

## PEMODELAN DATA GEOSPASIAL BALITA KURANG GIZI DENGAN PENDEKATAN *GEOGRAPHICALLY WEIGHTED REGRESSION PRINCIPAL COMPONENT ANALYSIS*

Cinta Rizki Oktarina<sup>1</sup>, Idhia Sriliana<sup>2\*</sup>, Esa Nur Fadhillah Sidik<sup>3</sup>, Muhammad Akbar Firmansyah<sup>4</sup>

<sup>1,2,3,4</sup> Program Studi S2 Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Bengkulu

\*e-mail: [idhiasriliana@unib.ac.id](mailto:idhiasriliana@unib.ac.id)

DOI: 10.14710/j.gauss.13.2.339-350

### Article Info:

Received: 2024-06-06

Accepted: 2024-11-14

Available Online: 2024-11-15

### Keywords:

Malnutrition; Principal Components; Spatial Heterogeneity; Multicollinearity

**Abstract:** This study examines the factors that influence under-five malnutrition in Indonesia using the Geographically Weighted Regression Principal Component Analysis (GWRPCA) method. The GWRPCA approach was chosen to overcome the problems of multicollinearity and spatial heterogeneity that often occur in geospatial data analysis. This study uses 10 factors that are thought to influence under-five malnutrition used as the dependent variable. Malnutrition in under-fives increases the risk of disease, death, and stunting, with data showing a caseload of 10.2% or 805,000 under-fives experiencing malnutrition. GWRPCA analysis successfully reduced data dimensionality and spatial heterogeneity by selecting four principal components that explained 81.4% of the total data variance, encompassing important information from the independent variables associated with under-five malnutrition. By using these principal components, the study was able to more efficiently identify the main determinants of undernutrition among children under five.

## 1. PENDAHULUAN

Analisis regresi biasanya digunakan untuk melihat pengaruh variabel independen terhadap variabel respons. Namun, dalam kasus data geospasial, seperti kekurangan gizi pada balita di Indonesia, terdapat tantangan berupa multikolinieritas dan heterogenitas spasial. Multikolinieritas, yang merupakan adanya korelasi kuat antar variabel independen, serta heterogenitas spasial yang menunjukkan perbedaan karakteristik antar wilayah, membuat model regresi tradisional kurang tepat untuk diterapkan. Untuk mengatasi tantangan tersebut, pendekatan *Geographically Weighted Regression Principal Component Analysis* (GWRPCA) dianggap lebih sesuai. Pendekatan ini mampu menangani isu multikolinieritas dan heterogenitas spasial dengan menggabungkan *Geographically Weighted Regression* (GWR) dan *Principal Component Analysis* (PCA). GWRPCA tidak hanya mereduksi dimensi data, tetapi juga secara efektif menangkap variasi lokal dalam data.

Urgensi penelitian ini ada pada tingginya prevalensi balita kurang gizi yang tercatat mencapai 10,2% atau sekitar 805.000 anak, yang memiliki risiko terhadap masalah kesehatan serius seperti stunting dan kematian. Faktor-faktor penyebab kekurangan gizi tidak hanya terkait dengan aspek nutrisi, tetapi juga melibatkan kondisi kesehatan, lingkungan, dan ekonomi yang bervariasi secara signifikan antar provinsi. Data ini menunjukkan bahwa penanganan yang seragam tidak dapat memenuhi kebutuhan spesifik setiap wilayah. Oleh karena itu, diperlukan pemahaman yang lebih dalam mengenai faktor-faktor lokal yang berpengaruh, yang dapat diidentifikasi melalui pendekatan GWRPCA.

Keunggulan GWRPCA terletak pada kemampuannya untuk mengatasi multikolinieritas dan menangkap keragaman lokal di setiap wilayah. Dengan mereduksi data menjadi empat komponen utama yang mencakup lebih dari 80% variabilitas, metode ini memungkinkan analisis yang efisien sambil tetap mempertahankan informasi penting. Melalui identifikasi faktor-faktor penentu kekurangan gizi di setiap daerah, GWRPCA memberikan dasar bagi intervensi kebijakan yang lebih efektif dan spesifik sesuai dengan kebutuhan masing-masing provinsi. Pendekatan ini menghasilkan wawasan yang mendalam untuk upaya peningkatan kesehatan gizi balita secara komprehensif dan berkelanjutan.

## 2. TINJAUAN PUSTAKA

PCA adalah suatu metode yang menjelaskan struktur varian dan kovarian dari sekelompok variabel dengan lebih sedikit kombinasi linier dari variabel tersebut, dengan tujuan mereduksi dan menafsirkan data (Johnson & Wichern, 2007). Misalnya vector random  $\mathbf{X}' = [X_1, X_2, \dots, X_n]$  memiliki suatu kovarian matrik  $\Sigma$  dengan nilai eigen  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$ , dapat didefinisikan kombinasi liniernya dengan bentuk:

$$\begin{aligned} PC_1 &= \mathbf{e}'_1 \mathbf{Z} = e_{11}Z_1 + e_{12}Z_2 + \dots + e_{1n}Z_n \\ PC_2 &= \mathbf{e}'_2 \mathbf{Z} = e_{21}Z_1 + e_{22}Z_2 + \dots + e_{2n}Z_n \\ &\vdots \\ PC_n &= \mathbf{e}'_n \mathbf{Z} = e_{n1}Z_1 + e_{n2}Z_2 + \dots + e_{np}Z_n \end{aligned} \quad (1)$$

Dengan Varian  $(PC_j) = \mathbf{a}'_j \Sigma \mathbf{a}_j$  dimana  $j = 1, 2, \dots, n$

Sehingga komponen varian total didefinisikan sebagai berikut:

$$\sum_{j=1}^k Var(PC_j) = tr(\Sigma) = tr(\Lambda) = \sigma_{11} + \sigma_{22} + \dots + \sigma_{nn} = \lambda_1 + \lambda_2 + \dots + \lambda_k \quad (2)$$

