Jurnal Litbang Provinsi Jawa Tengah
Vol. 22, No. 2, December 2024, p. 231 – 241

Spatial Probit Regression Analysis Using Recursive Importance Sampling on the Human Development Index in Central Jawa Province

Analisis Regresi Probit Spasial Menggunakan Recursive Importance Sampling terhadap Indeks Pembangunan Manusia di Provinsi Jawa Tengah

Novia Amilatus Solekha[1], Nabilla Rida Tri Nisa[2]
[1]Depertment of Statistics, Faculty Science and Data Analytics, Institut Teknologi Sepuluh Nopember Indonesia
[2]Graphic Engineering Study Program, Departement of Industrial Technology, Politeknik Negeri Media Kreatif Indonesia

Article Info
Received: 22 August 2024
Revised: 13 December 2024
Accepted: 27 December 2024

Keywords:
Human Development Index (HDI)
Recursive Importance Sampling
Spatial Probit

Abstract

This research aims to analyze the Human Development Index (HDI) in Central Java Province using two modeling approaches: the probit model and the spatial probit model. HDI, a crucial indicator of quality of life, is categorized into two groups: medium-low and high. Among the 35 districts and cities in Central Java, 9 fall into the medium-low category (35%), while 26 are classified as high (65%). Brebes Regency has the lowest HDI at 69.54, whereas Sukoharjo City boasts the highest HDI at 77.73. The results reveal that the spatial probit model provides a superior classification of HDI compared to the probit model. The mapping of prediction results highlights the discrepancies between actual and predicted HDI categories. This study underscores the advantage of incorporating a spatial approach, enhancing the accuracy of HDI analysis, and offering a more nuanced understanding of spatial distribution and dependencies in human development within Central Java. Further research is recommended to include additional variables, such as poverty levels or access to education, which spatial factors may also affect. Moreover, this method can be applied to other regions or enhanced using more complex spatial models, such as spatial autoregressive (SAR) or geographically weighted regression (GWR), for deeper and more comprehensive analysis.

Kata kunci:
Indeks Pembangunan Manusia (IPM)
Probit Spasial
Recursive Importance Sampling

Abstrak

Penelitian ini bertujuan untuk menganalisis Indeks Pembangunan Manusia (IPM) di Provinsi Jawa Tengah dengan menggunakan dua pendekatan pemodelan: model probit dan model probit spasial. IPM, sebagai indikator penting kualitas hidup, dikategorikan ke dalam dua kelompok: sedang, rendah dan tinggi. Dari 35 kabupaten dan kota di Jawa Tengah, 9 daerah (35%) masuk dalam kategori sedang-rendah, sedangkan 26 daerah lainnya (65%) diklasifikasikan sebagai tinggi. Kabupaten Brebes memiliki IPM terendah sebesar 69,54%, sementara Kota Sukoharjo memiliki IPM tertinggi sebesar 77,73%. Hasil penelitian menunjukkan bahwa model probit spasial memberikan klasifikasi IPM yang lebih baik dibandingkan dengan model probit. Pemetaan hasil prediksi mengungkapkan adanya perbedaan

DOI: https://doi.org/10.36762/jurnaljateng.v22i2.1230

Corresponding Author:
Novia Amilatus Solekha
Noviasolekhah51@gmail.com
085600174585

antara kategori IPM aktual dan yang diprediksi. Studi ini menekankan keuntungan dari pendekatan spasial yang dapat meningkatkan akurasi analisis IPM serta memberikan pemahaman yang lebih mendalam mengenai distribusi spasial dan ketergantungan dalam pembangunan manusia di Jawa Tengah. Penelitian lebih lanjut disarankan untuk memasukkan variabel tambahan, seperti tingkat kemiskinan atau akses terhadap pendidikan, yang juga dapat dipengaruhi oleh faktor spasial. Selain itu, metode ini dapat diterapkan di wilayah lain atau dikembangkan dengan model spasial yang lebih kompleks, seperti Spatial Autoregressive (SAR) atau Geographically Weighted Regression (GWR), guna memperoleh analisis yang lebih mendalam dan komprehensif.

INTRODUCTION

The probit regression model is used to identify the relationship between predictor variables and categorical (qualitative) response variables. This model is known for employing the approach of the cumulative distribution function (CDF) of the normal distribution, which is designed to overcome the weaknesses of the linear probability model. According to Greene (2008), qualitative response variables in probit regression originate from unobserved latent variables. This model was first introduced by Chester Bliss in 1934.

