Vol. 12 No. 1, pISSN: 2303-0577 eISSN: 2830-7062 http://dx.doi.org/10.23960/jitet.v12i1.3945

GWRPCA ALGORITHMIC FRAMEWORK: ANALYZING SPATIAL DYNAMICS OF POVERTY IN EAST JAVA PROVINCE

Harun Al Azies[1,2*], Noval Ariyanto[1]

[1]Study Program in Informatics Engineering, Faculty of Computer Science, Universitas Dian Nuswantoro, 50131, Semarang, Indonesia
[2]Research Center for Materials Informatics, Faculty of Computer Science, Universitas Dian Nuswantoro, 50131, Semarang, Indonesia

Riwayat artikel:

Received: 29 November 2022
Accepted: 29 Desember 2023
Published: 1 Januari 2024

Keywords:
Geographically Weighted Regression;
Principal Component Analysis;
Regression Algorithms;
Spatial Diversity.

Corespondent Email:
harun.alazies@dsn.dinus.ac.id

Abstract. *This study employs Regression Principal Component Analysis (RPCA) and Geographically Weighted Regression Principal Component Analysis (GWRPCA) algorithms to analyze poverty in East Java Province, using data from Statistics Indonesia (BPS). The research investigates regency/city-level poverty percentages and identifies influential factors such as education, literacy rates, housing conditions, and economic indicators. The results reveal that GWRPCA, with an 85.10% R^2 value, outperforms RPCA, highlighting its effectiveness in capturing spatial diversity and providing a nuanced portrayal of poverty characteristics across regencies/cities in East Java. In conclusion, GWRPCA emerges as a powerful algorithmic tool for informing targeted poverty alleviation policies, offering insights into spatial variations. The study suggests future research directions to explore evolving spatial patterns and consider additional variables for a more comprehensive analysis. The findings emphasize the significance of spatial considerations in devising effective, context-specific strategies for each regency/city in East Java.*

Abstrak. Penelitian ini menggunakan algoritma Regression Principal Component Analysis (RPCA) dan Geographically Weighted Regression Principal Component Analysis (GWRPCA) untuk menganalisis kemiskinan di Provinsi Jawa Timur, dengan menggunakan data dari Badan Pusat Statistik (BPS). Penelitian ini menyelidiki persentase kemiskinan tingkat kabupaten/kota dan mengidentifikasi faktor-faktor yang berpengaruh seperti pendidikan, tingkat melek huruf, kondisi perumahan, dan indikator ekonomi. Hasilnya menunjukkan bahwa GWRPCA, dengan nilai R^2 sebesar 85,10%, mengungguli RPCA, hal ini menunjukkan efektivitasnya dalam menangkap keragaman spasial dan memberikan gambaran yang berbeda mengenai karakteristik kemiskinan di seluruh kabupaten/kota di Jawa Timur. Kesimpulannya, GWRPCA muncul sebagai alat algoritmik yang ampuh untuk menginformasikan kebijakan pengentasan kemiskinan yang ditargetkan, dan menawarkan wawasan tentang variasi spasial. Studi ini menyarankan arah penelitian di masa depan untuk mengeksplorasi pola spasial yang berkembang dan mempertimbangkan variabel tambahan untuk analisis yang lebih komprehensif. Temuan ini menekankan pentingnya pertimbangan spasial dalam merancang strategi yang efektif dan spesifik konteks untuk setiap kabupaten/kota di Jawa Timur.

1. INTRODUCTION

In an era of sustainable development, poverty is still a major problem for many nations, including Indonesia [1]. One of Indonesia's most populous provinces, East Java, exhibits complicated dynamics in its attempts to combat poverty. East Java's regencies and cities exhibit notable disparities in income distribution and resource accessibility, notwithstanding the country's swift economic expansion [2]. There are geographical differences in poverty in East Java, and spatial data is essential to comprehending these dynamics [3]. Regional and observational data are combined to make spatial data, which has the potential to provide both spatial dependency and spatial variability [4], [5]. Geographically Weighted Regression (GWR) becomes a useful technique for handling spatial variability in this situation [6].

Previously, several researchers conducted studies on poverty indicators in Indonesia, such as Pratama's study in Lampung Province (2015-2019) utilizes spatial analysis, revealing clustered patterns and positive Moran's I value. The findings emphasize the importance of inter-regional policies for addressing multidimensional poverty [7]. Meanwhile, Belantika's research on Java Island, using geographically weighted regression (GWR) and 2020 data from BPS and government websites, categorizes affected districts based on factors like HDI, TPT, MSEs, and UMK. The study suggests targeted improvements in education, public health, and job training to alleviate poverty in these regions [8]. As technology progresses, the use of spatial algorithms and models to analyze and evaluate differences in conditions in each place becomes increasingly vital. Geographically weighted regression principal component analysis, which combines the benefits of GWR and principal component analysis [9], is one of the techniques under consideration in this study.

The study's context is hampered by East Java Province's vast size and socioeconomic diversity between regions. Furthermore, the presence of geographical influences and disparities in poverty levels across regencies and cities necessitates a more sophisticated analytical approach. At the local level, GWRPCA is seen to be an efficient strategy for dealing with spatial heterogeneity and local multicollinearity in poverty data. This study also depends on previous research, which explored poverty issues in East Java using spatial models such as GWR. This work, on the other hand, goes a step further by combining GWR with PCA, allowing for the discovery of critical aspects that traditional models may not adequately capture.