Dengan nilai eigen  $\lambda_j$  yang merupakan komponen utama ke- $j$ ,  $j = 1, 2, \dots, n$

Komponen utama dapat ditentukan dengan jumlahnya dengan mengamati persentase kumulatif varian yang mampu menerangkan total varian data 70% sampai 80%, dapat juga dengan melihat pola penurunan pada *scree plot* (Simamora, 2005). *Scree plot* adalah alat visual dalam PCA yang digunakan untuk menentukan jumlah komponen utama yang signifikan. Grafik ini menunjukkan proporsi variansi yang dijelaskan oleh setiap komponen, dengan fokus pada titik siku ("*elbow point*") sebagai penanda komponen signifikan. Komponen sebelum titik siku biasanya dianggap penting karena menjelaskan sebagian besar variansi dalam data. Asumsi utama PCA meliputi:

- Uji Kaiser-Meyer-Olkin (KMO) digunakan untuk mengukur kecukupan sampel dalam analisis faktor. Nilai KMO berkisar antara 0 dan 1, di mana nilai yang lebih tinggi menunjukkan bahwa pola korelasi dalam data cukup untuk analisis PCA. Nilai KMO di atas 0,6 dianggap memadai untuk analisis lebih lanjut (Nunes et al., 2020). Penggunaan uji KMO diperlukan ketika akan melakukan analisis faktor atau PCA untuk memastikan bahwa sampel cukup dan variabel memiliki korelasi antar-komponen. Namun, uji ini tidak diperlukan jika analisis tidak mengharuskan adanya asumsi korelasi antar-variabel.
- Uji Bartlett digunakan untuk membuktikan bahwa matriks korelasi bukan merupakan matriks identitas, yang menunjukkan bahwa terdapat hubungan antara variabel dalam dataset. Hasil yang signifikan menunjukkan bahwa matriks korelasi tidak identik dan bahwa analisis faktor dapat dilakukan dengan data tersebut (Yıldırım et al., 2021).
- Uji Kolmogorov-Smirnov adalah uji non-parametrik yang digunakan untuk membandingkan distribusi sampel dengan distribusi teoretis atau untuk membandingkan dua distribusi sampel. Uji ini bertujuan untuk mengevaluasi apakah dua set data berasal dari

distribusi yang sama (Rajah et al., 2021). Uji Kolmogorov-Smirnov menilai perbedaan antara distribusi kumulatif empiris dari sampel data dan distribusi kumulatif teoretis yang diharapkan, yaitu distribusi normal. Nilai statistik ujinya adalah jarak maksimum antara dua distribusi kumulatif ini. Hipotesis nol yang menyatakan bahwa data berdistribusi normal ditolak jika nilai statistik uji lebih besar dari nilai kritisnya. Penelitian menunjukkan bahwa uji Kolmogorov-Smirnov memiliki beberapa keunggulan dibandingkan dengan metode uji normalitas lainnya, seperti uji Shapiro-Wilk dan Anderson-Darling, terutama untuk ukuran sampel yang besar (Emmanuel et al., 2020).

GWR merupakan model regresi lanjutan yang menghitung setiap parameter pada titik observasi yang ada, agar nilai parameter regresi dalam setiap titik observasi berbeda. (Fotheringham et al., 2002). Model GWR sebagai berikut:

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^k \beta_j(u_i, v_i)x_j + \varepsilon_i \quad (3)$$

GWR mengasumsikan bahwa  $\varepsilon \sim N(0, \sigma^2)$  dengan 0 merupakan nilai ekspektasi dari *error* dan  $\sigma^2$  yang merujuk pada varians dari *error* model.

Parameter  $\beta$  dilakukan penaksiran dengan metode *Weighted Least Square* (WLS) sehingga diperoleh:

$$\hat{\beta}(u_i, v_i) = [\mathbf{X}'\mathbf{W}(u_i, v_i)\mathbf{X}]^{-1}\mathbf{X}'\mathbf{W}\mathbf{y} \quad (4)$$

Dengan bentuk dari vektor dan matriksnya sebagai berikut:

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}, \hat{\beta}(u_i, v_i) = \begin{pmatrix} \beta_0(u_i, v_i) \\ \beta_1(u_i, v_i) \\ \vdots \\ \beta_p(u_i, v_i) \end{pmatrix} \quad (5)$$

$$\mathbf{W}(u_i, v_i) = \begin{pmatrix} w_{i1} & 0 & \cdots & 0 \\ 0 & w_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{in} \end{pmatrix} \quad (6)$$

fungsi pembobot digunakan untuk memberikan setiap lokasi penaksiran parameter yang berbeda, terdapat 2 jenis fungsi pembobot:

a. *Fixed Kernel*.

Setiap titik lokasi pengamatan memiliki *bandwidth* yang sama.

1. *Kernel Gaussian*

$$w_i = e \left[ -\frac{1}{2} \left( \frac{d_i}{b} \right)^2 \right] \quad (7)$$

2. *Kernel Bi-square*

$$w_i = \begin{cases} \left[ 1 - \left( \frac{d_i}{b} \right)^2 \right]^2, & \text{jika } d_i < b \\ 0, & \text{lainnya} \end{cases} \quad (8)$$

b. *Adaptive Kernel*.

Setiap titik lokasi pengamatan memiliki *bandwidth* yang berbeda.

1. *Adaptive kernel gaussian*

$$w_i = e \left[ -\frac{1}{2} \left( \frac{d_i}{b_{i(q)}} \right)^2 \right] \quad (9)$$

2. *Adaptive kernel Bi-square*

$$w_i = \begin{cases} \left[1 - \left(\frac{d_i}{b_{i(q)}}\right)^2\right]^2, & \text{jika } d_i < b \\ 0, & \text{lainnya} \end{cases} \quad (10)$$

$b_{i(q)}$  merupakan *bandwidth adaptive* yang menetapkan 1 sebagai jarak ke tetangga terdekat dari titik lokasi pengamatan ke- $i$ , dan  $d_i$  merupakan jarak *Euclidean* antara titik lokasi pengamatan ke- $i$  dan titik lokasi pengamatan lainnya (Fotheringham et al., 2002). Teknis penentuan *bandwidth* adaptif dilakukan dengan menentukan jarak ke beberapa tetangga terdekat untuk setiap titik pengamatan, di mana jarak ini akan bervariasi sesuai kepadatan data lokal, sehingga *bandwidth* menyesuaikan jumlah titik pengamatan di sekitarnya dan mempertahankan jumlah tetangga tetap pada setiap titik.