Spatial probit regression is an extension of probit regression that incorporates spatial effects in the relationship between independent and dependent variables (Lesage & Pace, 2019). These spatial effects often arise when the data on categorical response variables are influenced by habits or opinions from surrounding environments, indicating the presence of spatial autocorrelation (Anselin, 1988). According to Fotheringham et al. (2002), spatial data refers to data with geographic references influenced by the location or position of the observed objects. If spatial autocorrelation is ignored, the use of a standard probit model may result in biased and inconsistent parameter estimates. To address this issue, spatial probit regression analysis was developed to explicitly include spatial elements in the model.

One of the estimation methods in spatial probit regression that has proven efficient is Recursive Importance Sampling (RIS). The RIS method simplifies the parameter estimation process by repeatedly sampling from the prior distribution. In the Human Development Index (HDI) analysis context, RIS enables more accurate estimation for regions with large datasets (Yang & He, 2020).

HDI measures human development achievements based on indicators of health (life expectancy), education (literacy rate and average years of schooling), and economy (purchasing power) (BPS, 2022). An example of applying spatial models is demonstrated by Beron and Verberg (2004), who found that similar decisions often influence a city's decision to increase taxes in neighboring areas. This research highlights the importance of considering spatial autocorrelation in economic models. McMillen (1992) also emphasized that ignoring spatial influences can lead to biased parameter estimates, which can be addressed through the development of spatial probit regression.

Previous studies have shown the advantages of spatial probit regression. Wang, Iglesias, and Wooldridge (2009) found that this model provides higher log-likelihood values compared to the standard probit model, indicating better performance. Research by Kurniawan & Hartono (2019) compared spatial and non-spatial models in the context of regional development in Indonesia and found that the spatial model was more accurate in capturing interregional influences. Dewanto (2018) employed a spatial probit regression model with the RIS approach to analyze factors influencing health indices in Papua Island. The results revealed that variables such as access to healthcare services and education levels had significant impacts on health indices. Tumanggor & Simamora (2021) applied spatial probit regression to study factors affecting HDI in Indonesia. The highlighted the importance of spatial effects in understanding the dynamics of human development, although they did not use RIS.

Other studies, such as Calabrese & Elkink (2014), compared various estimation methods for binary spatial autoregressive models, including RIS, and concluded that RIS excels in estimation accuracy. Bivand & Piras (2015) also evaluated the implementation of various estimation methods for spatial econometrics, providing valuable insights into tools for spatial probit regression analysis.

This study aims to evaluate the probit model and spatial probit regression using the RIS approach with HDI data in Central Java in 2021. Furthermore, this study will analyze significant factors influencing HDI in the region and compare the performance of the standard probit model with the spatial probit model. The results of this research are expected to provide significant contributions to regional development strategies, particularly through more accurate spatial approaches in understanding interregional dependencies.

RESEARCH METHODS

This research uses secondary data. The data are obtained from Statistics Indonesia (BPS) Central Java Province in 2021. The research variables consist of one response variable and six predictor variables (BPS, 2022). The response variable (Y) is the Human Development Index (HDI). This response variable is categorical as explained in Table 1.

Table 1. Research Variable

Variable	Information	Type Data
Y*	Human Development Index (HDI/IPM)	Nominal
X1	Life Expectancy (UHH)	Ratio
X2	Expected Years of Schooling (HLS)	Ratio
X3	Per Capita Income	Ratio
X4	Open Unemployment Rate	Ratio
X5	Percentage of Population Living in Poverty	Ratio
X6	Percentage of Population with Health Complaints	Ratio

Source: Statistics Indonesia, 2022

Based on latest categorization according to Statistics Indonesia (BPS, 2015), Human Development Index (Y*) are categorized as below:

i. 0 = group low if HDI < 60
ii. 1 = medium group If 60 ≤ HDI ≤ 70
iii. 2 = high group if 70 ≤ HDI ≤ 80
iv. 3 = group very tall HDI ≥ 80

In this research, if the Human Development Index (HDI) is classified according to BPS regulations into four categories, some categories will be empty. In 2021, regencies and cities in Central Java fell into the low, medium, and upper-medium categories.