This research is presented to make a substantial new contribution to a deeper understanding of regional disparities in poverty causes by focusing on the application of GWRPCA in poverty analysis in East Java. The findings of this study are supposed to provide more precise and complete knowledge, as well as a solid foundation for establishing policies that are more targeted and effective in reducing poverty levels at the local level.

2. LITERATURE REVIEW

2.1. Classic Regression Models

The classical regression model, often known as classical linear regression, is a statistical method for modeling the connection between a dependent variable and one or more independent variables [10]. This model assumes that all locations or regions have the same constant parameters [11]. Linear regression models in general:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \cdots + \beta_p x_p + \varepsilon \qquad (1)$$

In the given model context. $\beta_1, \beta_2, \cdots, \beta_p$ represent the model parameters. ε represents the observation error or error term. However, in the context of geographical analysis, such as this study, traditional regression can be limited if spatial patterns and local variability are not taken into consideration appropriately[12].

2.2. Geographically Weighted Regression

Geographically weighted regression is a regression method that considers spatial variations in variable relationships[13], [14]. GWR models in general:

$$\hat{y}_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i)\, x_{ik} + \varepsilon_i \quad (2)$$

y_i the observed value of the response variable for location i. u_i, v_i, denotes the coordinate point (longitude, latitude) for location i. $\beta_k(u_i, v_i)$, the regression coefficient of predictor variable k for location i. GWR weights each data point differently according to its spatial distance, allowing the model to adjust to local changes in the relationship structure[15]. In other words, GWR enables regression parameters to vary geographically, allowing for a better understanding of changing patterns across regions [16].

2.3. *Principal Component Analysis*

The principal component analysis approach is a statistical method for reducing the dimensionality of connected variables in a dataset [17]. The original data is converted into principal components, which are a linear combination of the starting variables, by PCA[18]. The application of PCA can aid in the identification of patterns and structures in complicated data as well as the reduction of multicollinearity issues[19].

2.4. *Geographically Weighted Regression Principal Component Analysis*

Geographically weighted regression principal component analysis is a type of principal component analysis. The advantages of GWR and PCA are combined in GWRPCA [20]. To understand complicated spatial patterns, GWRPCA allows for spatial variation in variable interactions while lowering variable dimensions. The GWRPCA is expected to provide a greater understanding of the factors that drive poverty in each region by emphasizing local trends. GWRPCA models in general:

$$\hat{y}_i = \beta_0(u_i, v_i) + \sum_{j=1}^{k} \beta_j(u_i, v_i)\, PC_{ij} + \varepsilon_i \quad (3)$$

For i = 1, 2, ..., n and j = 1, 2, ..., k, the variables and parameters are defined as follows, y_i the observed value of the response variable for location i. PC_{ij} the observed value of the j-th principal component at observation location i. $\beta_0(u_i, v_i)$, the constant/intercept at observation location i. u_i, v_i, denote the coordinate point (longitude, latitude) for location i. $\beta_j(u_i, v_i)$, the regression coefficient of the j-th independent variable at observation location i. ε_i, the observation error at location i.

The combination of GWR and PCA in GWRPCA is predicted to overcome earlier models' weaknesses, particularly when dealing with geographical heterogeneity and local multicollinearity [21]. This strategy is projected to contribute significantly to clarifying the dynamics of poverty issues at the local level, as well as offering more precise and detailed knowledge to promote more effective policy decision-making.

2.5. *Related Work*

Poverty, as a complex global challenge, continues to be a primary focus in the sustainable development agenda, particularly in Indonesia. East Java Province, being the second most populous province after West Java, faces intricate dynamics between rapid economic growth and inequality in resource distribution. This phenomenon underscores the need for an in-depth understanding of the factors influencing poverty rates in this region.

Several previous studies, such as those conducted by Mahara et al., have contributed to analyzing the impact of the Human Development Index (HDI) and the total population percentage on poverty in Central Java Province. The utilization of Ordinary Least Squares (OLS) and Geographically Weighted Regression (GWR) models in this research highlighted the increase in the number of impoverished people in Central Java Province in 2020, concluding that the GWR model, particularly with the Adaptive Kernel Bisquare weighting function, yielded superior results compared to the OLS model [22]. In line with this, Xindong's study in Sichuan Province, China, explored the relationship between poverty and geographic factors using OLS and GWR. The research findings demonstrated that the GWR model was better suited to explaining the association between poverty levels and geographic features in Sichuan, with six physical variables remaining significant in both models [23]. Rinaldi et al., brought a different perspective by comparing the Spatial Error Model (SEM) and Geographic Weighted Regression (GWR) in analyzing poverty, emphasizing the importance of considering spatial effects and regional heterogeneity. This study affirmed that SEM could address spatial dependence in observational data errors, while GWR was effective in handling spatial heterogeneity [24].

This research, as a response and extension of previous studies, seeks to introduce an innovative approach by integrating Geographically Weighted Regression Principal Component Analysis (GWRPCA) in the context of East Java. Involving the strengths of GWR and Principal Component Analysis (PCA), the study aims to provide a more detailed and profound insight into the factors influencing poverty at the local level. Thus, it is expected that the findings of this research will make a significant contribution to the development of more adaptive policies to reduce poverty in East Java Province.

