MIKAILALSYS
Journal of Multidisciplinary Sciences

e-ISSN : 2987-2286
p-ISSN : 2987-3924

Index: Harvard, Boston, Sydney, Copernicus, Sinta 4, Scilit, Semantic, Google, etc

https://doi.org/10.58578/mikailalsys.v4i3.11902

Modeling Human Development Index Based on Socio-Economic Factors in West Sumatra Using Local Polynomial Regression

Alya Zafirah & Fadhilah Fitri
Universitas Negeri Padang, Indonesia
zafirahalya07@gmail.com; fadhilahfitri@fmipa.unp.ac.id

Article Info:

Submitted:	Revised:	Accepted:	Published:
Jul 13, 2026	Aug 10, 2026	Aug 22, 2026	Aug 27, 2026

Abstract

Human development is a central concern in regional policy because economic advancement does not necessarily translate into improved quality of life. However, research modelling nonlinear relationships between the Human Development Index (HDI) and socioeconomic factors across the regencies and municipalities of West Sumatra using Local Polynomial Regression (LPR) remains limited. This study aims to model HDI based on the senior high school net enrolment rate (NER), gross regional domestic product (GRDP), and population density. A quantitative, nonexperimental, cross-sectional design was employed using 2023 secondary data from Statistics Indonesia covering 19 regencies and municipalities. The data were analysed using nonparametric LPR with a local-linear estimator and optimal bandwidth selection. The findings indicate that all three predictors are positively associated with HDI, with senior high school NER exhibiting the strongest correlation. The model achieved an R^2 of 0.9147, a root mean squared error of 1.3458, and a mean absolute error of 1.0292, demonstrating its ability to capture nonlinear patterns and explain regional variation in HDI. This study contributes to regional development modelling by demonstrating the applicability of LPR to multidimensional human development data. The findings provide an empirical basis for more context-sensitive development policies that integrate educational participation, economic capacity, and population distribution.

Keywords: Human Development Index; Local Polynomial Regression; Net Enrolment Rate; Gross Regional Domestic Product; Population Density

INTRODUCTION

Human development has increasingly become a central concern in development policy because economic progress alone does not necessarily translate into wider opportunities for people to live healthy, educated, and decent lives. At the international level, recent human development trends reveal an uneven recovery from global disruptions, with persistent disparities in capabilities and living standards across countries and population groups. The *Human Development Report 2023/2024* emphasizes that progress in human development remains unequal and that widening development gaps can weaken the capacity of societies to respond collectively to economic and social challenges (Development, 2024). A comparable concern is evident in Indonesia. Although the national Human Development Index (HDI) reached 74.39 in 2023, representing an increase of 0.84% from the previous year, improvements have not occurred at the same pace across regions (Badan Pusat Statistik, 2024). Regional disparities therefore remain relevant to the Indonesian development agenda, as differences within and between regions continue to shape the distribution of human development achievements (Ginanjar et al., 2024).

Human development is more meaningful when it is understood not merely as an improvement in aggregate economic indicators but as an expansion of people's real opportunities to pursue lives they value. Sen (1999) places human capabilities at the center of development, arguing that income is important primarily because of what it enables people to achieve. This perspective is reflected in the HDI, which brings together dimensions of health, education, and a decent standard of living rather than relying solely on economic output (Programme, 2024). Such an understanding suggests that regional development should be examined through the interaction of social and economic conditions that shape people's access to education, productive opportunities, public services, and material well-being. Empirical evidence from Indonesia similarly indicates that human development is associated with a broad set of educational, health, economic, and social characteristics rather than with a single development indicator (Masduki et al., 2022; Tyas & Sukartini, 2022).

West Sumatra provides an important regional setting for examining these relationships. The province recorded an HDI of 75.64 in 2023, increasing from 75.16 in the preceding year, yet development outcomes still varied among its regencies and municipalities (Badan Pusat Statistik Provinsi Sumatera Barat, 2024). Such variation deserves attention because a provincial average can conceal differences in educational participation, economic capacity, population concentration, infrastructure, and access to services at the local level. Education is particularly relevant because wider participation in schooling can strengthen knowledge, skills, and future economic opportunities. Recent Indonesian studies have also identified education-related factors as closely connected with human development outcomes (Baskara & Dahlan, 2024; Tyas & Sukartini, 2022). Economic capacity provides another important dimension. Evidence from a dynamic panel analysis indicates that Gross Regional Domestic Product (GRDP) is associated with improvements in HDI through pathways that include household welfare and access to education and health services (Astuti et al., 2025). Population characteristics also deserve consideration because population density may improve the efficiency of service provision in some areas while creating pressure on infrastructure and public services in others (Anggraini, 2024).

Socio-economic indicators may consequently relate to HDI through relationships that are neither uniform nor proportional across regions. An increase in economic output, for example, does not automatically generate an equivalent increase in human well-being when the benefits of growth are unevenly distributed or when access to essential services remains limited. Educational participation may also produce different development outcomes according to the social and economic conditions surrounding students and households. Population concentration can create economies of scale for schools, health facilities, transportation, and other public services, yet very high or very low density may generate different accessibility challenges. Studies conducted in Indonesia have demonstrated the relevance of population density, schooling, poverty, and other socio-economic indicators to variations in HDI (Aliu et al., 2023; Alwi et al., 2021; Anggraini, 2024). These findings reinforce the need to examine the shape of the relationship itself rather than assuming in advance that every predictor affects HDI through a constant linear pattern.