Therefore, for the purposes of this study, the HDI distribution is simplified into two categories:

- 0 = Intermediate Low
  Combined from category low and currently that is If HDI< 70
- 1 = Intermediate On
  Categories tall and very high i.e. If HDI ≥ 70

The methods and stages of analysis to achieve the objectives of this research are as follows:

1. Conduct descriptive analysis of each variable as an initial description of HDI in Central Java.

2. Identify and handle multicollinearity problems by looking at the VIF criteria
3. Create a weighting matrix between locations using queen contiguity for a spatial probit model.
4. Estimating parameters and forming a probit model then interpreting the results.
5. Estimating parameters and forming a spatial probit model then interpreting the results.
6. Comparing the results of the probit model and the spatial probit model then concluding the results that have been obtained.

RESULTS AND DISCUSSION

1. Descriptive Analysis

Central Java consists of 35 districts/cities, including 29 districts and 6 cities. In this research, Human Development Index (HDI) achievements are categorized into two groups: low-medium and high. The low-medium category includes 9 districts/cities, while the high category comprises 26 districts/cities. In percentage terms, 35% fall into the low-medium category, and 65% are in the high category. The region in Central Java Province with the lowest HDI achievement is Brebes Regency, with a score of 69,54; while the highest HDI achievement is in Sukoharjo City, with a score of 77,73.

The distribution of the Human Development Index in Central Java Province by district/city in 2021 is shown in Figure 2. The color gradient in the figure indicates that lighter colors represent regions with low-medium HDI values, while darker colors represent regions with high HDI values. Additionally, based on Figure 2, it can be observed that adjacent areas tend to have relatively similar HDI values, indicating spatial dependence in the HDI data for Central Java in 2021.

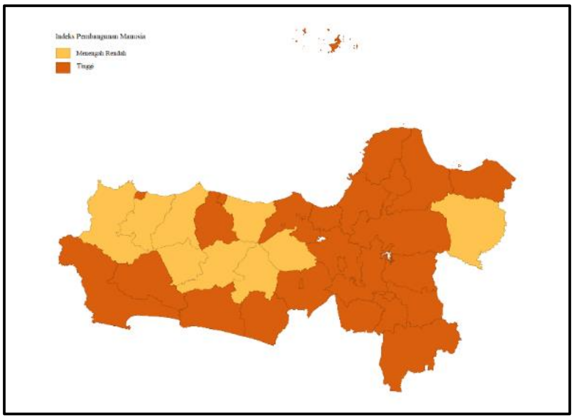

Figure 1. HDI Distribution Map by Regency/City in Central Java in 2021
Source: ArcMap 10.7.1 (processed)

In this research, there are six predictor variables which have an influence on HDI in Central Java in 2020, namely Life Expectancy (X1), Expected Years of Schooling (X2), Per Capita Income (X3), Open Unemployment Rate (X4), Percentage of Population Living in Poverty (X5), and Percentage of Population with Health Complaints (X6). The descriptive data for each variable attached are described in Table 2.

Table 2. Descriptive analysis of variables

Variable	Min	Median	Max	Mean
Y	66,32	72,36	83,60	72,85
X1	69,54	74,84	77,73	75,01
X2	11,63	12,85	15,53	12,97
X3	8573,00	10534	15843	11139
X4	3,85	6,070	9,83	6,40
X5	4,56	10,68	17,83	11,39
X6	4,00	32,00	102,00	35,11

Source: R Studio 4.3.0 (processed)

Based on Table 2, the highest HDI value is 83,60 and the lowest HDI value is 66,32. Besides that, the average is 72,85. Based on this data, it can be concluded that the HDI data is relatively homogeneous.

2. Multicollinearity Identification

According to regression analysis, multicollinearity can significantly impact the accuracy of parameter estimates. The Variance Inflation Factor (VIF) is used to measure the severity of multicollinearity in regression analysis. Therefore, it is important to assess whether there is a multicollinearity problem among the predictor variables. Multicollinearity can be detected by observing the VIF value. If any predictor variable has a VIF value greater than 10, it indicates the presence of multicollinearity in the model. VIF values for each predictor variable was calculated using the software detailed in Appendix 3. The results are shown in Table 3.

Table 3. Multicollinearity Test Result

Variable	VIF Value
X1	3.752
X2	4.131
X3	2.688
X4	2.104
X5	2.262
X6	1.362

Source: R Studio 4.3.0 (processed)

Based on the table above, it can be concluded that all variables have a VIF value <10. This means that multicollinearity does not occur between variables.

