

## ***GEOGRAPHICALLY WEIGHTED REGRESSION PRINCIPAL COMPONENT ANALYSIS (GWRPCA) PADA PEMODELAN PENDAPATAN ASLI DAERAH DI JAWA TENGAH***

Nurmalita Sari<sup>1</sup>, Hasbi Yasin<sup>2</sup>, Alan Prahutama<sup>3</sup>

<sup>1</sup>Mahasiswa Departemen Statistika FSM Universitas Diponegoro  
<sup>2,3</sup>Staff Pengajar Departemen Statistika FSM Universitas Diponegoro

### **ABSTRACT**

Linear Regression Analysis is a method for modeling the relation between a response variable with two or more independent variables. Geographically Weighted Regression (GWR) is a development of the regression model where each observation location has different regression parameter values because of the effects of spatial heterogeneity. Regression Principal Component Analysis (PCA) is a combination of PCA and are used to remove the effect of multicollinearity in regression. Geographically Weighted Regression Principal Component Analysis (GWRPCA) is a combination of PCA and GWR if spatial heterogeneity and local multicollinearity occurred. Estimation parameters for the GWR and GWRPCA using Weighted Least Square (WLS). Weighting use fixed gaussian kernel function through selection of the optimum bandwidth is 0,08321242 with minimum Cross Validation (CV) is 3,009035. There are some variables in PCA that affect locally-generated revenue in Central Java on 2012 and 2013, which can be represented by  $PC_1$  that explained the total variance data about 71,4%. GWRPCA is a better model for modeling locally-generated revenue for the districts and cities in Central Java than RPCA because it has the the smallest Akaike Information Criterion (AIC) and the largest  $R^2$ .

**Keywords** : Spatial Heterogeneity, Local Multicollinearity, Principal Component Analysis, Geographically Weighted Regression Principal Component Analysis.

### **1. PENDAHULUAN**

Analisis regresi digunakan untuk mengetahui pengaruh dua atau lebih variabel bebas terhadap sebuah variabel respon. Model regresi dikatakan baik atau cocok, jika memenuhi asumsi-asumsi ideal, antara lain tidak adanya autokorelasi, heteroskedastisitas dan multikolinieritas. Multikolinieritas terjadi ketika di dalam model regresi ada beberapa atau semua variabel bebas mempunyai hubungan linier sempurna atau hampir sempurna. Metode alternatif yang dapat digunakan untuk mengatasi permasalahan multikolinieritas dalam analisis regresi yaitu *Principal Component Analysis* (PCA).

Selain itu, pada model regresi diasumsikan nilai parameter regresi akan selalu tetap. Parameter regresi akan mempunyai nilai yang sama untuk setiap wilayah pengamatan. Apabila ada pengaruh unsur wilayah dalam model regresi, maka metode regresi akan kurang tepat dilakukan. Perbedaan karakteristik wilayah yang satu dengan wilayah yang lainnya diketahui dengan uji heterogenitas spasial. Salah satu metode untuk mengatasi masalah heterogenitas spasial yaitu *Geographically Weighted Regression* (GWR).

Pendapatan Asli Daerah (PAD) biasanya ditunjukkan dengan kecenderungan wilayah yang mempunyai kekayaan dan potensi-potensi yang melimpah, sehingga perlu ditambahkan informasi geografis dari lokasi daerah yang diamati. Hal ini melibatkan efek spasial dalam menganalisis faktor-faktor yang mempengaruhi PAD sehingga menimbulkan kondisi heterogenitas spasial. Banyaknya faktor yang mempengaruhi PAD memungkinkan terdapatnya multikolinieritas lokal. Sehingga kemungkinan pada data PAD terdapat heterogenitas spasial dan multikolinieritas lokal. Salah satu metode untuk mengatasi

masalah heterogenitas spasial dan multikolinieritas lokal yaitu *Geographically Weighted Regression Principal Component Analysis* (GWRPCA). Berdasarkan uraian tersebut, peneliti ingin membandingkan RPCA dan GWRPCA mengenai faktor-faktor yang mempengaruhi PAD di Jawa Tengah.