3. RESEARCH METHOD

The Research Methodology chapter is essential for understanding how this study is developed, how data is collected, and how analyses are carried out to answer the research questions. The actions used in this research method have a considerable impact on the validity and dependability of the results. This chapter provides a full overview of the data sources, research variables, and analytical methodologies used to identify the determinants influencing poverty at the regency/city level in East Java.

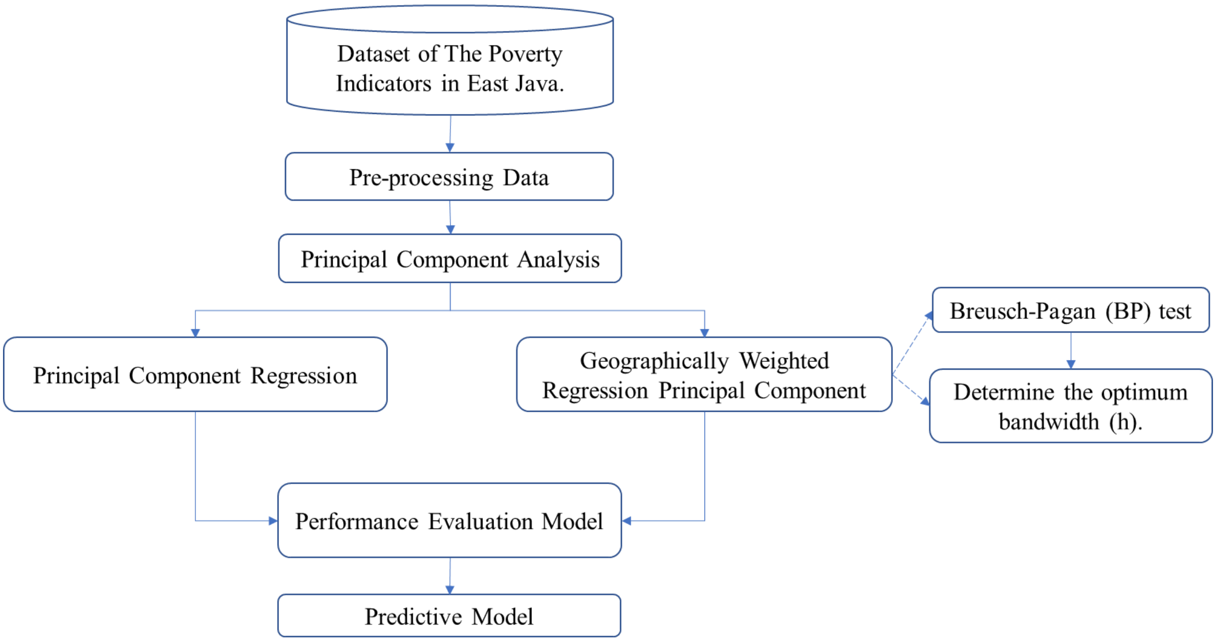

Figure 1. Framework of Analysis on Spatial Dynamics of Poverty in East Java Province

3.1. Data Collection

The initial step in this research is the data collection process. The primary data source is the Statistics Indonesia for the year 2016, providing a comprehensive overview of the socio-economic conditions in the region. The research variables encompass the percentage of poverty among the population at the regency/city level. Additionally, several predictor variables are considered, including average years of schooling, literacy rates, floor area per capita, residential status, access to electricity, per capita expenditure on food, types of residential land, and Regional Gross Domestic Product.

3.2. Data Preprocessing

Following the collection of data, the next stage is data preprocessing. Normalization is used to alter the variable scales to avoid undue influence on the analysis, while multicollinearity detection is used to find and address significant correlations across predictor variables. Normalization includes putting variable values into a standardized range [25], whereas multicollinearity detection requires assessing the strength of correlation among variables using statistical approaches such as variance inflation factor (VIF) analysis [26], [27]. The purpose of this stage is to improve the accuracy and reliability of the results so that they are ready for further analysis.

3.3. Model Development

After completing the data preprocessing stage, the next step in model development is implementing PCA analysis. PCA is utilized to reduce the dimensionality of variables while retaining significant information. This process is followed by the application of PCA regression, where the relationship between the principal variables and poverty is further analyzed. Subsequently, GWRPCA is employed to understand the spatial impact of these factors at the regency/city level in East Java. This step provides a more detailed overview of how poverty variability is influenced by the factors identified through PCA. The process explores spatial aspects and localities in the relationship between these variables, enriching the understanding of poverty dynamics in the region.

Model development continues with GWRPCA, involving the Breusch-Pagan test for residual homoskedasticity and the selection of optimal bandwidth [28]. The Breusch-Pagan test ensures the consistency of residual variance, while the selection of optimal bandwidth through cross-validation ensures the model can capture spatial variation without overfitting or underfitting. The integration of these two stages enhances the accuracy of the analysis and understanding of the factors influencing poverty rates in East Java spatially. The application of this model development aids in investigating the complexity of relationships between variables and provides in-depth insights into the factors influencing poverty rates in East Java.