3. IPM Spatial Dependency Test

When analyzing cross-sectional data, the effects of spatial autocorrelation or spatial dependency can manifest in two forms. The first is spatial lag, which occurs when information in one observation area is correlated with that in other areas intersecting with it. The second is spatial error, indicating that correlation between regions affects the response variables and the predictor variables (Anselin, 1988).

This research uses the Lagrange Multiplier to conduct the spatial dependency test. The spatial model used in this study is limited to the spatial lag model, so only the Lagrange Multiplier Lag test is employed to identify spatial lag dependencies in the Human Development Index in Central Java. The spatial weighting matrix (W) is provided in Appendix 2.

The testing steps used for this research are as follows:

a. Hypothesis used:

H_0 : $\rho = 0$ or the absence of spatial lag dependencies

H_1 : $\rho \neq 0$ or there is a spatial lag dependency

b. Level of Significance:

In this research, the level of significance αused is 0.05

c. Test Statistics:

$$LM = \frac{\left(\frac{e^t W y}{\sigma^2}\right)^2}{D} \quad (1)$$

In this study, the LM-lag value was 0.771

d. Rejection criteria

Ho is rejected if LM > $\chi_{(\ ,1)}$

e. Conclusion

Based on the test results in Appendix 4, it was found that the LM-lag value = 0.771 > $\chi_{(1)}$ = 3.841. It indicates that H_0 is rejected, which shows that in the model, there is a spatial lag effect in modelling the human development index in Central Java in 2021.

4. HDI with Probit Model

The first step in modeling HDI with a probit model is to estimate the parameters of the probit model. The following are the results of estimating the probit model parameters.

Table 4. Model Parameter Estimation

Variable	Coefficient	Std Error	Z-value	P-value	Decision
Intercept	-845.900	399200.000	-0.002	0.998	Failed to Reject H $_0$
X1	7.092	5.568	0.001	0.999	Failed to Reject H $_0$
X2	15.000	25120.000	0.001	1.000	Failed to Reject H $_0$
X3	0.012	9.626	0.001	0.999	Failed to Reject H $_0$
X4	-2.848	3524.000	-0.001	0.999	Failed to Reject H $_0$
X5	0.954	2748.000	0.000	1.000	Failed to Reject H $_0$
X6	0.618	461.700	0.001	0.999	Failed to Reject H $_0$

Source: R Studio 4.3.0 (processed)

Based on Table 4, it can be concluded that there are no significant variables in the district or city Human Development Index category at a significance level of α=5%. The resulting probit model is as follows:

$$\widehat{Y}_i = -845.900 + 7.092X_1 + 15.00X_2 + 0.012X_3 - 2.848X_4 + 0.954X_5 + 0.618X_6$$

From the model above, predictions of the actual HDI classification can be calculated and then used to assess classification accuracy. The accuracy of the classification between actual data and predicted data is shown in Table 5. This table indicates that the model correctly classifies 57,1% of the observations.

Table 5. Accuracy of Probit Model Classification

Actual	**Predicted**	
	0	**1**
0	6	0
1	15	14

Source: R Studio 4.3.0 (processed)

5. HDI Modelling with Spatial Probit Models

The first step in modeling HDI classification using spatial probit regression is to estimate the parameter values of the spatial probit model. The significance test of these parameters aims to determine whether there is a linear relationship between the dependent variable (y) and the independent variables (x). Parameter testing in bivariate probit regression is conducted in two stages: simultaneous parameter testing and partial parameter testing. The spatial probit regression model is formed as follows:

$$[\boldsymbol{\Gamma X\beta}]_i = \boldsymbol{\beta_0} + \boldsymbol{\rho}\left(\sum_{i=1,j\neq i}^{n} \boldsymbol{w_{ij} y_i}^*\right) + \boldsymbol{\beta_1 X_{1i+} \beta_2 X_{2i+} \beta_3 X_{3i}} + \boldsymbol{\beta_5 X_{5i}} \quad (2)$$

a. Simultaneous Testing

Simultaneous parameter testing in spatial probit regression is used to test the role of the coefficient $\boldsymbol{\beta}$.