Earlier empirical studies have approached HDI using a range of statistical frameworks. Tyas & Sukartini (2022) employed ordinary least squares and fixed-effects models to examine determinants of HDI and its dimensions across Indonesian provinces, while Aliu et al. (2023) applied ordinal logistic regression to classify HDI across Indonesian regencies and municipalities and identified variables such as average years of schooling and population density as significant predictors. Alwi et al. (2021) took a different direction by applying nonparametric spline regression to Indonesian HDI data and demonstrated that relationships between HDI and several explanatory variables did not necessarily follow a clearly specified parametric form. Other studies have relied on panel-data approaches to explore links between human development, regional economic performance, public expenditure, and socio-economic conditions (Sijabat, 2024). These contributions provide valuable evidence, yet they also show that the analytical conclusions obtained from HDI data are closely connected to the functional assumptions and estimation procedures used in the model.

Regression analysis offers a useful framework for describing relationships between a response variable and one or more predictors, but the choice between parametric and nonparametric approaches becomes important when the underlying relationship is not known in advance. Parametric regression requires a predetermined functional structure, which can be efficient when the assumed form adequately reflects the observed data. Real socio-economic relationships, however, may change across different ranges of a predictor and may exhibit curvature, thresholds, or locally varying tendencies that are difficult to represent through a single global function. Nonparametric regression provides greater flexibility because the regression curve can be estimated with fewer restrictions on its functional form (Eubank, 1988; Härdle & Linton, 1994). This flexibility is particularly relevant to regional human development, where improvements in education or economic capacity may generate different marginal patterns depending on the socioeconomic position of individual regencies or municipalities.

Local polynomial regression (LPR) offers one nonparametric strategy for addressing such complexity. Rather than imposing one polynomial equation on the entire range of observations, LPR estimates the regression function within neighborhoods surrounding particular evaluation points, allowing the fitted relationship to respond to local features of the data (Fan & Gijbels, 1996). Kernel weighting gives observations closer to a local point greater influence in the estimation process, making the resulting curve sensitive to local structure while retaining smoothness (Härdle & Linton, 1994). The bandwidth is especially important because it determines the size of the neighborhood used in estimation. A bandwidth that is too small may produce an excessively variable fit, whereas an overly large bandwidth may smooth away meaningful features of the relationship (Fan & Gijbels, 1996). Empirical applications of local polynomial methods in Indonesia have demonstrated the importance of selecting an appropriate bandwidth and model specification to balance accuracy and smoothness (Hendrian et al., 2021)(Suparti & Prahutama, 2016).

The methodological gap emerges from the intersection between the substantive problem of regional human development and the statistical characteristics of its relationships with socio-economic indicators. Indonesian HDI research has employed linear panel models, ordinal logistic regression, and nonparametric spline techniques (Aliu et al., 2023; Alwi et al., 2021; Tyas & Sukartini, 2022) , while local polynomial regression has been more commonly demonstrated in other empirical settings, such as electricity-load and commodity-price modeling (Hendrian et al., 2021; Suparti & Prahutama, 2016). Evidence specifically applying a multivariate local polynomial framework to model HDI across the regencies and municipalities of West Sumatra remains limited. This distinction is important because the contribution of LPR does not lie in introducing a new statistical estimator, but in using an established nonparametric framework to investigate regional human development without forcing the relationships between HDI and its socio-economic predictors into a single prespecified global form. The approach is theoretically consistent with the data-adaptive properties of local polynomial estimation described by Fan & Gijbels (1996) and with the broader rationale for nonparametric modeling articulated by (Eubank, 1988) and (Härdle & Linton, 1994).

The present study focuses on the 19 regencies and municipalities of West Sumatra using 2023 socio-economic data and examines HDI in relation to the senior high school Net Enrollment Rate (NER), GRDP per capita, and population density. These indicators represent educational participation, regional economic capacity, and demographic concentration, three dimensions that can shape people's opportunities to obtain knowledge, material resources, and access to public services. By employing multivariate local polynomial regression, the study seeks to preserve possible nonlinear and locally varying patterns that may be overlooked under a strictly parametric specification. Accordingly, this study aims to model the Human Development Index across regencies and municipalities in West Sumatra based on selected socio-economic factors using Local Polynomial Regression, determine an appropriate smoothing specification through bandwidth selection, and evaluate how well the resulting model represents the observed variation in regional human development.

METHODS

This study employed a quantitative approach with a nonexperimental, cross-sectional design using secondary regional data. The study covered all 19 regencies and municipalities in West Sumatra Province. Since all administrative units were included in the analysis, no sampling procedure was applied. The data represented socio-economic conditions in 2023 and were obtained from official publications of the Central Statistics Agency (*Badan Pusat Statistik*, BPS) of West Sumatra Province.

The study initially involved five variables consisting of one response variable and four predictor variables. The Human Development Index (HDI) was specified as the response variable (Y), while the predictor variables consisted of Labor Force Participation Rate (LFPR), Senior High School Net Enrollment Rate (NER), population density, and Gross Regional Domestic Product (GRDP). All variables were measured on a ratio scale, as presented in Table 1.

Table 1. Research Variables

Variable	Description	Scale
Y	Human Development Index (HDI)	Ratio
X_1	Labor Force Participation Rate (LFPR)	Ratio
X_2	Senior High School Net Enrollment Rate (NER)	Ratio
X_3	Population Density	Ratio
X_4	Gross Regional Domestic Product (GRDP)	Ratio

Data were collected through documentation and systematic extraction of the relevant 2023 indicators for each regency and municipality. Because the study relied entirely on official secondary statistical data rather than questionnaires, interviews, or researcher-developed instruments, conventional instrument validity and reliability testing was not required. The data were organized and analyzed using RStudio.