*Cross Validation* (CV) digunakan untuk menentukan bandwidth optimum, yaitu:

$$CV = \sum_{i=1}^n [y_i - \hat{y}_{\neq i}(b)]^2 \quad (11)$$

Kecocokan suatu model dapat diuji dengan menggunakan hipotesis berikut:

$H_0: \beta_k(u_i, v_i) = \beta_k$ , dimana setiap  $k = 0, 1, 2, \dots, p$  dan  $i = 1, 2, \dots, n$

$H_1$ : Terdapat satu atau lebih  $\beta_k(u_i, v_i) \neq \beta_k$ ,  $k = 0, 1, 2, \dots, p$

Penentuan statistik uji berdasarkan *Residual Sum of Square* (RSS). Statistik uji:

$$F_1 = \frac{RSS(H_1)/df_1}{RSS(H_0)/df_2} \quad (12)$$

Keterangan:

$$RSS(H_1) = \mathbf{y}'(\mathbf{I} - \mathbf{S})'(\mathbf{I} - \mathbf{S})\mathbf{y}$$

$$RSS(H_0) = \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}$$

$$\delta_1 = \text{tr}((\mathbf{I} - \mathbf{S})'(\mathbf{I} - \mathbf{S}))$$

$$\delta_2 = \text{tr}((\mathbf{I} - \mathbf{S})'(\mathbf{I} - \mathbf{S}))^2$$

$$df_1 = n - p - 1 - \delta_1$$

$$df_2 = n - p - 1 - 2\delta_1 + \delta_2$$

Jika tingkat signifikansi sama dengan  $\alpha$  maka tolak  $H_0$  jika  $F_1 < F_{1-\alpha; (df_1, df_2)}$ .

Jika terjadi multikolinieritas pada kasus GWR, variabel independen yang saling berkorelasi dapat disederhanakan dengan metode PCA. Multikolinieritas merupakan kondisi ketika variabel independen saling berkorelasi, sehingga nilai residu dari estimasi parameter yang dihasilkan besar (Myers, 1990). Multikolinieritas local diketahui dengan menghitung nilai VIF > 10 sebagai berikut:

$$VIF_j(u_i, v_i) = \frac{1}{1 - R_j^2(u_i, v_i)} \quad (13)$$

dengan  $R_j^2(u_i, v_i)$  adalah koefisien determinan setiap variabel bebas  $X_j$  dengan variabel bebas lain dalam setiap lokasi pengamatan  $(u_i, v_i)$ .

Menurut (Fotheringham et al., 2002), GWRPCA dapat didefinisikan dengan:

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^k \beta_j(u_i, v_i) PC_{ij} + \varepsilon_i; \quad i = 1, 2, \dots, n \quad (14)$$

Keterangan:

$y_i$  : nilai pengamatan variabel respon ke- $i$

$PC_{ij}$  : nilai pengamatan komponen utama ke- $j$  pada lokasi pengamatan ke- $i$

$\beta_0(u_i, v_i)$  : konstanta atau intercept dalam pengamatan ke- $i$

$(u_i, v_i)$  : titik koordinat (*longitude, latitude*) lokasi ke- $i$

$\beta_j(u_i, v_i)$  : koefisien regresi untuk variabel bebas ke- $j$  di lokasi pengamatan ke- $i$

$\varepsilon_i$  : error pengamatan ke- $i$

Dalam GWRPCA, proses PCA dilakukan secara local pada setiap lokasi geografis untuk menangkap heterogenitas spasial (Chen dkk, 2020).

Asumsi-asumsi GWRPCA:

- a. Non-Stasionaritas Spasial  
Asumsi bahwa hubungan antara variabel dependen dan independen bervariasi di seluruh ruang geografis. Koefisien regresi  $\beta_j(u_i, v_i)$  diperkirakan berubah di setiap lokasi (Chen et al., 2021).
- b. Multikolinearitas lokal  
PCA digunakan untuk mengatasi multikolinearitas antara variabel independen, memungkinkan model GWR untuk bekerja lebih efektif. Matriks kovarians  $PC$  digunakan untuk menghitung *eigenvectors* dan mengubah data menjadi komponen utama yang tidak berkorelasi (Li et al., 2020).
- c. Heterogenitas Spasial  
GWRPCA mengasumsikan bahwa ada heterogenitas spasial dalam komponen utama yang harus dimodelkan secara eksplisit untuk mendapatkan estimasi yang akurat. Model regresi spasial berbobot digunakan untuk memperhitungkan variasi lokal (Han et al., 2022).