The testing steps are as follows:

1. Hypothesis used:
   H$_0$: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$
   H$_1$: There is at least one $\beta_j \neq 0$ for *j=1,2,3,4,5,6*
2. Level of Significance:
   In this research, the level of significance α used is 0,05
3. Test Statistics:
   $$\boldsymbol{G^2 = -2ln\left[\frac{L(\widehat{\omega})}{L(\widehat{\Omega})}\right]} \quad (3)$$
   In this research, a value $G^2$of 45.649 was obtained.
4. Rejection criteria
   Ho is rejected if $G^2 > \chi_{(\ ,5)}$
5. Conclusion
   Based on the test results in Appendix 4, it was found that the value G^2= 45.649 > $\chi_{(\alpha,5)}$ = 11.07. Consequently, the null hypothesis H$_0$ is rejected, indicating the presence of at least one significant parameter in the model at a 95% confidence level. As a result, the next step in the modeling process is to conduct partial parameter testing to identify the predictor variables that significantly influence the response variable.

b. Partial Testing

Simultaneous partial testing is a testing to examine whether each parameter has a significant effect on the response variable. The following are the results of spatial autoregressive (SAR) probit parameter estimation which are presented in Table 4. The test steps used for this research are as follows:

1. Hypothesis used:
   H$_0$: $\beta_j = 0$
   H$_1$: $\beta_j \neq 0$for *j=1,2,3,4,5,6*
2. Level of Significance:
   In this research, the level of significance α used is 0,05
3. Test Statistics:
   $$\mathbf{Z} = \frac{\widehat{\boldsymbol{\beta}}_{\mathbf{j}}}{\mathbf{SE}(\widehat{\boldsymbol{\beta}}_{\mathbf{j}})} \quad (4)$$
   Where $\widehat{\beta_j}$ is the parameter estimator β_j and $SE(\widehat{\beta_j})$ is $\sqrt{\widehat{Var}(\widehat{\beta_j})}$
4. Rejection criteria
   $Z_{hitung} < -Z_{\alpha/2}$ or $Z_{hitung} > Z_{\alpha/2}$ or p-value < α
5. Conclusion

Based on the test results in Appendix 4, the following significant test results were obtained:

Table 6. Significance Test Result

Variable	Min	Median	Max
Intercept	0.31434	16.75053	8.6 x 10 $^{-16}$
Y	-0.25252	-6.20391	1.2 x 10 $^{-06}$
X1	0.47319	23.35408	1.9 x 10 $^{-19}$
X2	0.00173	7.38070	6.1 x 10 $^{-08}$
X3	-0.46408	-6.44580	6.5 x 10 $^{-07}$
X4	-0.23357	-13.58642	1.3 x 10 $^{-13}$
X5	0.05532	6.10622	1.5 x 10 $^{-06}$
X6	0.31434	16.75053	8.6 x 10 $^{-16}$

Source: R Studio 4.3.0 (processed)

Table 6 shows that all variables are significant in the district or city human development index category at the real level or α=5%. It can be seen that all variables X have a significant effect on the model. So, the model resulting from health index modelling using SAR probit is as follows:

$$[\boldsymbol{\Gamma X\beta}]_i = 0.31434 - \mathbf{0.2487}\left(\sum_{i=1,j\neq i}^{35} \boldsymbol{w_{ij}y_i}^*\right)$$
$$-0.25252\boldsymbol{X_{1i}} + 0.47319\boldsymbol{X_{2i}} + 0.00173\,\boldsymbol{X_{3i}}$$
$$-0.46408\,\boldsymbol{X_{4i}} - 0.23357\,\boldsymbol{X_{5i}} + 0.05532\,\boldsymbol{X_{6i}}$$

After obtaining the model, the next step is to calculate the probability prediction values. The complete results of these calculations, performed using software, are provided in Appendix 6. These prediction values can be used to create spatial effects maps.

A confusion matrix is commonly used to evaluate the performance of classification models in machine learning. This table provides detailed information on the number of correctly or incorrectly classified data points. It can be used to assess the accuracy of predictions against observations (real-world data or research results). From this table, three key metrics accuracy, sensitivity, and specificity can be derived to evaluate the effectiveness of a model in classification. The complete confusion matrix for this study is presented in Table 7.

Table 7. Health Index Confusion Matrix Table

Actual results	Prediction Results	
	P	N
P	*True Positive* (9)	*False Negatives* (0)
N	*False Positives* (13)	*True Negative* (13)

Source: R Studio 4.3.0 (processed)

Based on Table 7, the values for accuracy, sensitivity, and specificity can be calculated as follows:

Accuracy

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$
$$= \frac{9+13}{35} = 0.628$$

Based on the results, the model demonstrates an accuracy of 0,628, or 62,8%. It indicates that the model classifies 62,8% of the tested objects correctly, reflecting its overall predictive reliability.