The analytical procedure consisted of several stages: (1) collecting HDI data and the socio-economic indicators considered relevant to human development; (2) examining the characteristics of the data using descriptive analysis; (3) constructing scatter plots and examining the relationships between the response and predictor variables; (4) determining the optimal bandwidth; (5) estimating the regression model using Local Polynomial

Regression; (6) visualizing the estimated results; (7) evaluating and interpreting the fitted model; and (8) drawing conclusions based on the analytical findings.

a. Local Polynomial Regression

Local Polynomial Regression (LPR) is a nonparametric regression method in which the regression function is estimated locally using a polynomial function. This method does not require the relationship between the response and predictor variables to follow a predetermined global functional form. Instead, the regression curve is allowed to adjust to the local characteristics of the observed data (Eubank, 1988; Fan & Gijbels, 1996).

In general, a nonparametric regression model can be expressed as follows:

$$\mathbf{Y_i = f(X_i) + \varepsilon_i, \quad i = 1, 2, \ldots, n} \qquad (1)$$

where Y_i represents the response variable, X_i represents the predictor variable, $f(X_i)$ is an unknown regression function, and ε_i represents the random error.

Local Polynomial Regression is based on a Taylor series expansion around a local evaluation point x_0. If $f(x)$ is approximated by a polynomial of degree m around x_0, the Taylor expansion can be expressed as follows (Fan & Gijbels, 1996):

$$\mathbf{f(x) \approx f(x_0) + f'(x_0)(x - x_0) + [f''(x_0)/2!](x - x_0)^2 + \ldots + [f^m(x_0)/m!](x - x_0)^m} \qquad (2)$$

By defining:

$$\mathbf{\beta_j = f^j(x_0)/j!, \quad j = 0, 1, 2, \ldots, m}$$

the local polynomial approximation can be written more simply as:

$$\mathbf{f(x) \approx \beta_0 + \beta_1(x - x_0) + \beta_2(x - x_0)^2 + \ldots + \beta_m(x - x_0)^m} \qquad (3)$$

Therefore, the Local Polynomial Regression model can be expressed as:

$$\mathbf{Y_i = \beta_0 + \beta_1(X_i - x_0) + \beta_2(X_i - x_0)^2 + \ldots + \beta_m(X_i - x_0)^m + \varepsilon_i} \qquad (4)$$

In this study, a local-linear specification was employed. Therefore, the polynomial degree was set to m = 1. For a model involving p predictor variables, the local-linear model is expressed as:

$$\mathbf{Y_i = \beta_0 + \Sigma_{j=1}^{p} \beta_j(X_{ij} - x_{0j}) + \varepsilon_i} \qquad (5)$$

where X_{ij} represents the value of predictor j for observation i, x_{0j} represents the local evaluation point for predictor j, β_0 represents the local intercept, and β_j represents the local regression parameter.

The model can generally be represented in matrix form as:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{6}$$

where Y represents the vector of response observations, X represents the local design matrix, β represents the vector of regression parameters, and ε represents the vector of random errors.

b. Estimation of the Local Polynomial Regression Model

Parameter estimation in Local Polynomial Regression was performed using Weighted Least Squares (WLS). In this approach, observations located closer to the local evaluation point receive greater weights than observations located farther from that point. The weights are determined using a kernel function (Eubank, 1988).

The kernel function with bandwidth h can be expressed as:

$$K_h(X_i - x_0) = (1/h)\, K[(X_i - x_0)/h] \tag{7}$$

where K represents the kernel function and h represents the bandwidth.

Several kernel functions can be used in Local Polynomial Regression, including Uniform, Triangular, Epanechnikov, and Gaussian kernels (Härdle, 1991).

The Uniform kernel is defined as:

$$K(u) = 1/2, \quad \text{for } |u| \le 1; \quad K(u) = 0, \quad \text{for } |u| > 1 \tag{8}$$

The Triangular kernel is defined as:

$$K(u) = 1 - |u|, \quad \text{for } |u| \le 1; \quad K(u) = 0, \quad \text{for } |u| > 1 \tag{9}$$

The Epanechnikov kernel is defined as:

$$K(u) = 3/4(1 - u^2), \quad \text{for } |u| \le 1; \quad K(u) = 0, \quad \text{for } |u| > 1 \tag{10}$$

The Gaussian kernel is defined as:

$$K(u) = [1/\sqrt{(2\pi)}]\, \exp(-u^2/2), \quad -\infty < u < \infty \tag{11}$$

A kernel function is generally real-valued, bounded, continuous, and symmetric, and satisfies the following condition (Härdle & Linton, 1994):

$$\int_{-\infty}^{\infty} K(u)\, du = 1 \tag{12}$$

For a multivariate Local Polynomial Regression model, the kernel weight for observation i can be expressed as:

$$w_i(x_0) = \Pi_{j=1}^{p} \{(1/h_j)\ K[(X_{ij} - x_{0j})/h_j]\} \qquad (13)$$

where h_j represents the bandwidth corresponding to predictor j.

The regression parameters are estimated by minimizing the weighted sum of squared errors:

$$\Sigma_{i=1}^{n}\ w_i(x_0)[Y_i - \beta_0 - \Sigma_{j=1}^{p}\ \beta_j(X_{ij} - x_{0j})]^2 \qquad (14)$$

In matrix form, the weighted least squares objective function can be expressed as:

$$(Y - X\beta)^T\ W\ (Y - X\beta) \qquad (15)$$

where W represents the diagonal matrix containing the kernel weights.