## 2. TINJAUAN PUSTAKA

### 2.1 Analisis Regresi Linier Berganda

Analisis regresi linier berganda merupakan suatu metode yang digunakan untuk memodelkan hubungan antara sebuah variabel respon  $y$  dengan dua atau lebih variabel bebas  $x$  [5]. Model regresi linier berganda adalah sebagai berikut:

$$y_i = \beta_0 + \sum_{j=1}^k \beta_j X_{ij} + \varepsilon_i \quad ; \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, k$$

Model persamaan regresi linear berganda dapat ditulis dalam bentuk matriks  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$  [5] dimana :  $\mathbf{y}$  = vektor ( $n \times 1$ ) dari observasi

$\mathbf{X}$  = matriks  $n \times (k+1)$  dari variabel bebas

$\boldsymbol{\beta}$  = vektor  $(k+1) \times 1$  dari koefisien regresi

$\boldsymbol{\varepsilon}$  = vektor ( $n \times 1$ ) dari *error* random

Metode yang digunakan untuk menaksir parameter  $\beta$  adalah metode kuadrat terkecil dengan meminimumkan jumlah kuadrat *error* sehingga diperoleh nilai estimasi  $\beta$  [5].

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

#### 2.1.1 Uji Hipotesis dalam Regresi Linier Berganda

Uji signifikansi regresi bertujuan untuk menentukan apakah ada hubungan linier antara respon dengan variabel bebas  $X_1, X_2, \dots, X_k$  secara bersama-sama [5].

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1 : \text{minimal ada satu } \beta_j \neq 0 \text{ dengan } j = 1, 2, \dots, k$$

Statistik uji :

$$F_{hitung} = \frac{SSR/df_1}{SSE/df_2}$$

dengan  $SSR = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 = \mathbf{y}^T (\mathbf{H} - \mathbf{I} \frac{1}{n}) \mathbf{y}$ ,  $SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \mathbf{y}^T (\mathbf{I} - \mathbf{H}) \mathbf{y}$ ,  $df_1 =$  banyaknya variabel bebas dan  $df_2 = n - k - 1$ , dimana  $\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ ,  $n$  adalah banyaknya data dan  $k$  adalah banyaknya variabel bebas.

Tolak  $H_0$  jika nilai  $F_{hitung} > F_{(\alpha; df_1; df_2)}$ .

Sedangkan uji signifikansi koefisien regresi secara individu dilakukan dengan hipotesis sebagai berikut [5]:

$$H_0 : \beta_j = 0 \text{ (tidak ada pengaruh variabel bebas } X_j \text{ terhadap variabel respon } Y)$$

$$H_1 : \beta_j \neq 0 \text{ dengan } j = 1, 2 \dots k \text{ (ada pengaruh variabel bebas } X_j \text{ terhadap variabel respon } Y)$$

Statistik uji :

$$t_{hitung} = \frac{\hat{\beta}_j}{\sqrt{Se(\hat{\beta}_j)}}$$

dengan  $Se(\hat{\beta}_j)$  adalah  $\sqrt{\hat{\sigma}^2 C_{jj}}$ ,  $C_{jj}$  adalah elemen diagonal dari  $(\mathbf{X}^T \mathbf{X})^{-1}$

Tolak  $H_0$  jika nilai  $|t_{hitung}| > t_{(\frac{\alpha}{2}; n-k-1)}$ .

#### 2.1.2 Uji Asumsi Analisis Regresi Linier Berganda

Multikolinieritas berarti adanya hubungan linear yang sempurna atau pasti, di antara beberapa atau semua variabel bebas dari model regresi ganda. Jika nilai VIF  $> 10$ , maka menunjukkan bahwa multikolinieritas adalah masalah yang pasti terjadi antar variabel bebas. Menghitung nilai VIF diperoleh dengan rumus [2]:

$$VIF_j = \frac{1}{1 - R_j^2}$$

$R_j^2$  adalah koefisien determinasi antara  $X_j$  dengan variabel bebas lainnya dan  $j= 1, 2, \dots, k$ .