Sensitivity

$$Sensitivity = \frac{TP}{TP+FN} \quad (6)$$
$$= \frac{9}{9} = 1$$

The model's sensitivity is 1, or 100%. This indicates that the model correctly identifies 100% of the health indexes that are truly in the high category.

Specificity

$$Specificity = \frac{TN}{TN+FP} \quad (7)$$
$$= \frac{13}{26} = 0.5$$

It can be obtained that the specificity value for the model is 0,5 or 50%. This means that the proportion of health indexes in the low-medium category are also identified by the model as objects that have a high-category health index of 50 percent.

The marginal effect or spill over is the average effect at the population level or the magnitude of the impact of changes in the dependent variable in region-i, due to changes in predictors in region-j. Marginal effects for binary variables lead to discrete changes. It can be said that the marginal effect is used to identify the size of the influence of changes in each predictor variable on the response variable only if the other variables are constant. It can be known according to the value of the marginal effect which is attached in Appendix 7. For example, the value of the marginal effect of life expectancy (X1) for Cilacap Regency is -0,02720. This shows that every change in the life expectancy variable (X1) by one unit will reduce the probability of the district achieving a high category health index by 2,72%. The relationship between the per capita expenditure variable and the health index shows a negative relationship. It indicates that if the value of per capita expenditure increases, the health index value decreases, and vice versa.

6. Comparison of Probit Model Result and Spatial Probit Models

The variables influencing the Human Development Index (HDI) in Central Java were identified after conducting simulation using both the probit and spatial probit models. The results indicate that the spatial probit model outperforms the probit model in terms of classification accuracy and its superior ability to correctly categorize HDI. To provide a clearer perspective, Figure 2 below illustrates a map displaying the categorization of HDI in Central Java based on the SAR Probit model. The map shows two HDI categories: high and low-medium. Regions shaded in darker colors represent areas with a high HDI, while lighter shades correspond to low-medium HDI. According to the model's predictions, 13 districts/cities fall into the high HDI category, while 22 districts/cities are classified as low-medium HDI. This visualization emphasizes the spatial probit model's effectiveness in capturing spatial dependencies and providing an accurate classification of HDI across Central Java.

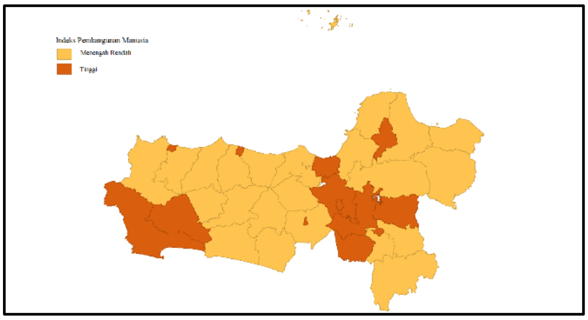

Figure 2. Map of Human Development Index Prediction Results Spatial Probit Regression Model
Source: ArcMap 10.7.1 (processed)

The comparison between actual and predicted Human Development Index (HDI) categories for districts and cities is presented in Table 8, providing a detailed breakdown of how each region aligns with or deviates from the model's predictions. For instance, Jepara Regency, categorized as having a low to medium HDI based on actual data, is predicted to belong to the high HDI category when analyzed using the SAR Probit model. This intriguing discrepancy is not isolated, as similar patterns are observed in other districts and cities, highlighting the spatial model's capability to offer unique perspectives and insights.

Table 8 reveals four distinct groupings that illustrate the relationship between actual data and predictions. Group 1 consists of 9 districts and cities that align perfectly with the actual HDI data, demonstrating the model's reliability in correctly classifying these regions. Interestingly, Group 2 is entirely empty, signifying no instances of overestimation where districts/cities with low HDI were predicted to be in the same category but incorrectly classified. On the other hand, Groups 3 and 4 each comprise 13 districts/cities, reflecting areas where the model captured varying levels of development dynamics.