The estimator of the Local Polynomial Regression parameter is:

$$\hat{\beta} = (X^TWX)^{-1}X^TWY \qquad (16)$$

The estimated regression function at the local evaluation point x_0 is obtained from the estimated intercept:

$$\hat{f}(x_0) = \hat{\beta}_0 \qquad (17)$$

This estimation procedure is repeated at different evaluation points to obtain the fitted Local Polynomial Regression function.

c. Bandwidth Selection

Bandwidth is an important component of Local Polynomial Regression because it determines the size of the neighborhood used in estimating the regression function. A bandwidth that is too large may cause *oversmoothing*, resulting in an estimated curve that is excessively smooth and unable to capture important local patterns. This condition generally produces relatively high bias and low variance. Conversely, a bandwidth that is too small may cause *undersmoothing*, producing a highly fluctuating estimated curve with relatively low bias but high variance (Fan & Gijbels, 1996; Suparti & Prahutama, 2016).

Therefore, an appropriate bandwidth must be selected to achieve a balance between bias and variance. In this study, the Local Polynomial Regression model used a local-linear specification (regtype = "ll"). The optimal bandwidth for each predictor was selected through cross-validation using an AIC-based criterion (bwmethod = "cv.aic"). The selected bandwidths were subsequently used in estimating the final multivariate Local Polynomial Regression model.

d. Model Evaluation

The performance of the Local Polynomial Regression model was evaluated using Mean Square Error (MSE), the coefficient of determination (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE).

Mean Square Error measures the average squared difference between the observed and predicted response values and is expressed as:

$$\mathbf{MSE = (1/n)\, \Sigma_{i=1}^{n}\, (Y_i - \hat{Y}_i)^2} \qquad (18)$$

where Y_i represents the observed HDI value and $\hat{Y}_i$ represents the predicted HDI value.

The coefficient of determination represents the proportion of variation in the response variable that can be represented by the fitted model. It is calculated as:

$$\mathbf{R^2 = 1 - [\Sigma_{i=1}^{n}\, (Y_i - \hat{Y}_i)^2 \,/\, \Sigma_{i=1}^{n}\, (Y_i - \bar{Y})^2]} \qquad (19)$$

where $\bar{Y}$ represents the mean of the observed HDI values. An R^2 value closer to 1 indicates that a greater proportion of the variation in HDI is represented by the fitted model. Root Mean Square Error is calculated as:

$$\mathbf{RMSE = \sqrt{[(1/n)\, \Sigma_{i=1}^{n}\, (Y_i - \hat{Y}_i)^2]}} \qquad (20)$$

Mean Absolute Error is calculated as:

$$\mathbf{MAE = (1/n)\, \Sigma_{i=1}^{n}\, |Y_i - \hat{Y}_i|} \qquad (21)$$

Lower MSE, RMSE, and MAE values indicate smaller discrepancies between the observed and predicted HDI values. In addition to these quantitative measures, the fitted model was evaluated graphically by comparing the actual HDI values with the HDI values predicted by the Local Polynomial Regression model.

RESULTS

1. Intervariable Correlation

The analysis was conducted using secondary data obtained from the Central Statistics Agency (Badan Pusat Statistik/BPS) of West Sumatra Province for 19 regencies and municipalities in 2023. The Human Development Index (HDI) was treated as the response variable, while the initial set of predictors consisted of the Labor Force Participation Rate (LFPR), Senior High School Net Enrollment Rate (NER), Gross Regional Domestic Product (GRDP), and population density. The dataset was processed in R using the dplyr and readxl packages. Correlation analysis was first performed to examine the direction and strength of the association between HDI and each potential predictor before developing the multivariate nonparametric model.

Table 2. Correlation between HDI and the Predictor Variables

Variable	HDI	LFPR	Senior High School NER	GRDP	Population Density
HDI	1.0000	-0.4624	0.7339	0.7186	0.7271

Table 2 shows that Senior High School NER, GRDP, and population density were positively associated with HDI. The strongest correlation was observed between HDI and Senior High School NER, with a coefficient of 0.7339. Population density showed a similarly strong positive correlation of 0.7271, while GRDP had a correlation coefficient of 0.7186. These results indicate that regencies and municipalities with higher levels of upper-secondary school participation, stronger regional economic performance, and greater population concentration generally tended to record higher HDI values in the 2023 dataset.

LFPR displayed a different pattern, with a correlation coefficient of -0.4624. The negative coefficient indicates that its observed relationship with HDI moved in the opposite direction from those of the other three predictors. Based on the modeling specification adopted in this study, LFPR was therefore not carried forward into the subsequent multivariate Local Polynomial Regression model. The final modeling stage consequently focused on Senior High School NER, GRDP, and population density as the three predictors used to explain variation in HDI across West Sumatra.

2. Patterns of the Univariate Relationships

A graphical examination was conducted before constructing the multivariate model to provide a clearer view of how HDI varied across the range of each selected predictor. The relationships were visualized using univariate nonparametric regression with a Local Constant Kernel Regression approach. This stage was intended to reveal patterns that might not be adequately represented by a single straight-line relationship.