3.4. Model Evaluation

The final stage in model development is the evaluation, where the best model is determined based on the R-squared value [29]. R-squared is a statistical measure that depicts the extent to which the variability of the response variable can be explained by predictor variables in the model. The higher the R-squared value, the better the model is at explaining data variation. In this context, after undergoing PCA analysis, PCA regression, and GWRPCA with the Breusch-Pagan test and optimal bandwidth selection, the evaluation step is carried out by comparing the R-squared values of each model. The model with the highest R-squared will be chosen as the best model [30], indicating its superior ability to explain and model the factors influencing the poverty rate in East Java. The selection of the best model based on the R-squared value is a critical concluding step to ensure accurate and relevant analysis results in the context of poverty research in the region [31].

4. RESULTS AND DISCUSSION

4.1. Overview of Poverty Conditions in East Java

The distribution of poverty in each regency and city in East Java is illustrated in Figure 2.

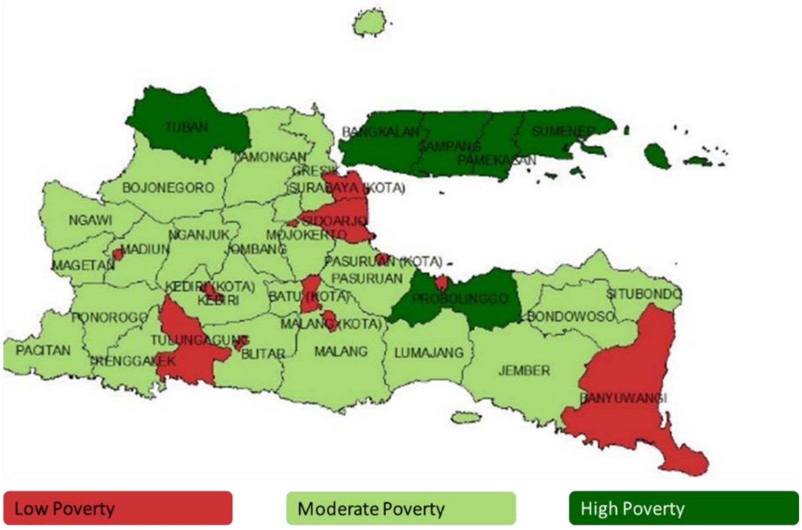

Figure 2 Distribution of Poverty in East Java

Figure 2 depicts the distribution of poverty in the province of East Java in detail. Six of the 38 regencies/cities are classified as high, with poverty rates ranging from 15.49% to 24.11%. Tuban Regency, Probolinggo, and four Madura Island regencies are among them. Surabaya City, Madiun, Pasuruan, Kediri, Blitar, Probolinggo, Batu, Malang, Tulungagung, Sidoarjo, and Banyuwangi Regencies, on the other hand, have low poverty rates ranging from 4.33% to 8.79%. This poverty distribution research serves as a foundation for evaluating the socioeconomic conditions in each region. Following that, these findings will be incorporated into a more detailed analytical model to explain the causes impacting the poverty rate in East Java at the regency/city level.

4.2. Model of Regression Principal Component Analysis (RPCA)

The classical regression model can be used to investigate the relationship between each regency or city in East Java Province's poverty status and the factors that influence the poverty rate. The predictor variables in the regression model are not multicollinear. The Variance Inflation Factors (VIF) criterion is used in this study to identify the presence of multicollinearity among predictor variables. VIF values greater than 10 imply collinearity between predictor variables.

Table 1. Multicollinearity Detection Results

Variable	VIF Value
X1	**18.365
X2	4.349
X3	1.780
X4	8.822
X5	1.416
X6	4.472
X7	2.002
X8	2.188

According to Table 1, the predictor variable with a VIF value greater than 10 is X1 (average schooling time at junior high school age). This signifies that there is a violation of the assumption in the regression model, particularly the presence of multicollinearity in the X1 variable. Principal Component Analysis (PCA) or Principal Component Analysis was utilized to eliminate multicollinearity in the regression model. Table 2 shows the findings of the Principal Component Analysis.

Table 2. Principal Component Analysis Results

Principal Component	Standard deviation	Proportion of Variance	Cumulative Proportion of Variance
1	1.947**	0.474**	0.474**
2	1.207**	0.182**	0.656**
3	0.965	0.116	0.772
4	0.905	0.102	0.875
5	0.700	0.061	0.936
6	0.561	0.039	0.975
7	0.403	0.020	0.996
8	0.186	0.004	1

Note: **) represents the selected principal components with eigenvalues greater than 1.

Table 2 shows the findings of principal component analysis (PCA), which produced two new variables (principal components) with eigenvalues or standard deviations larger than one. The first principal component (PC1) eigenvalue is 1.947, explaining 47.4% of the variance. The second principal component (PC2) has an eigenvalue of 1.207, explaining 65.6% of the variation. Meanwhile, because their eigenvalues or standard deviations are less than one, the third and subsequent principal components, up to and including the eighth, do not produce new variables. For these two reasons, PC1 and PC2 were chosen to replace the variables determining poverty in East Java. After collecting the variables PC1 and PC2, regression on the poverty response variable is performed, yielding the RPCA model as follows.

$$\hat{y} = 11.87 - 2.01PC_1 - 0.93PC_2 \qquad (1)$$

The choice to reject H_0 is made in the RPCA significance test with a significance level of 10% since the calculated F-value = 46.93 > $F_{(0,1;2;35)}$ = 2.461, and p-value = 0.000 < 0.1. As a result, it is inferred that the poverty response variable and PC1 or PC2 together have at least some associations. Meanwhile, the decision to reject H_0 is taken in the RPCA individual coefficient test. As illustrated in Table 3 below.