### 3. METODOLOGI

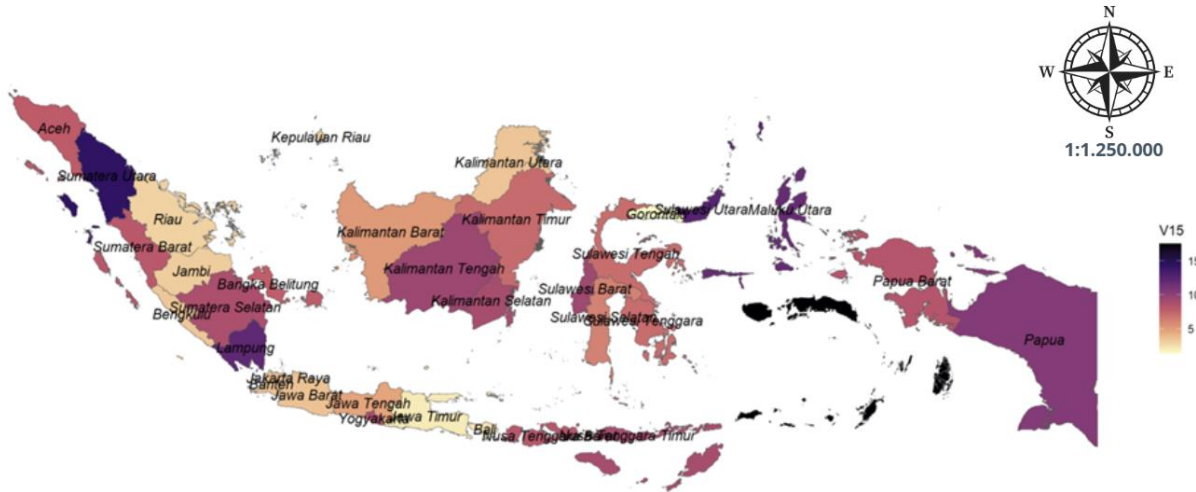
Pemilihan variabel independen dalam penelitian ini didasarkan pada tinjauan literatur yang mengidentifikasi faktor utama penyebab kurang gizi pada balita. Jumlah layanan kesehatan ( $X_1$ ) dan layanan kesehatan dasar pada balita ( $X_2$ ) dianggap penting karena ketersediaan fasilitas kesehatan berperan dalam pemenuhan gizi balita (Nursiyono & Apriyani, 2023). Selain itu, akses sanitasi dasar ( $X_3$ ) dan fasilitas kesehatan dasar ( $X_4$ ) membantu mencegah penyakit yang memperburuk gizi balita (Basyariyah, Diyanah, & Pawiyra, 2022). Dukungan sosial melalui realisasi KPM bantuan pangan ( $X_5$ ) dan anggaran bantuan pangan ( $X_6$ ) mengurangi risiko kurang gizi pada kelompok rentan. Variabel sosial ekonomi lain seperti kepemilikan rumah kontrak ( $X_7$ ), buta aksara usia 45+ ( $X_8$ ), tingkat pengangguran terbuka ( $X_9$ ), dan persentase penduduk miskin ( $X_{10}$ ) juga dipilih karena memengaruhi akses pangan bergizi dan layanan kesehatan (Aini & Margareta, 2023). Variabel-variabel ini diharapkan memberikan pemahaman komprehensif tentang faktor-faktor yang memengaruhi gizi balita. Adapun tahapan analisisnya adalah sebagai berikut:

1. Analisis dengan Metode PCA
  - Uji asumsi PCA: uji KMO, uji Bartlett, dan uji Kolmogrov-Smirnov
  - Standarisasi data
  - Pembentukan komponen utama dengan menggunakan Persamaan (1)
  - Pemilihan komponen utama berdasarkan  $PKV \geq 80\%$  dengan menggunakan Persamaan (2)
2. Pengujian heterogenitas spasial komponen utama terpilih
3. Pengujian Multikolinieritas Lokal dengan menggunakan Persamaan (13)
4. Analisis dengan Metode GWRPCA
  - Menghitung jarak untuk metode GWRPCA
  - Menentukan *Bandwidth* optimum GWRPCA dengan menggunakan Persamaan (11)
  - Menentukan matriks pembobot spasial dengan fungsi *fixed Gaussian* dengan menggunakan Persamaan (7)

- Pemodelan GWRPCA dengan menggunakan Persamaan (14)
- Interpretasi model

5. Kesimpulan

4. HASIL DAN PEMBAHASAN



Gambar 1. Peta Sebaran Balita Kurang Gizi

Gambar 1. memvisualisasikan data kurang gizi (kekurangan gizi) pada balita di Indonesia pada tahun 2022. Berdasarkan peta tersebut, tampak variasi signifikan tingkat kurang gizi di seluruh provinsi. Provinsi-provinsi di Kalimantan, Sulawesi, dan Papua Barat memiliki tingkat kekurangan gizi yang lebih tinggi dibandingkan dengan provinsi lainnya. Sebaliknya, provinsi di Jawa, Bali, dan Kepulauan Riau menunjukkan tingkat kekurangan gizi yang relatif lebih rendah. Perbedaan ini menunjukkan bahwa program dan intervensi perbaikan gizi perlu disesuaikan dengan kebutuhan spesifik setiap daerah dan tidak dapat disamaratakan. Peta ini memberikan gambaran menyeluruh mengenai sebaran masalah kurang gizi pada balita di Indonesia, yang dapat menjadi dasar bagi perumusan kebijakan dan strategi yang lebih efektif dalam mengatasi masalah tersebut.

Uji KMO digunakan untuk mengukur kecukupan sampel dan seberapa cocok data untuk analisis PCA. Nilai KMO berkisar antara 0 hingga 1, di mana nilai yang lebih tinggi (umumnya > 0,5) menunjukkan bahwa PCA adalah metode yang tepat untuk data tersebut. Uji ini dilakukan untuk memastikan bahwa korelasi antar variabel cukup kuat untuk dianalisis dengan PCA. Diperoleh hasil nilai KMO sebagai berikut:

Tabel 1. Nilai Kaiser-Meyer-Olkin

Nilai KMO Keseluruhan: 0.61									
$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$	$X_{10}$
0.52	0.59	0.63	0.55	0.48	0.48	0.67	0.64	0.80	0.82

Nilai MSA keseluruhan sebesar 0.61 menunjukkan bahwa data dapat dilakukan analisis PCA. Nilai MSA individual bervariasi, dengan  $X_1$  (0.52) dan  $X_2$  (0.59) di bawah ambang batas 0.6, mengindikasikan korelasi yang agak lemah dengan variabel lainnya. Sebaliknya,  $X_9$  (0.80) dan  $X_{10}$  (0.82) menunjukkan kecukupan yang sangat baik, menandakan korelasi kuat dengan variabel lain. Mayoritas variabel memiliki nilai MSA di atas atau mendekati 0.6, sehingga data tetap dianggap cukup untuk melanjutkan analisis PCA.