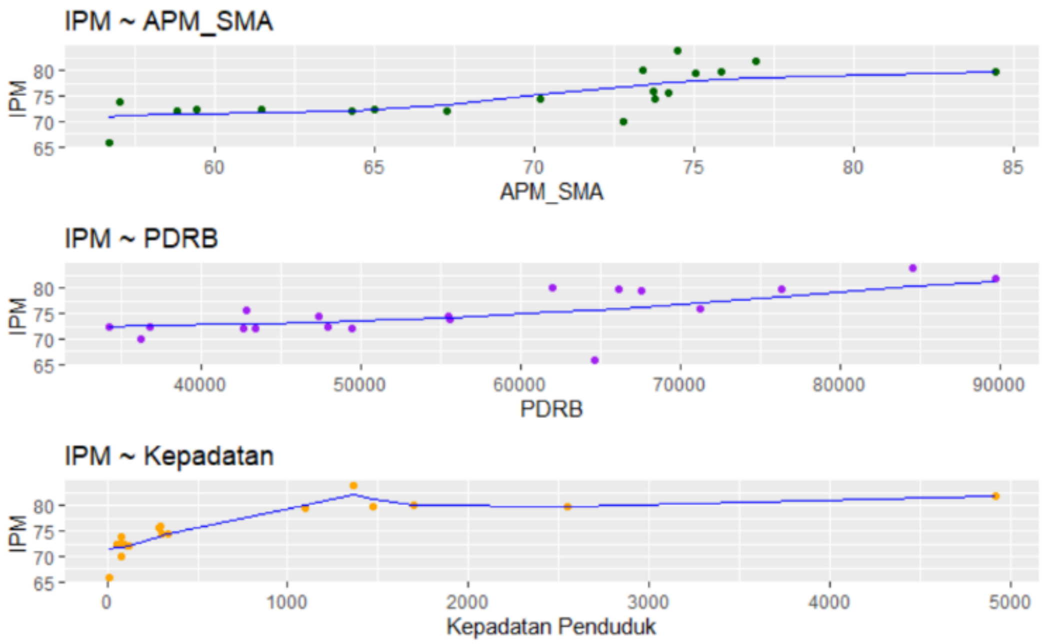

Figure 1. Relationship of HDI with Senior High School NER, GRDP, and Population Density

Figure 1 presents three fitted relationships between HDI and the selected socio-economic indicators. The first panel shows a generally positive pattern between Senior High School NER and HDI. Higher levels of school participation at the senior high school level were accompanied by higher HDI values across the observed regencies and municipalities. The fitted curve suggests that educational participation was closely associated with differences in human development outcomes in West Sumatra. The pattern is particularly meaningful because education constitutes one of the dimensions represented directly in human development measurement.

The relationship between GRDP and HDI also tended to be positive, although the fitted curve was not completely linear. At lower to middle levels of GRDP, increases in regional economic output were accompanied by a more visible rise in HDI. The curve became flatter at higher GRDP levels, indicating that comparable increases in economic capacity were not always accompanied by proportional changes in HDI. The observed pattern suggests diminishing variation in HDI at the upper range of GRDP rather than a constant linear increase across the entire predictor range.

Population density produced a more clearly nonlinear pattern. The fitted curve indicates that increases in population density were initially associated with higher HDI values, particularly across areas where a more concentrated population may coincide with greater accessibility to public facilities and services. The relationship became less pronounced after a certain range of density. This pattern shows that the association between population concentration and human development was not uniform across all observed areas. Taken together, the three panels in Figure 1 provide empirical support for the use of a flexible regression approach capable of accommodating curvature and locally changing relationships.

3. Multivariate Local Polynomial Regression Model

The three selected predictors were subsequently incorporated into a multivariate Local Polynomial Regression (LPR) model to estimate their simultaneous relationship with HDI. The model used a local-linear specification (regtype = "ll"). The local-linear form was applied to allow the fitted values to adjust to local characteristics of the observations while maintaining a smoother estimate than that produced by the initial local-constant visualization.

Bandwidth selection was performed through cross-validation using the minimum AIC criterion (bwmethod = "cv.aic"). Bandwidth determines how broadly neighboring observations contribute to the estimate around a particular point. Smaller bandwidth values provide a more locally responsive fit, whereas larger values produce greater smoothing. The estimation procedure generated a separate optimal bandwidth for each predictor, as presented in Table 3.

Table 3. Estimated Optimal Bandwidths for the LPR Model

Predictor Variable	Optimal Bandwidth
Senior High School NER	5.631048
GRDP	39,534,636,144
Population Density	1,562.601

Table 3 shows that the estimated optimal bandwidth differed considerably among the three predictor variables, reflecting differences in their measurement scales and observed distributions. Senior High School NER had an optimal bandwidth of 5.631048, while the optimal bandwidth for GRDP was 39,534,636,144. Population density produced an optimal bandwidth of 1,562.601. These bandwidths were then used in the final multivariate local-linear model to generate predicted HDI values for each regency and municipality based on the combined information contained in the three predictors.

4. Model Performance and Predictive Fit

The performance of the final LPR model was evaluated using the coefficient of determination ((R^2)), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). These measures were used to examine both the proportion of variation in HDI represented by the model and the magnitude of the differences between observed and predicted values.

Table 4. Performance Evaluation of the Local Polynomial Regression Model

Predictor Variable	Optimal Bandwidth
Senior High School NER	5.631048
GRDP	39,534,636,144
Population Density	1,562.601

As reported in Table 4, the LPR model produced an (R^2) value of 0.9147. This value indicates that approximately 91.47% of the observed variation in HDI across the 19 regencies and municipalities was represented by the model using Senior High School NER, GRDP, and population density as predictors. Only a relatively small proportion of the observed variation remained outside the fitted model.

The prediction errors were also relatively low, with an RMSE of 1.3458 and an MAE of 1.0292. The RMSE indicates that the typical magnitude of prediction error, with greater weight assigned to larger deviations, was approximately 1.35 HDI points. The MAE shows that the absolute difference between the observed and predicted HDI values averaged approximately 1.03 points. Considered together with the high (R^2), these error measures indicate that the fitted values were generally close to the observed HDI values in the dataset.