Selain itu pada model regresi diasumsikan bahwa nilai estimasi parameter regresi akan tetap (konstan), artinya parameter regresi akan sama untuk setiap lokasi pengamatan<sup>[8]</sup>. Apabila diduga ada pengaruh unsur wilayah atau spasial dalam model regresi, maka metode regresi akan kurang tepat dilakukan. Maka selanjutnya akan dilakukan uji heterogenitas spasial menggunakan *Breusch Pagan Test*<sup>[7]</sup>.

$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2$  (homoskedastisitas)

$H_1 : \text{Minimal ada satu } \sigma_i^2 \neq \sigma^2 \text{ dengan } i=1, 2, \dots, n$  (heteroskedastisitas)

Statistik uji :

$$BP = \frac{1}{2} \mathbf{f}' \mathbf{Z} (\mathbf{Z}' \mathbf{Z})^{-1} \mathbf{Z}' \mathbf{f} \sim \chi^2_{(\alpha; k)}$$

dengan  $\mathbf{f} = (f_i)^T$  dimana  $f_i = \left(\frac{e_i^2}{\sigma^2} - 1\right)$ ,  $e_i$  merupakan residual untuk pengamatan ke- $i$  dan  $\mathbf{Z}$ = matriks berukuran  $n \times (k+1)$  dari variabel bebas yang berisi vektor yang sudah dinormal-standarkan untuk setiap pengamatan

Keputusan menolak  $H_0$  jika nilai  $BP > \chi^2_{(\alpha; k)}$

## 2.2 Uji Asumsi *Principal Component Analysis* (PCA)

Hipotesis untuk uji kecukupan sampel secara keseluruhan menggunakan *Kaiser-Mayer-Olkin* (KMO) sebagai berikut:

$H_0$  : ukuran data cukup untuk dianalisis komponen utama

$H_1$  : ukuran data tidak cukup untuk dianalisis komponen utama

Statistik Uji:

$$KMO = \frac{\sum_{j=1}^p \sum_{k=1}^p r_{jk}^2}{\sum_{j=1}^p \sum_{k=1}^p r_{jk}^2 + \sum_{j=1}^p \sum_{k=1}^p a_{jk}^2}, \text{ untuk } j= 1, 2, 3, \dots, p \text{ dan } k= 1, 2, 3, \dots, p$$

Tolak  $H_0$  jika nilai KMO kurang dari 0,5<sup>[8]</sup>.

Sedangkan hipotesis uji kecukupan data untuk masing-masing variabel menggunakan *Measure Sampling of Adequacy* (MSA) sebagai berikut:

$H_0$  : variabel memadai untuk dianalisis lebih lanjut

$H_1$  : variabel tidak memadai untuk dianalisis lebih lanjut

Statistik Uji:

$$MSA_j = \frac{\sum_{k=1}^p r_{jk}^2}{\sum_{k=1}^p r_{jk}^2 + \sum_{k=1}^p a_{jk}^2}$$

dengan  $r_{jk}$ = koefisien korelasi antara variabel  $j$  dan  $k$  dan  $a_{jk}$ = koefisien korelasi parsial antara variabel  $j$  dan  $k$ .

Tolak  $H_0$  jika nilai MSA kurang dari 0,5.

## 2.2 *Principal Component Analysis* (PCA)

*Principal Components Analysis* (PCA) merupakan suatu teknik statistika untuk mentransformasi variabel-variabel asli yang saling berkorelasi satu dengan yang lain menjadi satu set variabel baru yang tidak berkorelasi lagi. Misalkan vektor random  $\mathbf{X}^T = [X_1, X_2, \dots, X_k]$  mempunyai matriks kovarian  $\Sigma$  dengan nilai eigen  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k \geq 0$  maka bentuk kombinasi liniernya sebagai berikut<sup>[3]</sup>:

$$PC_1 = \mathbf{a}_1^T \mathbf{X} = a_{11}X_1 + a_{12}X_2 + \dots + a_{1k}X_k$$

$$PC_2 = \mathbf{a}_2^T \mathbf{X} = a_{21}X_1 + a_{22}X_2 + \dots + a_{2k}X_k$$

⋮

$$PC_k = \mathbf{a}_k^T \mathbf{X} = a_{k1}X_1 + a_{k2}X_2 + \dots + a_{kk}X_k$$

$$\text{dengan } \text{Var}(PC_j) = \mathbf{a}_j^T \Sigma \mathbf{a}_j \quad j = 1, 2, \dots, k$$