Uji Bartlett digunakan untuk menguji hipotesis bahwa matriks korelasi adalah matriks identitas, yang berarti bahwa variabel tidak berkorelasi satu sama lain.

Tabel 2. Pengujian korelasi antar variabel prediktor

$H_0: R = I$	
Statistik	Nilai
Bartlett's K-squared	14024
Df	9
p-value	<0.001
Keputusan	Tolak $H_0$

Uji Bartlett menunjukkan nilai Bartlett K-squared sebesar 14024 dengan p-value < 0.05. Karena p-value sangat kecil, hipotesis nol ( $R = I$ ) ditolak. Ini menunjukkan adanya korelasi signifikan antara variabel-variabel, sehingga data layak untuk analisis PCA.

Uji Kolmogorov-Smirnov digunakan untuk menguji normalitas data. Normalitas data ini penting dalam PCA karena memastikan distribusi varian yang konsisten, sehingga komponen utama dapat menangkap variabilitas data secara linier dengan lebih akurat. Distribusi yang normal juga menghasilkan matriks kovarians yang stabil, yang sangat penting untuk proses dekomposisi dan interpretasi komponen utama. Tanpa normalitas, hasil PCA rentan terhadap distorsi akibat outlier atau pola nonlinear, yang dapat mengurangi keandalan dan makna komponen utama. Jika data tidak berdistribusi normal, transformasi data mungkin diperlukan untuk mencapai normalitas.

Tabel 3. Uji normalitas

$H_0$ : Error model berdistribusi normal	
Statistik	Nilai
D	0.0882
p-value	0.9996
Keputusan	Terima $H_0$

Uji normalitas menunjukkan nilai D sebesar 0.0882 dengan p-value 0.9996. Karena p-value sangat tinggi,  $H_0$  diterima. Ini menunjukkan bahwa data berdistribusi normal, sehingga cocok untuk analisis lebih lanjut.

Standarisasi data dilakukan untuk memastikan bahwa setiap variabel memiliki skala yang sama. Langkah ini penting karena PCA sangat peka terhadap skala variabel. Proses standarisasi mengubah data sehingga memiliki rata-rata 0 dan standar deviasi 1, sehingga variabel dengan skala lebih besar tidak akan mendominasi hasil analisis. Komponen utama (principal components) kemudian dibentuk sebagai kombinasi linear dari variabel asli. Setiap komponen utama merupakan vektor eigen dari matriks korelasi, dan komponen-komponen ini disusun berdasarkan besarnya nilai eigen. Komponen utama pertama menjelaskan varians terbesar, diikuti oleh komponen utama kedua, dan seterusnya. Berikut adalah persamaan untuk komponen utama ( $PC_1$  hingga  $PC_{10}$ ):

Tabel 4. Pembentukan komponen utama

Komponen	$PC_1$	$PC_2$	$PC_3$	$PC_4$	$PC_5$	$PC_6$	$PC_7$	$PC_8$	$PC_9$	$PC_{10}$
Persentase Kumulatif Varians	0.345	0.563	0.716	0.814	0.883	0.936	0.967	0.989	0.999	1.000

Berdasarkan persentase kumulatif varians, komponen utama dipilih untuk menjelaskan sebanyak mungkin informasi dalam data. Dari hasil analisis PCA, empat komponen utama pertama menjelaskan 81.4% dari total varians, yang berarti lebih dari 80% informasi dalam data dapat dijelaskan oleh lima komponen utama tersebut. Oleh karena itu, Empat komponen utama

dipilih karena mencakup sebagian besar variabilitas dalam data, memungkinkan analisis yang lebih efisien tanpa kehilangan informasi yang signifikan. Setelah melalui tahapan-tahapan sebelumnya, sejumlah kecil komponen utama yang signifikan berhasil diperoleh. Komponen-komponen utama ini secara efektif mengurangi dimensi data asli sambil mempertahankan sebanyak mungkin informasi. Sehingga diperoleh empat komponen utama sebagai berikut:

$$PC_1 = 0.24X_1 - 0.13X_2 - 0.39X_3 - 0.37X_4 - 0.03X_5 - 0.02X_6 - 0.29X_7 + 0.40X_8 - 0.42X_9 + 0.45X_{10} \quad (17)$$

$$PC_2 = -0.04X_1 - 0.12X_2 - 0.05X_3 + 0.00X_4 + 0.67X_5 + 0.67X_6 - 0.13X_7 + 0.12X_8 + 0.21X_9 + 0.02X_{10} \quad (18)$$

$$PC_3 = 0.56X_1 + 0.16X_2 + 0.61X_3 + 0.27X_4 + 0.19X_5 + 0.07X_6 - 0.93X_7 + 0.87X_8 + 0.33X_9 - 0.77X_{10} \quad (19)$$

$$PC_4 = 0.07X_1 + 0.95X_2 - 0.23X_3 - 0.07X_4 + 0.06X_5 - 0.06X_6 - 0.09X_7 + 0.06X_8 + 0.01X_9 - 0.10X_{10} \quad (20)$$

Uji Breusch-Pagan (BP) digunakan untuk menguji heterogenitas spasial *error* model dari komponen utama yang telah terbentuk. Uji ini menentukan apakah terdapat variabilitas yang signifikan dalam hubungan antar *error model* komponen utama yang dapat dijelaskan oleh lokasi geografis. Jika hasil uji BP signifikan ( $p\text{-value} < 0,05$ ), maka terdapat bukti kuat adanya heterogenitas spasial dalam *error model* komponen utama, yang mengindikasikan perlunya analisis lebih lanjut dengan metode *Geographically Weighted Regression* (GWR). Setelah diperoleh *error model* komponen utama, maka akan dilakukan pengujian heterogenitas spasial sebagai berikut:

Tabel 5. Pengujian heterogenitas spasial

$H_0: \sigma_1 = \sigma_2 = \dots = \sigma_n = \sigma$	
Statistik	Nilai
BP	14.181
p-value	0.0067
Keputusan	Tolak $H_0$

Uji BP menghasilkan nilai 14.181 dengan p-value 0.0067. Karena  $p\text{-value} < 0.05$ ,  $H_0$  ditolak, menunjukkan adanya heterogenitas spasial dalam *error* model komponen utama. Oleh karena itu, penggunaan metode GWR diperlukan untuk menangani variasi spasial dan heteroskedastisitas komponen utama.