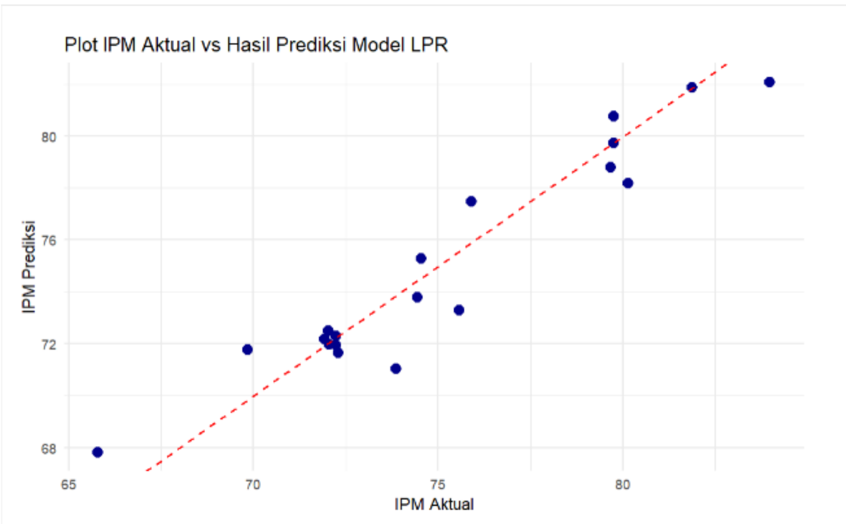

Figure 2. Actual HDI versus Predicted HDI from the Local Polynomial Regression Model

Figure 2 provides a visual comparison between the actual HDI values and those predicted by the LPR model. Most observations are located relatively close to the identity line, represented by the dashed diagonal line, indicating substantial agreement between the observed and predicted values. This graphical evidence is consistent with the quantitative evaluation reported in Table 4 and shows that the model was able to reproduce the general distribution of HDI across the observed regencies and municipalities with relatively limited prediction error.

Several observations deviate more visibly from the identity line, particularly within the middle range of HDI values. These deviations indicate that the model did not reproduce every regional observation with the same degree of precision. Such differences may reflect distinctive combinations of socio-economic characteristics among particular regencies or municipalities that are not fully represented by the three predictors included in the model. The overall concentration of points around the identity line nevertheless shows that the multivariate LPR specification captured the main empirical pattern in the 2023 West Sumatra HDI data.

Overall, the results demonstrate that the relationships between HDI and the selected socio-economic factors were not adequately characterized by a single uniform pattern. Senior High School NER, GRDP, and population density were positively associated with HDI, while graphical inspection revealed curvature and local variation, particularly in the relationships involving GRDP and population density. The multivariate local-linear LPR model accommodated these features and produced a high coefficient of determination together with relatively low prediction errors. These findings provide the empirical basis for examining how educational participation, regional economic conditions, and population concentration relate to differences in human development across West Sumatra.

DISCUSSION

The findings show that variations in the Human Development Index (HDI) across the 19 regencies and municipalities of West Sumatra are closely associated with differences in Senior High School Net Enrollment Rate (NER), Gross Regional Domestic Product (GRDP), and population density. The Local Polynomial Regression (LPR) model produced an (R^2) of 0.9147, with an RMSE of 1.3458 and an MAE of 1.0292, indicating that the model represented a substantial proportion of the observed variation in HDI while maintaining relatively small prediction errors. More importantly, the graphical results suggest that these relationships do not follow an entirely uniform linear pattern. GRDP and population density, in particular, display changes in their relationship with HDI across different ranges of the predictors. This pattern supports the use of a flexible nonparametric approach, as regional development is rarely shaped by social and economic factors in exactly the same way across areas with different characteristics. Because the analysis does not report variable-specific inferential tests or p-values, the results are more appropriately interpreted as strong empirical and model-based associations rather than as evidence of causal or statistically significant effects for each individual predictor.

Senior High School NER showed the strongest positive correlation with HDI (($r=0.7339$)), suggesting that districts and municipalities with broader participation in upper-secondary education tended to have higher levels of human development. This relationship is substantively meaningful because participation in education extends beyond school attendance itself. Greater access to secondary education can strengthen knowledge, skills, employability, and the capacity of young people to participate more fully in economic and social life. A similar pattern was reported by Suryani & Rinaldy (2023) in Padang City, where educational attainment, represented by average years of schooling, had a positive and significant relationship with HDI and emerged as the most dominant variable in their model. Recent evidence reported by Rahmarisa et al. (2026) likewise indicates that educational indicators are closely associated with improvements in population quality as measured by HDI. The present findings extend this discussion by showing that educational participation remains strongly connected with human development even when examined simultaneously with economic and demographic conditions. From a policy perspective, improving secondary-school participation should therefore be understood not simply as an educational target, but as part of a broader effort to expand people's opportunities and strengthen the quality of life within each region.

GRDP also demonstrated a strong positive association with HDI (($r=0.7186$)), although the estimated relationship was not fully linear. HDI tended to increase as regional economic capacity improved, but the fitted curve became less steep at higher levels of GRDP. This finding suggests that additional economic output may be associated with increasingly smaller differences in human development after a certain level has been reached. Comparable evidence has been reported in Indonesian studies. Suryani & Rinaldy (2023) found that GRDP had a positive and significant relationship with HDI in Padang City, while

Astuti et al. (2025), using dynamic panel data for Indonesian provinces, reported that GRDP contributed positively to HDI, particularly through economic capacity that supports household consumption, education, and health. The nonlinear pattern observed in West Sumatra adds an important qualification to these findings. Economic growth remains relevant, but higher regional output alone may not be sufficient to sustain proportional improvements in human development. The social value of economic growth depends on how effectively economic resources are translated into accessible education, health services, employment opportunities, and decent living conditions for the population.