Varian total dari komponen-komponen yang terbentuk adalah:

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

dengan  $\lambda_j$  adalah nilai eigen dari komponen utama ke- $j$ ,  $j = 1, 2, \dots, k$

Ada tiga kriteria dalam pemilihan komponen utama yang digunakan yaitu jika  $\lambda_i > 1$ , melihat sudut pada *scree plot* yang menunjukkan perubahan nilai eigen yang besar, dan proporsi variansi kumulatif mencapai 70% sampai dengan 80%<sup>[3]</sup>.

### 2.3 Geographically Weighted Regression (GWR)

*Geographically Weighted Regression* (GWR) merupakan pengembangan dari model regresi dimana setiap parameter dihitung pada setiap lokasi pengamatan, sehingga setiap lokasi pengamatan mempunyai nilai parameter regresi yang berbeda-beda<sup>[1]</sup>. Model GWR dapat dinyatakan dalam bentuk:

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

Estimasi parameter model GWR menggunakan metode *Weighted Least Square* (WLS) dengan memberikan pembobot yang berbeda untuk setiap lokasi amatan<sup>[1]</sup>. Fungsi pembobot untuk setiap lokasi  $(u_i, v_i)$  adalah  $w_m(u_i, v_i)$  dimana  $m = 1, 2, \dots, n$  dan dapat ditulis dalam bentuk matriks sebagai berikut:

$$\mathbf{W}(u_i, v_i) = \begin{bmatrix} w_{i1} & 0 & \dots & 0 \\ 0 & w_{i2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_{in} \end{bmatrix}$$

Selanjutnya meminimumkan jumlah kuadrat residual, sehingga diperoleh estimasi parameter model GWR sebagai berikut:

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

Untuk memberikan hasil penaksiran parameter yang berbeda pada lokasi yang berbeda maka dibutuhkan fungsi pembobot. Menentukan besarnya pembobot untuk masing-masing lokasi yang berbeda pada model GWR dapat menggunakan fungsi kernel. Ada dua jenis fungsi kernel yang digunakan dalam GWR, yaitu *adaptive* dan *fixed*<sup>[1]</sup>. Fungsi kernel memberikan pembobot sesuai *bandwidth* optimum yang nilainya sesuai dengan kondisi data. Salah satu metode yang digunakan untuk memilih *bandwidth* optimum menggunakan metode *Cross Validation* (CV)<sup>[1]</sup>.

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

### 2.4 Geographically Weighted Regression Principal Component Analysis (GWRPCA)

GWRPCA merupakan gabungan dari metode GWR dan PCA. Setelah dilakukan uji heterogenitas spasial dan diketahui terdapat perbedaan karakteristik suatu wilayah dengan wilayah lainnya maka metode GWR dapat digunakan. Apabila terdapat multikolinieritas lokal pada model GWR, selanjutnya dilakukan analisis komponen utama untuk menyederhanakan variabel-variabel bebas yang saling berkorelasi dengan menggunakan metode PCA. Multikolinieritas lokal merupakan korelasi antara variabel bebas satu dengan variabel bebas lainnya dalam setiap lokasi pengamatan. Kemudian dilakukan regresi antara variabel respon  $y$  dengan komponen utama terpilih dengan menambahkan *bandwidth*. Uji heterogenitas spasial dapat menggunakan uji *Breusch Pagan*<sup>[7]</sup>. Sedangkan untuk mengetahui adanya multikolinieritas lokal dapat dihitung dari nilai VIF > 10 dengan rumus:

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

dengan  $R_j^2(u_i, v_i)$  merupakan koefisien determinasi antara variabel bebas  $X_j$  dengan variabel bebas lainnya untuk setiap lokasi pengamatan  $(u_i, v_i)$ .