Table 3. Results of Simultaneous Significance Test of RPCA

Variable	Coefficient	t-value	P-value
Intercept	11.8747	28.256	0.000**
PC_1	-2.0098	-9.310	0.000**
PC_2	-0.9321	-2.677	0.011**

Note: ** indicates significant regression coefficients that are not equal to zero at p-value < 0.1.

This decision is based on the fact that the |t-value| for PC1 and PC2 is greater than $t_{(0,05;38)}$= 2.024, and p-value = 0.000 < 0.05. Thus, it is concluded that there is an influence between the poverty response variable and both PC1 and PC2 individually.

4.3. Model of Geographically Weighted Regression Principal Component Analysis (GWRPCA)

The Geographically Weighted Regression Principal Component Analysis model is a local model built for data with impacts of spatial heterogeneity. Following the collection of PC1 and PC2, the Breusch-Pagan (BP) test is used to determine changes in features between regions. The BP value is 5.279 based on the Breusch-Pagan test results, whereas the crucial Chi-square value at a significance level of 10% is 4.605, showing the presence of spatial heterogeneity. The GWRPCA approach is used to handle spatial heterogeneity issues in the RPCA model.

Weighting is used in the GWRPCA model based on the geographical location of each regency or city. The first step is to determine each regency or city's geographical position (longitude and latitude) in East Java Province. Following that, the Euclidean distance is determined for each regency or city in East Java Province based on geographical locations. Based on Euclidean distance, Region I can be allocated an order of other surrounding regions, resulting in a succession of nearest neighbors for all observation areas. The optimal bandwidth for each regency or city is then determined using a kernel function. To select the optimum kernel approach, a model is developed for each weight to acquire the weight's cross-validation (CV) value. The optimal weight for building the model is the weight with the lowest CV value. Table 4 displays the CV values for each weight. The optimal weight is chosen based on the CV values for each weight, as shown in Table 4.

Table 4. Bandwidth Selection Results

Kernel Weight	CV
Adaptive Gaussian	283.5714**
Adaptive Bi-square	287.689
Fixed Gaussian	287.5021
Fixed Bi-square	289.5331

Note: ** represents the optimum bandwidth value as it has the minimum CV.

Smaller CV values characterize optimal weights. Table 4 shows that the Adaptive Gaussian bandwidth is the best weight for GWRPCA modeling since it has a lower CV value than other weights. After determining the optimal bandwidth, parameter estimation for each region is obtained. Table 5 shows the minimum and maximum GWRPCA model estimations. The following is a statistical overview of parameter estimator results for the GWRPCA model.

Table 5. GWRPCA Modeling Results

Parameter	Minimum	Maximum
β0	10.416	12.308
PC1	-3.075	-1.336
PC2	-2.395	0.6705

Note: ** represents the optimum bandwidth value as it has the minimum CV.

Table 5 shows that for all regencies and cities in East Java, the parameter estimations for each PC1 variable exhibit negative regression coefficient values, whereas the estimated values for the PC2 variable range between -2.39518 and 0.67050. At the 10% level of significance, the decision fails to reject H_0 because F1 = 1.0122 is less than $F_{(0,1;38;19.3)}$ = 1.734. As a result, there is no discernible difference between the GWRPCA model and the RPCA model.

4.4. Model Performance Evaluation

The performance of various regression models, notably RPCA (Regression Principal Component Analysis) and GWRPCA (Geographically Weighted Regression Principal Component Analysis), was comprehensively examined throughout the model evaluation phase. The R2 value, which shows the quality of fit for each model, was the major criterion for measuring model performance.

Table 6. Model Performance Evaluation Result

Model	R^2
RPCA	72.84%
GWRPCA	85.10%**

Note: ** indicates the best model with the highest R^2 value.

Table 6 illustrates the R^2 values for both the RPCA and GWRPCA models. Overall, the R^2 value generated by the GWRPCA model is higher than that of the RPCA model. This indicates that the GWRPCA model is more effective in modeling poverty in the East Java Province. The comparison indicates that the GWRPCA model significantly enhances the understanding of the factors influencing poverty in East Java. The higher R^2 value suggests that incorporating geographic variations and spatial relationships through the GWRPCA approach leads to a more accurate and localized poverty model. Therefore, the GWRPCA model is considered the more effective and robust choice for modeling poverty in the East Java Province.

4.5. Poverty Analysis with GWRPCA Model in East Java Province

The utilization of the GWRPCA model proves to be more effective in modeling poverty in East Java Province compared to the RPCA model. This superiority arises from the spatial diversity inherent in nonstationary spatial data, where variations in poverty dynamics differ across observation areas. Tailoring poverty alleviation programs to each region based on its dominant issues aims to optimize poverty alleviation efforts in East Java. Therefore, a partial test is conducted for each parameter in every observation area. The partial test employs the t-statistic for each parameter within the GWRPCA model. The following is a visualization of poverty analysis using the GWRPCA model.