Population density presented another notable result. Its correlation with HDI was positive and relatively strong (($r=0.7271$)), yet the graphical relationship was nonlinear. Moderate increases in density appeared to correspond with higher HDI, whereas the relationship became less pronounced beyond a certain range. This result can be understood through the spatial organization of public services. More concentrated populations may allow schools, health facilities, transportation systems, and other public infrastructure to serve residents more efficiently. Excessive concentration, by contrast, may place greater pressure on infrastructure, housing, mobility, environmental quality, and public-service capacity. Evidence from other Indonesian regions lends support to this interpretation. Anggraini (2024) identified population density as one of the variables significantly related to HDI in the Special Region of Yogyakarta, while Agesty et al. (2026) reported a strong positive relationship between population density and HDI in West Java, particularly in highly urbanized areas with comparatively better access to services and infrastructure. The West Sumatra findings therefore suggest that population density should not be viewed simply as beneficial or detrimental. Its relationship with human development depends on whether regional infrastructure and public services can respond to the needs created by the spatial concentration of the population.

The performance of the LPR model also carries an important methodological implication. Earlier research by Alwi et al. (2021) found that relationships between HDI and several socio-economic variables in Indonesia did not exhibit a clear predetermined pattern and consequently applied nonparametric spline regression. Their best spline specification achieved an (R^2) of 84.79%, providing evidence that flexible nonparametric techniques can be useful for human-development data. The present study follows a related methodological rationale but employs a local-linear polynomial approach, allowing estimation to adjust to the characteristics of observations within local neighborhoods. Local polynomial methods have also demonstrated useful predictive performance in other contexts. Hendrian et al. (2021), for example, applied local polynomial regression to world gold-price data and showed that appropriate selection of the kernel, polynomial degree, and bandwidth could produce low forecasting error. Mahdy (2023) similarly employed a local polynomial approach to model the relationship between health complaints and the Happiness Index in Indonesia and identified a relationship that varied across the observed data range. These studies reinforce the value of nonparametric estimation when the empirical relationship is too irregular to be represented adequately through a single global linear function. The (R^2) of 0.9147 obtained in the present study should nevertheless not be interpreted as directly superior to values reported in earlier studies, since differences in datasets, observation units, predictors, periods, and model specifications prevent a direct performance comparison.

The findings have both theoretical and practical implications for regional human-development planning. Theoretically, the results reinforce the view that human development emerges from the interaction of educational, economic, and demographic conditions rather than from a single isolated determinant. The nonlinear patterns further suggest that the association between socioeconomic resources and human development can change according to the level and context of those resources. Practically, this means that development policy across West Sumatra should be sensitive to differences among regencies and municipalities. Areas with lower secondary-school participation may require stronger efforts to reduce barriers to continued schooling, while regions experiencing economic expansion need mechanisms that translate growth into wider access to essential services and household welfare. Densely populated areas may require greater investment in service capacity and infrastructure, whereas less densely populated areas may face a different challenge: ensuring that education, health care, and other essential services remain geographically accessible. A locally responsive policy orientation is therefore more consistent with the patterns identified by the LPR model than a uniform intervention applied in the same manner across all regions.

Several limitations should be considered when interpreting these findings. The study is based on a cross-sectional dataset covering only 19 regencies and municipalities in West Sumatra for 2023, which limits the extent to which temporal changes and longer-term development dynamics can be examined. The relatively small number of observations also requires caution when interpreting the high (R^2), particularly because model evaluation is based on the same regional dataset used for estimation rather than on an independent out-of-sample dataset. Predictor selection also excluded LFPR after its negative bivariate correlation with HDI was observed; a negative relationship, however, does not necessarily imply that a variable lacks explanatory relevance when considered jointly with other predictors. Other dimensions that may shape regional human development, including health conditions, poverty, income inequality, government expenditure, infrastructure, and the quality of public services, were not incorporated into the final model. Future research could therefore extend the analysis through panel data covering multiple years, a broader set of socioeconomic and health indicators, spatial dependence among neighboring regions, alternative bandwidth-selection procedures, and out-of-sample or cross-validation strategies. Such developments would provide a more comprehensive understanding of how local socioeconomic conditions are connected to human development and would strengthen the evidence available for region-specific policy formulation.

CONCLUSION

Based on the results and discussion, the multivariate nonparametric model using Local Polynomial Regression (LPR) was effective in representing the relationship between the Human Development Index (HDI) and selected socio-economic indicators across the regencies and municipalities of West Sumatra. Senior High School Net Enrollment Rate (NER), Gross Regional Domestic Product (GRDP), and population density all showed positive relationships with HDI, with Senior High School NER exhibiting the strongest association. The model achieved an (R^2) value of 0.9147, indicating that approximately 91.47% of the variation in HDI was represented by the three predictors included in the model. The relatively low prediction errors, with an RMSE of 1.3458 and an MAE of 1.0292, further indicate that the estimated HDI values were generally close to the observed values.

The findings also demonstrate the value of LPR for capturing nonlinear and locally varying relationships that may not be adequately represented through a conventional linear specification. The observed patterns suggest that improvements in educational participation and regional economic capacity, together with appropriate management of population concentration and public-service accessibility, are closely related to better human development outcomes. These findings provide an empirical basis for more context-sensitive regional development policies in West Sumatra, particularly policies aimed at expanding educational participation, strengthening inclusive economic development, and improving access to public services. Future studies are encouraged to extend the analysis by incorporating longer time periods, additional socio-economic and health indicators, spatial characteristics, and out-of-sample validation to provide a broader and more robust understanding of regional human development dynamics.