Model GWRPCA dapat dinyatakan dalam bentuk:

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

dengan:  $y_i$  : nilai observasi variabel respon ke- $i$

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

- $\beta_0(u_i, v_i)$ : konstanta / *intercept* pada pengamatan ke- $i$
- $(u_i, v_i)$  : menyatakan titik koordinat (*longitude, latitude*) lokasi ke- $i$
- $\beta_j(u_i, v_i)$  : koefisien regresi variabel bebas ke- $j$  pada lokasi pengamatan ke- $i$
- $\varepsilon_i$  : *error* pengamatan ke- $i$

### 2.5.1 Pengujian Kesesuaian Model GWRPCA

Uji kesesuaian model dapat dilakukan dengan hipotesis sebagai berikut<sup>[7]</sup>:

$H_0$ :  $\beta_j(u_i, v_i) = \beta_j$ , untuk setiap  $j = 0, 1, 2, \dots, k$  dan  $i = 1, 2, \dots, n$  (tidak ada perbedaan antara RPCA dengan GWRPCA)

$H_1$ : Minimal ada satu  $\beta_j(u_i, v_i) \neq \beta_j$ , (ada perbedaan antara RPCA dengan GWRPCA)

Statistik Uji:

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

dengan  $RSS(H_0) = \mathbf{y}^T(\mathbf{I} - \mathbf{H})\mathbf{y}$ ,  $RSS(H_1) = \mathbf{y}^T(\mathbf{I} - \mathbf{L})^T(\mathbf{I} - \mathbf{L})\mathbf{y}$ ,  $df_1 = \frac{\delta_1^2}{\delta_2^2}$ ,  $df_2 = n - p - 1$ , dimana

$$\delta_i^2 = \text{tr}[(\mathbf{I} - \mathbf{L})^T(\mathbf{I} - \mathbf{L})^i] \text{ untuk } i = 1, 2, \text{ dan } \mathbf{L} = \begin{bmatrix} \mathbf{x}_1^T [\mathbf{X}^T \mathbf{W}(u_1, v_1) \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{W}(u_1, v_1) \\ \mathbf{x}_2^T [\mathbf{X}^T \mathbf{W}(u_2, v_2) \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{W}(u_2, v_2) \\ \vdots \\ \mathbf{x}_n^T [\mathbf{X}^T \mathbf{W}(u_n, v_n) \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{W}(u_n, v_n) \end{bmatrix}$$

$\mathbf{x}_i^T = (1, x_{i1}, x_{i2}, \dots, x_{ik})$  adalah elemen baris ke- $i$  pada matriks  $\mathbf{X}$

$H_0$  akan ditolak jika nilai  $F_1 \geq F_{\alpha, df_1, df_2}$

### 2.5.2 Pengujian Parsial Parameter Model GWRPCA

Uji parsial parameter model GWR dilakukan dengan hipotesis sebagai berikut<sup>[7]</sup>:

$H_0$ :  $\beta_j(u_1, v_1) = \beta_j(u_2, v_2) = \dots = \beta_j(u_n, v_n)$  untuk  $j = 0, 1, 2, \dots, k$  (tidak ada perbedaan pengaruh yang signifikan dari variabel  $x_j$  antara satu lokasi dengan lokasi lainnya)

$H_1$ : Minimal ada satu  $\beta_j(u_i, v_i)$  untuk  $i = 1, 2, \dots, n$  yang berbeda (ada perbedaan pengaruh yang signifikan dari variabel  $x_j$  antara satu lokasi dengan lokasi lainnya)