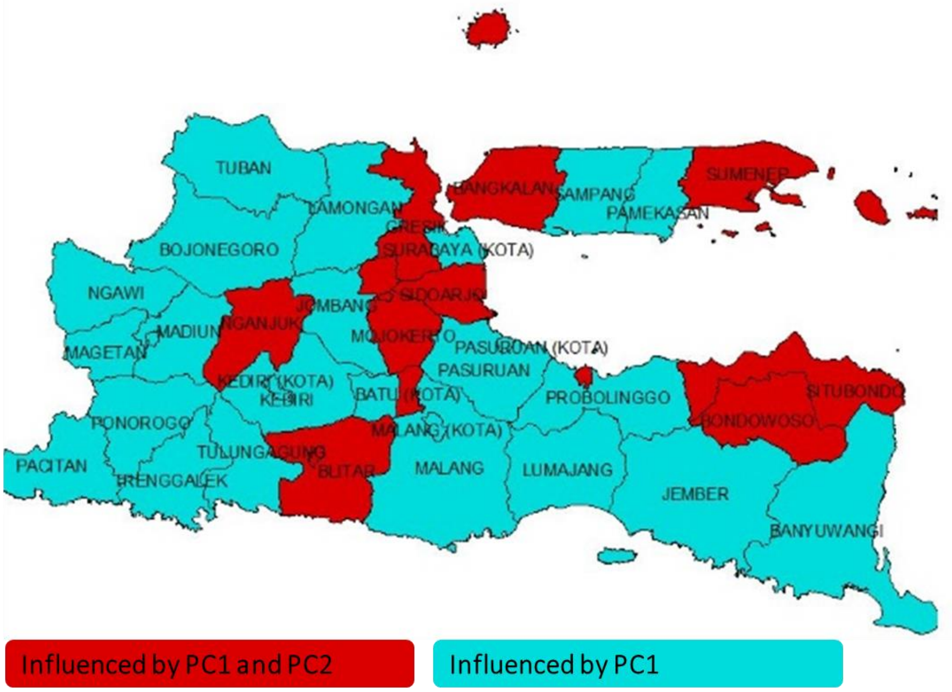

Figure 3. Visualization of Regions Based on the Significance of Principal Components

Based on Figure 3, East Java's regencies/cities are categorized into two regions: those influenced by PC1 and those influenced by both PC1 and PC2. There are 14 regions influenced by PC1 and PC2, including Surabaya, Probolinggo, Batu, Blitar, Nganjuk, Blitar Regency, Tulungagung, Mojokerto, Gresik, Sidoarjo, Bondowoso, Situbondo, Bangkalan, and Sumenep. This spatial differentiation underscores the need for targeted and nuanced poverty alleviation strategies tailored to the specific challenges faced by each region.

Table 7. GWRPCA Model Coefficient Values

Regency/City	βPC_1	βPC_2
Pacitan Regency	-2.24	-0.52
Ponorogo Regency	-2.46	-0.3
Trenggalek Regency	-2.33	-1.08
Tulungagung Regency	-2.85	-1.42
Lumajang Regency	-2.2	-0.16
Bondowoso Regency	-1.57	-1.08
Pasuruan Regency	-2.38	-0.22
Jombang Regency	-1.91	0.44
Nganjuk Regency	-1.43	-0.89
Madiun Regency	-2.34	-0.37
Magetan Regency	-1.37	-0.68
Ngawi Regency	-2.46	-0.6
Bojonegoro Regency	-1.34	-0.24
Tuban Regency	-2.75	-1.16
Lamongan Regency	-1.53	0.02
Bangkalan Regency	-2.42	-2.4
Pamekasan Regency	-2.98	-0.76
Pacitan Regency	-2.28	-0.21
Ponorogo Regency	-1.82	-1.18
Trenggalek Regency	-2.41	-0.26
Tulungagung Regency	-1.48	-1.01
Lumajang Regency	-2.42	-0.29
Bondowoso Regency	-1.46	-0.9
Pasuruan Regency	-1.96	-0.57
Jombang Regency	-1.39	-0.78
Nganjuk Regency	-1.51	-0.92
Madiun Regency	-2.52	-2.04
Magetan Regency	-1.89	0.67
Ngawi Regency	-1.44	-0.93
Bojonegoro Regency	-2.27	-0.57
Tuban Regency	-1.44	-0.89
Lamongan Regency	-1.36	-0.31
Bangkalan Regency	-1.53	-0.73
Pamekasan Regency	-2.35	-0.53
Pacitan Regency	-3.07	-2.05
Ponorogo Regency	-1.48	-1.11
Trenggalek Regency	-1.42	-0.91
Tulungagung Regency	-1.56	-1.11

The parameter Table 7 above displays the GWRPCA model's estimation results for each regency/city in East Java. The GWRPCA model is intended to comprehend the influence of factors influencing poverty at the local level while taking spatial variety into account. In other words, each region has its model that accounts for the differences in traits and causes that influence the poverty rate in that region. The GWRPCA model is significant because it can capture spatial differences in the connection between predictor factors (PC1 and PC2) and the response variable (poverty rate). We can see how the influence of these elements differs from region to region using this modeling technique. As a result, the GWRPCA model enables us to comprehend the dynamics of poverty at the local level in greater depth and context. The variation in parameter values (PC1 and PC2) among regions suggests that the impact of PC1 and PC2 on the poverty rate may change across regions and cities. Some locations may be more influenced by one major component than others, and the estimated parameter values reflect this. As a result, by taking into account the spatial dimension, the GWRPCA model provides a deeper understanding of the underlying determinants of poverty in East Java.