This grouping provides a clear visualization of the SAR Probit model's strengths in detecting spatial dependencies and capturing regional variations in human development. However, the presence of discrepancies, such as Jepara Regency's misclassification, suggests that while the spatial model provides a significant improvement in classification accuracy, there remains room for refinement to enhance its predictive precision and better account for localized factors influencing HDI. These findings offer valuable insights for policymakers seeking to understand and address disparities in human development across Central Java.

Table 8. Results of Grouping Actual and Predicted Data for Districts/Cities in Central Java using the Probit Model

Actual	Prediction	Amount	Regency/City
1	1	6	Banjarnegara Regency, Batang Regency, Blora Regency, Brebes Regency, Pemalang Regency, Purbalingga Regency
1	0	0	Blank (None)
0	1	16	Demak Regency, Grobogan Regency, Jepara Regency, Karanganyar Regency, Kebumen Regency, Kendal Regency, Magelang Regency, Pati Regency, Pekalongan Regency, Purworejo Regency, Rembang Regency, Sukoharjo Regency, Wonogiri Regency, Tegal Regency, Temanggung Regency
0	0	14	Banyumas Regency, Boyolali Regency, Cilacap Regency, Klaten Regency, Kudus Regency, Semarang Regency, Sragen Regency, Magelang City, Pekalongan City, Salatiga City, Semarang City, Surakarta City, Tegal City, Wonosobo Regency

Source: R Studio 4.3.0 (processed)

Table 8 above can be visualized as a map, providing a clearer depiction of the distribution of districts and cities across different HDI categories based on actual and predicted data. For further analysis, it is observed that 57,1% of the classifications for Central Java Province align correctly with the actual HDI data, reflecting a strong level of accuracy in the model's predictions. However, 42,9% of the classifications are incorrect, highlighting areas where the model's performance can be refined.

The classification accuracy achieved in Central Java Province demonstrates the model's capability to effectively categorize HDI across regions. Nevertheless, the presence of misclassifications indicates that certain regional characteristics and spatial dependencies may not have been fully captured. For example, misclassified districts or cities might reflect unique local factors, such as disparities in resource allocation, economic activities, or policy implementations, which are not adequately accounted for in the model.

Table 9. Results of Grouping Actual and Predicted Data for Districts/Cities in Central Java using the Spatial Probit Model

Actual	Prediction	Amount	Regency/City
1	1	9	Banjarnegara Regency, Batang Regency, Blora Regency, Brebes Regency, Pemalang Regency, Purbalingga Regency, Tegal Regency, Temanggung Regency, Wonosobo Regency
1	0	0	Blank (None)
0	1	13	Demak Regency, Grobogan Regency, Jepara Regency, Karanganyar Regency, Kebumen Regency, Kendal Regency, Magelang Regency, Pati Regency, Pekalongan Regency, Purworejo Regency, Rembang Regency, Sukoharjo Regency, Wonogiri Regency
0	0	13	Banyumas Regency, Boyolali Regency, Cilacap Regency, Klaten Regency, Kudus Regency, Semarang Regency, Sragen Regency, Magelang City, Pekalongan City, Salatiga City, Semarang City, Surakarta City, Tegal City

Source: R Studio 4.3.0 (processed)

Table 9 presents the results of grouping actual and predicted HDI data for districts and cities in Central Java using the Spatial Probit Model. The table categorizes districts and cities into four groups based on their actual and predicted classifications. Group 1 consists of 9 districts and cities, including Banjarnegara, Batang, and Brebes Regency, correctly classified as having high HDI. Notably, Group 2, which would include regions misclassified as low HDI despite their actual classification as high, is empty, indicating no such errors.

Group 3 includes 13 districts and cities, such as Demak, Jepara, and Sukoharjo Regency, which were misclassified as having high HDI when their actual category is low-medium. Meanwhile, Group 4 comprises 13 regions, including Banyumas, Kudus, and Semarang City, correctly classified as low-medium HDI.

These groupings can be visualized on a map, as shown in Figure 3, providing a clear spatial representation of HDI classifications across Central Java. The analysis reveals that 62.8% of the classifications are accurate, while 37.2% are misclassified, indicating good overall accuracy for the Spatial Probit Model.