Statistik Uji:

$$F_3 = \frac{V_j^2 / \text{tr} \left( \frac{1}{n} \mathbf{B}_j^T \left[ \mathbf{I} - \frac{1}{n} \mathbf{J} \right] \mathbf{B}_j \right)}{RSS(H_1) / \delta_1}$$

dengan:  $V_j^2 = \frac{1}{n} \sum_{i=1}^n \left( \hat{\beta}_j(u_i, v_i) - \frac{1}{n} \sum_{i=1}^n \hat{\beta}_j(u_i, v_i) \right)^2 = \frac{1}{n} \mathbf{B}_j^T \left[ \mathbf{I} - \frac{1}{n} \mathbf{J} \right] \mathbf{B}_j$ ,

$$\mathbf{B}_j = \begin{bmatrix} \mathbf{e}_j^T [\mathbf{X}^T \mathbf{W}(u_1, v_1) \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{W}(u_1, v_1) \\ \mathbf{e}_j^T [\mathbf{X}^T \mathbf{W}(u_2, v_2) \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{W}(u_2, v_2) \\ \vdots \\ \mathbf{e}_j^T [\mathbf{X}^T \mathbf{W}(u_n, v_n) \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{W}(u_n, v_n) \end{bmatrix}, \mathbf{B}_j(u_i, v_i) = \begin{bmatrix} \beta_0(u_i, v_i) \\ \beta_1(u_i, v_i) \\ \vdots \\ \beta_k(u_i, v_i) \end{bmatrix}$$

$H_0$  akan ditolak jika nilai  $F_3 \geq F_{\alpha, df_1, df_2}$  dengan  $df_1 = \left( \frac{\gamma_1^2}{\gamma_2} \right)$ ,  $df_2 = \left( \frac{\delta_1^2}{\delta_2} \right)$ , dimana  $\gamma_i = \text{tr} \left( \frac{1}{n} \mathbf{B}_j^T \left[ \mathbf{I} - \frac{1}{n} \mathbf{J} \right] \mathbf{B}_j \right)^i$ ,  $i = 1, 2$ .

Hipotesis uji signifikansi parameter model pada setiap lokasi sebagai berikut<sup>[7]</sup>:

$H_0$ :  $\beta_j(u_i, v_i) = 0$ , dengan  $j = 1, 2, \dots, k$  dan  $i = 1, 2, \dots, n$  (variabel  $x_j$  pada lokasi ke- $i$  tidak mempengaruhi variabel respon di lokasi tersebut)

$H_1$ :  $\beta_j(u_i, v_i) \neq 0$ , (variabel  $x_j$  pada lokasi ke- $i$  mempengaruhi variabel respon di lokasi tersebut)

Statistik uji:

$$T_{hit} = \frac{\hat{\beta}_j(u_i, v_i)}{\hat{\sigma} \sqrt{c_{kk}}}$$

dengan  $c_{kk}$  elemen diagonal ke- $k$  dari matriks  $\mathbf{C}_i \mathbf{C}_i^T$  dan  $\mathbf{C}_i = (\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i)$  dan  $\hat{\sigma} = \sqrt{\frac{RSS(H_1)}{\delta_1}}$

$H_0$  akan ditolak jika nilai  $|t_{hit}| > t_{\frac{\alpha}{2}, df}$  dengan  $df = \left(\frac{\delta_1^2}{\delta_2}\right)$

## 2.6 Pemilihan Model Terbaik

Metode yang digunakan untuk mendapatkan model terbaik, diantaranya dengan nilai  $R^2$  terbesar dan AIC terkecil. Dalam GWRPCA terdapat dua koefisien determinasi yaitu koefisien determinasi global dan koefisien determinasi lokal. Koefisien determinasi global didefinisikan sebagai berikut<sup>[6]</sup>:

$$R^2 = \frac{JKR}{JKT} = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

Koefisien determinasi lokal menunjukkan bagaimana model lokal dapat meniru data yang tercatat dalam daerah sekitar titik regresi dengan baik dengan rumus<sup>[11]</sup>:

$$r_i^2 = \frac{(TSS^w - RSS^w)}{TSS^w}$$

Sedangkan untuk menghitung nilai AIC digunakan rumus sebagai berikut<sup>[11]</sup>:

$$AIC = 2n \ln(\hat{\sigma}) + n \ln(2\pi) + n + \text{tr}(\mathbf{L})$$

## 2.7 Pendapatan Asli Daerah

Pendapatan daerah meliputi semua penerimaan uang melalui rekening kas daerah, yang merupakan hak daerah dalam satu tahun anggaran dan tidak perlu dibayar kembali oleh daerah, salah satunya Pendapatan Asli Daerah<sup>[4]</sup>. Faktor-faktor yang mempengaruhi PAD yaitu jumlah penduduk, retribusi daerah, belanja daerah, PDRB, dan pajak daerah<sup>[9]</sup>.