5. CONCLUSIONS

Through the course of this research, an in-depth analysis has been conducted regarding poverty in East Java Province using two main models, namely Regression Principal Component Analysis (RPCA) and Geographically Weighted Regression Principal Component Analysis (GWRPCA). Both models were applied to detail the factors influencing poverty rates at the regency/city level. In this conclusion, we will present key findings and identify the strengths, limitations, and potential further developments of this study.

Research Results and Model Comparison:

a. GWRPCA provides a more detailed representation with a higher R2 value (85.10%) compared to RPCA (72.84%).
b. Spatial modeling of GWRPCA allows for a deeper understanding of poverty factors in each regency/city in East Java.
c. Ability to capture spatial diversity and illustrate differences in characteristics between regions in more detail.
d. Relevant for more contextual and targeted policy decision-making.

Limitations and Constraints:

a. GWRPCA requires complete spatial data for more accurate results.
b. The complexity of the analysis may demand significant computational resources.

Potential Future Developments:

a. Integration of data from subsequent years to monitor spatial changes and poverty dynamics.
b. Focus on the selection of kernel methods and additional variables for further development.

REFERENCES

[1] A. Nugroho, H. Amir, I. Maududy, and I. Marlina, "Poverty eradication programs in Indonesia: Progress, challenges and reforms," *J Policy Model*, vol. 43, no. 6, pp. 1204–1224, Nov. 2021, doi: 10.1016/J.JPOLMOD.2021.05.002.

[2] I. Nuraini, R. Hidayat, and S. W. Sulistyono, "INTERACTION DETERMINING FACTORS FROM REGENCY / CITY REGION IN EAST JAVA," *Jurnal Ekonomi Pembangunan*, vol. 17, no. 2, pp. 152–163, Jan. 2019, doi: 10.22219/JEP.V17I2.10998.

[3] S. R. Putri, A. W. Wijayanto, and A. D. Sakti, "Developing Relative Spatial Poverty Index Using Integrated Remote Sensing and Geospatial Big Data Approach: A Case Study of East Java, Indonesia," *ISPRS International Journal of Geo-Information 2022, Vol. 11, Page 275*, vol. 11, no. 5, p. 275, Apr. 2022, doi: 10.3390/IJGI11050275.

[4] L. Anselin, "Spatial Models in Econometric Research," *Oxford Research Encyclopedia of Economics and Finance*, Aug. 2021, doi: 10.1093/ACREFORE/9780190625979.013.643.

[5] Y. Song, J. Wang, Y. Ge, and C. Xu, "An optimal parameters-based geographical detector model enhances geographic characteristics of explanatory variables for spatial heterogeneity analysis: cases with different types of spatial data," *GIsci Remote Sens*, vol. 57, no. 5, pp. 593–610, Jul. 2020, doi: 10.1080/15481603.2020.1760434.

[6] W. Yang, M. Deng, J. Tang, and L. Luo, "Geographically weighted regression with the integration of machine learning for spatial prediction," *J Geogr Syst*, vol. 25, no. 2, pp. 213–236, Apr. 2023, doi: 10.1007/S10109-022-00387-5/FIGURES/12.

[7] A. D. Pratama, I. W. Suparta, and U. Ciptawaty, "Spatial Autoregressive Model and Spatial Patterns of Poverty In Lampung Province," *Eko-Regional: Jurnal Pembangunan Ekonomi Wilayah*, vol. 16, no. 1, pp. 14–28, Mar. 2021, doi: 10.20884/1.ERJPE.2021.16.1.1776.

[8] B. Tiyas Belantika *et al.*, "Factors affecting poverty using a geographically weighted regression approach (case study of Java Island, 2020)," *Optimum: Jurnal Ekonomi dan Pembangunan*, vol. 13, no. 2, pp. 141–154, Nov. 2023, doi: 10.12928/OPTIMUM.V13I2.7993.

[9] A. Shrestha and W. Luo, "Analysis of Groundwater Nitrate Contamination in the Central Valley: Comparison of the Geodetector Method, Principal Component Analysis and Geographically Weighted Regression," *ISPRS International Journal of Geo-Information 2017, Vol. 6, Page 297*, vol. 6, no. 10, p. 297, Sep. 2017, doi: 10.3390/IJGI6100297.

[10] P. Das, "Linear Regression Model: Relaxing the Classical Assumptions," *Econometrics in Theory and Practice*, pp. 109–135, 2019, doi: 10.1007/978-981-32-9019-8_4.

[11] D. H. Maulud and A. Mohsin Abdulazeez, "A Review on Linear Regression Comprehensive in Machine Learning," *Journal of Applied Science and Technology Trends*, vol. 1, no. 4, pp. 140–147, Dec. 2020, doi: 10.38094/jastt1457.

[12] M. Gregorich, S. Strohmaier, D. Dunkler, and G. Heinze, "Regression with highly correlated predictors: Variable omission is not the solution," *Int J Environ Res Public Health*, vol. 18, no. 8, p. 4259, Apr. 2021, doi: 10.3390/IJERPH18084259/S1.

[13] D. C. Wheeler, "Geographically Weighted Regression," *Handbook of Regional Science: Second and Extended Edition: With 238 Figures and 78 Tables*, pp. 1895–1921, Jan. 2021, doi: 10.1007/978-3-662-60723-7_77/COVER.