Hasil uji multikolinieritas lokal menggunakan  $VIF_j$  untuk 34 lokasi dengan 10 variabel prediktor menunjukkan variasi yang signifikan. Beberapa variabel, seperti variabel kelima dan keenam, menunjukkan nilai  $VIF_j$  yang sangat tinggi di sebagian besar lokasi, mengindikasikan adanya multikolinieritas yang kuat. Variabel lain juga menunjukkan multikolinieritas yang signifikan di beberapa lokasi, seperti variabel ketiga dan keempat. Karena adanya multikolinieritas yang tinggi di banyak lokasi, disarankan untuk menggunakan metode GWRPCA (*Geographically Weighted Regression Principal Component Analysis*). GWRPCA dapat mengurangi multikolinieritas secara lokal, sehingga memberikan analisis yang lebih akurat dan relevan secara spasial.

Langkah pertama dalam analisis GWRPCA adalah menghitung jarak antara titik-titik data. Jarak ini umumnya dihitung menggunakan jarak Euclidean, yang merupakan jarak lurus antara dua titik dalam ruang dua dimensi. Perhitungan jarak ini penting karena GWRPCA memperhitungkan lokasi geografis setiap titik data dalam membangun model lokal.



$$d_{ij} = \begin{bmatrix} 0 & 22.13 & 11.15 & 14.12 & \dots & 18.18 \\ 22.13 & 0 & 10.98 & 9.19 & \dots & 4.69 \\ 11.15 & 10.98 & 0 & 4.20 & \dots & 7.19 \\ 14.12 & 9.19 & 4.20 & 0 & \dots & 4.55 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 18.18 & 4.69 & 7.19 & 4.55 & \dots & 0 \end{bmatrix} \quad (21)$$

Langkah selanjutnya adalah membuat matriks pembobot spasial menggunakan fungsi *fixed kernel Gauss*. Fungsi ini membuat matriks pembobot spasial dalam analisis GWRPCA dengan memberikan bobot lebih tinggi pada titik-titik data yang lebih dekat dan bobot lebih rendah pada titik-titik yang lebih jauh. Pendekatan ini memungkinkan pengaruh data tetangga berkurang secara eksponensial seiring bertambahnya jarak, sehingga model dapat menangkap variasi spasial secara halus. Titik-titik data yang lebih dekat akan memiliki pengaruh lebih besar dalam analisis dibandingkan titik-titik yang lebih jauh, serupa dengan fungsi kernel Gauss, meskipun dengan distribusi penurunan bobot yang berbeda. Pada analisis ini, diperoleh bandwidth sebesar 16.33. Selain itu, fungsi kernel Gauss juga digunakan untuk membentuk matriks pembobot spasial dalam GWRPCA dengan prinsip serupa, yakni memberikan bobot lebih tinggi pada titik-titik data yang lebih dekat dan bobot lebih rendah pada titik yang lebih jauh, serupa dengan fungsi kernel bisquare namun dengan distribusi penurunan bobot yang berbeda. Sehingga diperoleh matriks pembobot sebagai berikut:

$$w_{ij} = \begin{bmatrix} 1 & 0.992 & 0.998 & 0.997 & \dots & 0.995 \\ 0.976 & 1 & 0.994 & 0.996 & \dots & 0.999 \\ 0.996 & 0.997 & 1 & 0.999 & \dots & 0.998 \\ 0.995 & 0.996 & 0.999 & 1 & \dots & 0.998 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0.988 & 0.999 & 0.998 & 0.997 & \dots & 1 \end{bmatrix}$$

Pemodelan GWRPCA dilakukan dengan menggabungkan prinsip Geographically Weighted Regression (GWR) dan Principal Component Analysis (PCA). Pertama, PCA diterapkan untuk mereduksi dimensi data dan mengatasi multikolinieritas dengan membentuk komponen utama. Selanjutnya, GWR digunakan pada komponen utama ini, menghasilkan model regresi lokal di setiap titik data. Pendekatan ini memungkinkan model menangkap variasi spasial dalam hubungan antar variabel dengan lebih akurat. Berdasarkan penjabaran di atas telah diperoleh  $PC_1$ ,  $PC_2$ ,  $PC_3$  dan  $PC_4$  selanjutnya dapat dilakukan pemodelan GWRPCA, selanjutnya akan memperoleh model lokal untuk masing-masing daerah. Contoh model GWRPCA di Provinsi Bengkulu:

$$y_5 = -3.32 \times 10^{11} PC_1 + 10057.22 PC_2 + 60.15 PC_3 + 19.38 PC_4 \quad (22)$$

Dengan masing-masing nilai  $PC$  sebagai berikut:

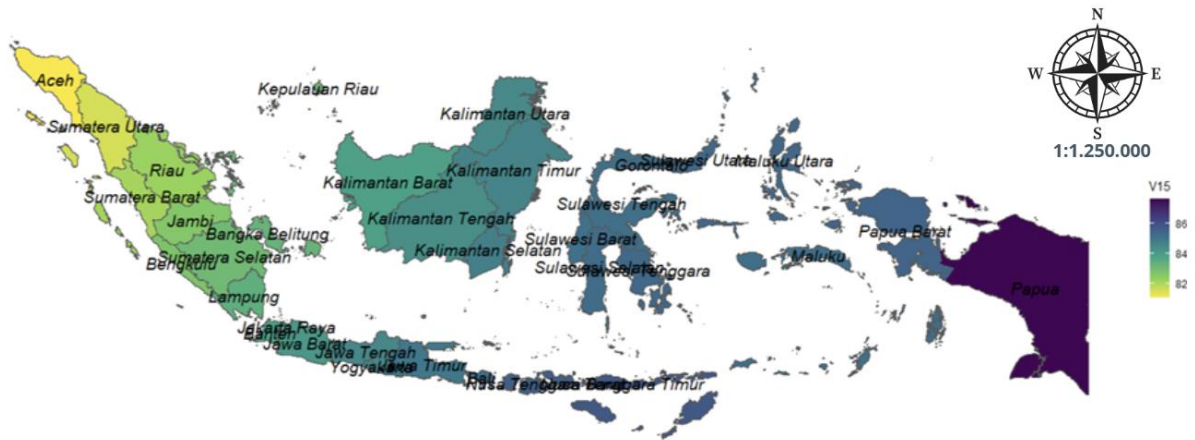
$$PC_1 = -0.27X_1 + 0.03X_2 + 0.42X_3 + 0.38X_4 + 0.03X_5 + 0.02X_6 + 0.25X_7 - 0.43X_8 + 0.41X_9 - 0.44X_{10} \quad (23)$$

$$PC_2 = 0.05X_1 - 0.07X_2 - 0.06X_3 + 0.03X_4 + 0.68X_5 + 0.68X_6 - 0.14X_7 + 0.11X_8 + 0.19X_9 + 0.002X_{10} \quad (24)$$

$$PC_3 = 0.36X_1 - 0.06X_2 + 0.45X_3 + 0.49X_4 - 0.04X_5 - 0.05X_6 - 0.53X_7 + 0.19X_8 - 0.28X_9 - 0.12X_{10} \quad (25)$$

$$PC_4 = 0.00X_1 + 0.95X_2 + 0.05X_3 - 0.05X_4 + 0.04X_5 + 0.05X_6 + 0.03X_7 + 0.23X_8 - 0.05X_9 - 0.19X_{10} \quad (26)$$

Model GWRPCA menunjukkan bahwa di Provinsi Bengkulu, komponen utama pertama ( $PC_1$ ) memiliki pengaruh negatif yang sangat besar terhadap  $y_5$  menandakan bahwa peningkatan variabel-variabel di dalam  $PC_2$  cenderung menurunkan  $y_5$ . Sementara itu, komponen utama  $PC_2$ ,  $PC_3$ , dan  $PC_4$  memiliki pengaruh positif, tetapi relatif kecil, yang menunjukkan pengaruh positif terhadap  $y_5$ . Berdasarkan model-model lokal yang telah diperoleh, total keragaman lokal untuk setiap komponen utama dapat dihitung. Perhitungan ini memungkinkan pemahaman mengenai seberapa besar keragaman lokal yang terdapat dalam berbagai komponen utama di masing-masing provinsi di Indonesia. Persentase total keragaman lokal dapat dihitung untuk memperoleh wawasan tentang variasi dan distribusi keragaman ini, yang kemudian divisualisasikan melalui peta pada Gambar 2.



Gambar 2. Peta persentase total keragaman lokal komponen utama

Gambar 2. menunjukkan persentase keragaman lokal untuk komponen utama di berbagai provinsi Indonesia. Ringkasan dari analisis ini dan interpretasinya dapat dilihat pada tabel di bawah, yang memberikan deskripsi singkat untuk beberapa wilayah berdasarkan persentase keragaman lokal.

Tabel 6. Interpretasi persentase keragaman lokal

Wilayah	Persentase Keragaman	Warna Peta
Sumatera Utara	Rendah	Hijau Terang - Kuning
Aceh	Rendah	Hijau Terang - Kuning
Papua	Tinggi	Ungu
Provinsi Sekitar Papua	Tinggi	Ungu
Kalimantan	Sedang	Hijau Sedang - Biru
Sulawesi	Sedang-Hingga-Tinggi	Hijau Sedang - Biru

Tabel di atas memberikan rangkuman terkait persentase keragaman lokal untuk komponen utama di berbagai provinsi di Indonesia. Rangkuman ini menyajikan tingkat keragaman dan kondisi di setiap wilayah, yang berguna untuk memahami perbedaan distribusi faktor lokal, seperti akses layanan kesehatan dan sumber daya gizi, yang memengaruhi kondisi gizi balita di wilayah-wilayah tersebut. Keunikan penelitian ini terletak pada penggunaan metode GWRPCA, yang mampu mengatasi multikolinieritas dan heterogenitas spasial dalam data geospasial. Metode GWRPCA ini menjadikan analisis lebih efisien dan menghasilkan empat

komponen utama yang menjelaskan lebih dari 80% variasi data. Dengan menghitung Average Squared Loading Error (ASLE), penelitian ini juga mampu mengevaluasi konsistensi spasial dari komponen utama di seluruh lokasi. Berikut nilai ASLE untuk setiap komponen:

Tabel 7. Evaluasi komponen utama

Komponen	$PC_1$	$PC_2$	$PC_3$	$PC_4$
Nilai	0.0993	0.1023	0.0976	0.1010

ASLE menunjukkan rata-rata kuadrat dari perbedaan antara loadings lokal dan loadings global untuk setiap komponen utama. Nilai ASLE yang lebih kecil menunjukkan bahwa struktur komponen utama di berbagai lokasi lebih konsisten dengan struktur globalnya, sedangkan nilai ASLE yang lebih besar menunjukkan adanya variasi spasial yang lebih tinggi dalam loadings untuk komponen tersebut.