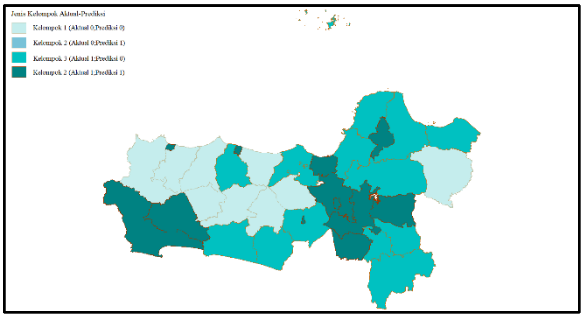

Figure 3. Grouping Map of Actual and Predicted Data on Human Development Index
Source: ArcMap 10.7.1 (processed)

CONCLUSION

The conclusions of this study highlight that the Spatial Probit Model significantly improves classification accuracy compared to the standard probit model, particularly in identifying spatial dependencies in Human Development Index (HDI) data. This model provides valuable insights into regional development patterns and a deeper understanding of the spatial distribution of human development, making it an effective tool for spatial data analysis.

In practical applications, this model can assist local governments and policymakers in designing more targeted development strategies. With high sensitivity in identifying regions with high HDI, the model enables better prioritization of resource allocation in areas requiring greater attention. Although there is room to improve specificity in classifying regions in the low-medium category, this approach offers a robust analytical framework for understanding the complex dynamics of human development. The integration of spatial approaches in development data analysis can support more efficient evidence-based policymaking, particularly in addressing development disparities across regions in Central Java.

SUGESSTION

Overall, this research strengthens the argument that spatial models are very relevant in analyzing data that has a geographic component. Future researchers can utilize these results to develop more sophisticated models or apply different methods to deeply explore some factors influencing HDI in other regions or with other variables that may also be influenced by spatial factors.

BIBLIOGRAPHY

Anselin, L. (1988). *Spatial Econometrics: Method and Models.* Dordrecht: Kluwe Academic Publishers.

Beron, K. J., & Verburg, T. (2004). "Spatial externalities and their estimation using spatial econometric models." *Environmental and Resource Economics*.

Bivand, R. S., & Piras, G. (2015). *Comparing Implementations of Estimation Methods for Spatial Econometrics. Journal of Statistical Software,* 63(18), 1-36.

Bliss, C. I. (1934). The Method of Probits. *Science*, 79(2037), 38–39.

BPS. (2020). *Survei Angkatan Kerja Nasional*. Jakarta: Badan Pusat Statistik.

BPS. (2021). *Kemiskinan dan Ketimpangan*. Jakarta: Badan Pusat Statistik.

BPS. (2022). *Harapan Lama Sekolah*. Jakarta: Badan Pusat Statistik.

BPS. (2022). *Indeks Pembangunan Manusia*. Jakarta: Badan Pusat Statistik.

BPS. (2022). *Penduduk yang Mempunyai Keluhan Kesehatan*. Jakarta: Badan Pusat Statistik.

BPS. (2022). *Pengeluaran per Kapita*. Jakarta: Badan Pusat Statistik.

BPS. (2022). *Usia Harapan Hidup*. Jakarta: Badan Pusat Statistik.

Calabrese, R., & Elkink, J. A. (2014). *Estimating Binary Spatial Autoregressive Models for Rare Events: A Comparative Study of Estimators*. *Political Analysis*, 22(2), 183-205.

Dewanto, T. F. (2018). *Model Regresi Probit Spasial dengan Pendekatan Recursive Importance Sampling (Studi Kasus: Indeks Kesehatan Tahun 2016 di Pulau Papua)*. Tesis, Institut Teknologi Sepuluh Nopember.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley.

Greene, W. H. (2008). *Econometric Analysis* (6th ed.). Pearson.

Kurniawan, A., & Hartono, D. (2019). "Penerapan Model Spasial dalam Analisis Pembangunan Regional." *Jurnal Ekonomi Regional*.

Lesage, J., & Pace, R. K. (2019). *Introduction to Spatial Econometrics*. CRC Press.

McMillen, D. P. (1992). Probit with Spatial Autocorrelation. *Journal of Regional Science*, 32(3), 335–348.

Tumanggor, J. R., & Simamora, A. H. (2021). *Pemodelan Faktor-Faktor yang Mempengaruhi Indeks Pembangunan Manusia di Indonesia dengan Pendekatan Regresi Probit Spasial*. *Jurnal Riset Matematika dan Ilmu Pengetahuan Alam*, 6(2), 123-134.

Yang, S., & He, H. (2020). Recursive Importance Sampling for Spatial Probit Models. *Journal of Computational Statistics*, 35(3), 451–467.