REFERENCES

Agesty, W., Felisa, N. B. S., Barokah, A. S., & Winata, H. M. (2026). Regional development in West Java Province: Clustering population density, human development index, and life expectancy. *Tamalanrea: Journal of Government and Development (JGD), 3*(1), 21–34. https://doi.org/10.69816/jgd.v3i1.48775

Aliu, M. A., Fitrianto, A., Erfiani, Indahwati, & Khusnia, N. K. (2023). Pemodelan regresi logistik ordinal *backward* dengan imputasi K-nearest neighbour pada Indeks Pembangunan Manusia di Indonesia tahun 2021. *Jurnal Statistika dan Aplikasinya, 7*(1), 49–61. https://doi.org/10.21009/JSA.07105

Alwi, W., Irwan, M., & Musfirah. (2021). Penerapan regresi nonparametrik spline dalam memodelkan faktor-faktor yang mempengaruhi Indeks Pembangunan Manusia (IPM) di Indonesia tahun 2018. *Jurnal MSA (Matematika dan Statistika serta Aplikasinya), 9*(2), 112–119. https://doi.org/10.24252/msa.v9i2.23055

Anggraini, J. M. (2024). The effect of total unemployment, regional minimum wage, population density, and original local government revenue on the Human Development Index in the Special Region of Yogyakarta, 2014–2021. *Efficient: Indonesian Journal of Development Economics, 7*(1), 32–43. https://doi.org/10.15294/efficient.v7i1.4831

Astuti, H., Susilo, J. H., Trifandha, S., & Nkembo, A. F. (2025). Dynamic panel data analysis of the Human Development Index in Indonesia. *Ekuilibrium: Jurnal Ilmiah Bidang Ilmu Ekonomi, 20*(2), 229–245. https://doi.org/10.24269/ekuilibrium.v20i2.2025.pp229-245

Baskara, A., & Dahlan, D. (2024). Pengaruh anggaran pendidikan terhadap *Human Development Index* (HDI) di Indonesia periode tahun 2004 sampai dengan tahun 2023. *PEKA, 12*(1), 39–53. https://doi.org/10.25299/peka.2024.vol12(1).17564

Eubank, R. L. (1988). *Spline smoothing and nonparametric regression*. Marcel Dekker.

Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications*. Chapman & Hall.

Ginanjar, R. A. F., Sutjipto, H., Suhendra, I., & Anwar, C. J. (2024). A Theil decomposition of regional grouping in Indonesia's Human Development Index. *Economics Development Analysis Journal, 13*(3), 338–354. https://doi.org/10.15294/edaj.v13i3.13802

Härdle, W., & Linton, O. (1994). Applied nonparametric methods. In R. F. Engle & D. L. McFadden (Eds.), *Handbook of econometrics* (Vol. 4, pp. 2295–2339). Elsevier. https://doi.org/10.1016/S1573-4412(05)80007-8

Hendrian, J., Suparti, & Prahutama, A. (2021). Pemodelan harga emas dunia menggunakan metode nonparametrik polinomial lokal dilengkapi GUI R. *Jurnal Gaussian, 10*(4), 605–616. https://doi.org/10.14710/j.gauss.10.4.605-616

Mahdy, I. F. (2023). Pemodelan keluhan kesehatan dan indeks kebahagiaan di Indonesia tahun 2021 menggunakan pendekatan *local polynomial*. *STATMAT: Jurnal Statistika dan Matematika, 5*(1), 37–44. https://doi.org/10.32493/sm.v5i1.27380

Masduki, U., Rindayati, W., & Mulatsih, S. (2022). How can quality regional spending reduce poverty and improve human development index? *Journal of Asian Economics, 82*, Article 101515. https://doi.org/10.1016/j.asieco.2022.101515

Rahmarisa, F., Nasution, I., & Tahir, M. (2026). Tingkat pendidikan dan kualitas penduduk: Analisis Indeks Pembangunan Manusia di Indonesia. *JEKKP (Jurnal Ekonomi, Keuangan dan Kebijakan Publik), 8*(1), 46–56. https://doi.org/10.30743/jekkp.v8i1.14065

Sen, A. (1999). *Development as freedom*. Oxford University Press.

Sijabat, R. (2024). Democracy, human development, income distribution, and regional economic performance: A panel data analysis of 34 provinces in Indonesia. *Jurnal Ilmu Sosial dan Ilmu Politik, 28*(1), 73–93. https://doi.org/10.22146/jsp.76139

Suparti, & Prahutama, A. (2016). Pemodelan regresi nonparametrik menggunakan pendekatan polinomial lokal pada beban listrik di Kota Semarang. *Media Statistika, 9*(2), 85–93. https://doi.org/10.14710/medstat.9.2.85-93

Suryani, Y., & Rinaldy, R. (2023). Analysis of the influence factors of the Human Development Index in Padang City, West Sumatra. *Economics, Business, Accounting & Society Review, 2*(2), 120–133. https://doi.org/10.55980/ebasr.v2i2.65

Tyas, D. P. P., & Sukartini, N. M. (2022). Determinants of the Human Development Index (HDI) in Indonesia, 2014–2021. *Media Trend, 17*(2), 481–495. https://doi.org/10.21107/mediatrend.v17i2.14610

United Nations Conference on Trade and Development. (2024). *World investment report 2024: Investment facilitation and digital government*. United Nations. https://unctad.org/system/files/official-document/wir2024_en.pdf

United Nations Development Programme. (2024). *Human development report 2023–24: Breaking the gridlock: Reimagining cooperation in a polarized world.* https://hdr.undp.org/content/human-development-report-2023-24