## 5. KESIMPULAN

Pendekatan *Geographically Weighted Regression Principal Component Analysis* (GWRPCA) dipilih untuk mengatasi masalah multikolinieritas dan heterogenitas spasial dalam analisis data geospasial. Masalah multikolinieritas terjadi ketika variabel-variabel independen memiliki korelasi tinggi, sedangkan heterogenitas spasial mengacu pada variasi data di berbagai lokasi. Model lokal GWRPCA berhasil mereduksi dimensi data, menghasilkan empat komponen utama yang menjelaskan 82-84% dari total variansi data. ilai ASLE yang rendah menunjukkan konsistensi spasial yang lebih tinggi, yang berarti struktur komponen utama relatif seragam di berbagai wilayah, sementara nilai ASLE yang lebih tinggi menunjukkan adanya variasi spasial antar lokasi. Pendekatan ini juga dapat menangkap variasi lokal di setiap titik geografis, memberikan visualisasi keragaman lokal yang lebih akurat dan mendukung intervensi kebijakan yang lebih efektif sesuai dengan kebutuhan masing-masing daerah.

## DAFTAR PUSTAKA

- Aini, Z.U.N. & Margareta, J., 2023. Analisis hubungan penduduk miskin, sanitasi dan imunisasi dasar dengan kejadian gizi buruk pada balita. *Hearty*, 11(1), pp.42-48.
- Alvin C. Rencher & Christensen, W. F. 1995. *Methods of Multivariate Analysis*. A John Wiley & Sons, Inc., Publication. Available at: <https://doi.org/10.5860/choice.33-1586>.
- Basyariyah, Q., Diyanah, K.C. & Pawitra, A.S., 2022. Hubungan ketersediaan sanitasi dasar terhadap status gizi baduta di Desa Pelem, Bojonegoro. *Jurnal Kesehatan Lingkungan Indonesia*, 21(1), pp.18-26.
- Chen, J., Qu, M., Zhang, J., Xie, E., Zhao, Y. & Huang, B., 2021. Improving The Spatial Prediction Accuracy of Soil Alkaline Hydrolyzable Nitrogen Using GWRPCA-GWRK. *Soil Science Society of America Journal*, Vol. 85, pp.879-892.
- Emmanuel, B. O., Maureen, N., & Wonu, N. 2020. Detection of Non-Normality in Data Sets and Comparison Between Different Normality Tests. *AJPAS*.
- Fotheringham, A. S., Brunsdon, C., & Charlton, M. 2002. *Geographically Weighted Regression. 1st ed.* John Wiley & Sons, Ltd.
- Han, J., Kang, X., Yang, Y. & Zhang, Y., 2023. Geographically And Temporally Weighted Principal Component Analysis: A New Approach for Exploring Air Pollution Non-Stationarity in China, 2015–2019. *Journal of Spatial Science*, Vol 68, pp.451-468.
- Johnson, R.A. & Wichern, D.W., 2002. *Applied Multivariate Statistical Analysis. 6th edition.* New Jersey: Pearson Prentice Hall.

- Kemenkes RI. 2020. *Buku Saku Pencegahan dan Tata Laksana Gizi Buruk Pada Balita di Layanan Rawat Jalan Bagi Tenaga Kesehatan*. Kemenkes RI: Jakarta.
- Kevin, A. 2020. *Dampak Gizi Buruk Terhadap Kesehatan dan Daya Tahan Tubuh - Alodokter*. Alodokter. Available at: <https://www.alodokter.com/dampak-gizi-buruk-terhadap-kesehatan-dan-daya-tahan-tubuh>.
- Li, G., Chen, W., Li, R., Chen, Y., Bi, H., Zhao, H. & Li, L., 2020. Prediction Of AOD Data By Geographical And Temporal Weighted Regression With Nonlinear Principal Component Analysis. *Arabian Journal of Geosciences*, Vol 23, p.876.
- Mei, C.L., He, S.Y. & Fang, K.T., 2004. A Note on The Mixed Geographically Weighted Regression Model. *Journal of Regional Science*, Vol 44, pp.143-157.
- Myers, R.H. & Myers, R.H., 1990. *Classical and Modern Regression with Applications*. 2nd edition. Belmont, CA: Duxbury Press.
- Nunes, A. F., Monteiro, P. L., & Nunes, A. S. 2020. Factor Structure of The Convergence Insufficiency Symptom Survey Questionnaire. *Plos one*, Vol. 15.
- Nursiyono, J.A. & Apriyani, M., 2023. Pengaruh jumlah pelayanan kesehatan dan jumlah balita kurus terhadap kejadian stunting pada balita di Jawa Timur 2020. *Spirakel*, 15(1), pp.1-8.
- Pratnyaningrum, N., Yasin, H. & Hoyyi, A., 2015. Pemodelan Persentase Balita Gizi Buruk di Jawa Tengah dengan Pendekatan Geographically Weighted Regression Principal Components Analysis (GWRPCA). *Jurnal Gaussian*, Vol. 4, pp.171-180.
- Putri, F. K., & Imro'ah, N. 2021. Pemodelan persentase angka kematian bayi di Kalimantan Barat dengan metode geographically weighted regression principal component analysis (GWRPCA). *Buletin Ilmiah Math. Stat. dan Terapan (Bimaster)*, Vol. 10, pp. 117–124.
- Sari, N., Yasin, H., & Prahutama, A. 2016. Geographically Weighted Regression Principal Component Analysis (GWRPCA) pada pemodelan pendapatan asli daerah di Jawa Tengah. *Jurnal Gaussian*, Vol.5, pp. 717–726. Available at: <http://ejournal-s1.undip.ac.id/index.php/gaussian>
- Simamora, B., 2005. *Analisis Multivariat Pemasaran*. Jakarta: Gramedia Pustaka Utama.
- Wang, Q., Jiang, D., Gao, Y., Zhang, Z. & Chang, Q., 2022. Examining the Driving Factors of SOM Using a Multi-Scale GWR Model Augmented by Geo-Detector and GWRPCA Analysis. *Agronomy*, Vol.12, p.1697.
- Yildirim, H., Isik, K., & Cengizhan, S. Ö. 2021. Psychometric evaluation of the healthy aging instrument in older adults. *Educational Gerontology*, Vol.47, 235-246.