[14] R. Cellmer, A. Cichulska, and M. Bełej, "Spatial Analysis of Housing Prices and Market Activity with the Geographically Weighted Regression," *ISPRS International Journal of Geo-Information 2020, Vol. 9, Page 380*, vol. 9, no. 6, p. 380, Jun. 2020, doi: 10.3390/IJGI9060380.

[15] O. Raza, M. A. Mansournia, A. Rahimi Foroushani, and K. Holakouie-Naieni, "Geographically weighted regression analysis: A statistical method to account for spatial heterogeneity," *Arch Iran Med*, vol. 22, no. 3, pp. 155–160, Mar. 2019.

[16] S. Sisman and A. C. Aydinoglu, "A modelling approach with geographically weighted regression methods for determining geographic variation and influencing factors in housing price: A case in Istanbul," *Land use policy*, vol. 119, p. 106183, Aug. 2022, doi: 10.1016/J.LANDUSEPOL.2022.106183.

[17] G. T. Reddy *et al.*, "Analysis of Dimensionality Reduction Techniques on Big Data," *IEEE Access*, vol. 8, pp. 54776–54788, 2020, doi: 10.1109/ACCESS.2020.2980942.

[18] N. B. Erichson, P. Zheng, K. Manohar, S. L. Brunton, J. N. Kutz, and A. Y. Aravkin, "Sparse Principal Component Analysis via Variable Projection," *https://doi.org/10.1137/18M1211350*, vol. 80, no. 2, pp. 977–1002, Apr. 2020, doi: 10.1137/18M1211350.

[19] J. Y. Le Chan *et al.*, "Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review," *Mathematics 2022, Vol. 10, Page 1283*, vol. 10, no. 8, p. 1283, Apr. 2022, doi: 10.3390/MATH10081283.

[20] R. Zhao, L. Zhan, M. Yao, and L. Yang, "A geographically weighted regression model augmented by Geodetector analysis and principal component analysis for the spatial distribution of PM2.5," *Sustain Cities Soc*, vol. 56, p. 102106, May 2020, doi: 10.1016/J.SCS.2020.102106.

[21] E. May Mujiarti, A. Vickylia Roos, F. Fran, U. H. Tanjungpura Jl Hadari Nawawi, K. Pontianak, and P. Korespondensi, "Application of the PCA Method to Identify Factors Affecting Poverty Rates in West Kalimantan," *Jurnal Forum Analisis Statistik (FORMASI)*, vol. 3, no. 1, pp. 11–22, Jul. 2023, doi: 10.xxxxx/formasi.2021.1.1.1-12.

[22] D. O. Mahara and A. Fauzan, "Impacts of Human Development Index and Percentage of Total Population on Poverty using OLS and GWR models in Central Java, Indonesia," *EKSAKTA: Journal of Sciences and Data Analysis*, vol. 2, no. 2, pp. 142–154, Sep. 2021, doi: 10.20885/EKSAKTA.VOL2.ISS2.ART8.

[23] X. He, X. Mai, and G. Shen, "Poverty and Physical Geographic Factors: An Empirical Analysis of Sichuan Province Using the GWR Model," *Sustainability 2021, Vol. 13, Page 100*, vol. 13, no. 1, p. 100, Dec. 2020, doi: 10.3390/SU13010100.

[24] A. Rinaldi *et al.*, "Spatial Modeling for Poverty: The Comparison of Spatial Error Model and Geographic Weighted Regression," *Al-Jabar : Jurnal Pendidikan Matematika*, vol. 12, no. 1, pp. 237–251, Jun. 2021, doi: 10.24042/AJPM.V12I1.8671.

[25] L. Huang, J. Qin, Y. Zhou, F. Zhu, L. Liu, and L. Shao, "Normalization Techniques in Training DNNs: Methodology, Analysis and Application," *IEEE Trans Pattern Anal Mach Intell*, vol. 45, no. 8, pp. 10173–10196, Aug. 2023, doi: 10.1109/TPAMI.2023.3250241.

[26] J. I. Daoud, "Multicollinearity and Regression Analysis," *J Phys Conf Ser*, vol. 949, no. 1, p. 012009, Dec. 2017, doi: 10.1088/1742-6596/949/1/012009.

[27] N. Shrestha, "Detecting Multicollinearity in Regression Analysis," *Am J Appl Math Stat*, vol. 8, no. 2, pp. 39–42, 2020, doi: 10.12691/ajams-8-2-1.

[28] N. Martin, "Robust and efficient Breusch-Pagan test-statistic: an application of the beta-score Lagrange multipliers test for non-identically distributed individuals," Jan. 2023, Accessed: Nov. 28, 2023. [Online]. Available: https://arxiv.org/abs/2301.07245v1

[29] J. Karch, "Improving on adjusted R-squared," *Collabra Psychol*, vol. 6, no. 1, Sep. 2020, doi: 10.1525/COLLABRA.343/114458.

[30] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput Sci*, vol. 7, pp. 1–24, Jul. 2021, doi: 10.7717/PEERJ-CS.623/SUPP-1.

[31] J. D. Rights and S. K. Sterba, "Quantifying explained variance in multilevel models: An integrative framework for defining R-squared measures," *Psychol Methods*, vol. 24, no. 3, pp. 309–338, Jun. 2019, doi: 10.1037/MET0000184.