, C. M. Ringle, and M. Sarstedt, "STRUCTURAL MODELING OF HETEROGENEOUS DATA WITH PARTIAL LEAST SQUARES," *Review of marketing research*, pp. 255–296, 2010. doi: https://doi.org/10.1108/S1548-6435(2010)0000007011
[29] K. P. Burnham and D. R. Anderson, "MULTIMODEL INFERENCE: UNDERSTANDING AIC AND BIC IN MODEL SELECTION," *Sociol Methods Res*, vol. 33, no. 2, pp. 261–304, 2004.doi: https://doi.org/10.1177/0049124104268644
[30] C. Hahn, M. D. Johnson, A. Herrmann, and F. Huber, "CAPTURING CUSTOMER HETEROGENEITY USING A FINITE MIXTURE PLS APPROACH," *Schmalenbach Business Review*, vol. 54, pp. 243–269, 2002. doi: https://doi.org/10.1007/BF03396655
[31] J. S. Muthmaina, "THE FEMININOMENON OF INEQUALITY: A DATA-DRIVEN ANALYSIS AND CLUSTER PROFILING IN INDONESIA," *arXiv preprint arXiv:2412.00012*, 2024.
[32] A. Rahmadian, V. Hannisa, N. R. Alifa, M. Zaky, and Y. F. Rahmah, "A COMPREHENSIVE ANALYSIS OF GENDER DISPARITIES IN INDONESIAN HUMAN DEVELOPMENT USING MACHINE LEARNING TECHNIQUES," in *International Conference on Islamic Economics (ICIE)*, 2024, pp. 691–696.