## 3. METODE PENELITIAN

Data yang digunakan dalam penelitian ini merupakan data sekunder tentang faktor-faktor yang diduga mempengaruhi PAD di Jawa Tengah tahun 2012 dan 2013. Data ini bersumber dari Jawa Tengah dalam Angka 2013 dan Jawa Tengah dalam Angka 2014. Variabel bebas yang digunakan dalam penelitian ini adalah jumlah penduduk ( $X_1$ ), retribusi daerah ( $X_2$ ), belanja daerah ( $X_3$ ), PDRB atas dasar harga konstan ( $X_4$ ), PDRB atas dasar harga berlaku ( $X_5$ ) dan pajak daerah ( $X_6$ ).

Langkah-langkah untuk menganalisis data dalam penelitian ini yaitu:

1. Melakukan analisis regresi linier berganda
2. Melakukan uji asumsi heterogenitas spasial dan multikolinieritas
3. Melakukan analisis menggunakan GWR, apabila terjadi heterogenitas spasial
4. Melakukan uji asumsi PCA dan analisis PCA, apabila terjadi multikolinieritas pada analisis regresi dan multikolinieritas lokal pada GWR
5. Melakukan analisis regresi PCA
6. Melakukan analisis GWRPCA
7. Pemilihan model terbaik antara model RPCA dan model GWRPCA

## 4. HASIL DAN PEMBAHASAN

### 4.1 Analisis Regresi Linier Berganda

Pada penelitian ini variabel respon PAD dihubungkan dengan 6 variabel bebas yang diduga mempengaruhinya. Model regresi linier berganda yang terbentuk sebagai berikut:

$$\hat{y} = -0,000000122 - 0,002X_1 + 0,077X_2 + 0,096X_3 + 0,317X_4 - 0,287X_5 + 0,829X_6$$



#### 4.1.1 Uji Hipotesis dalam Regresi Linier Berganda

Pada taraf signifikansi ( $\alpha$ ) sebesar 5% diketahui pada pengujian signifikansi regresi, keputusan menolak  $H_0$  karena nilai  $F_{hitung} = 401,968 > F_{(0,05;6;63)} = 2,246$  dan nilai signifikansi =  $0,000 < 0,05$ . Sehingga disimpulkan minimal ada satu  $\beta$  yang menggambarkan hubungan linier antara variabel respon dengan variabel bebas. Selanjutnya pada pengujian koefisien regresi secara parsial, diketahui keputusan menolak  $H_0$  untuk  $X_2, X_3, X_4, X_5$ , dan  $X_6$  karena nilai  $|t_{hitung}| > |t_{(0,05;63)}| = |-1,6694|$  dan nilai signifikansi  $< 0,1$ . Sedangkan keputusan menerima  $H_0$  untuk  $X_1$  karena nilai  $|t_{hitung}| < |t_{(0,05;63)}| = |-1,6694|$  dan nilai signifikansi  $> 0,1$ .

**Tabel 1.** Uji Koefisien Regresi

Variabel Bebas	Koefisien	$t_{hitung}$	Sig	Keputusan
$X_1$	-0,002	-0,053	0,958	Menerima $H_0$
$X_2$	0,077	1,834	0,071	Menolak $H_0$
$X_3$	0,096	1,768	0,082	Menolak $H_0$
$X_4$	0,317	2,071	0,043	Menolak $H_0$
$X_5$	-0,287	-1,986	0,051	Menolak $H_0$
$X_6$	0,829	19,291	0,000	Menolak $H_0$

Sehingga disimpulkan retribusi daerah ( $X_2$ ), belanja daerah ( $X_3$ ), PDRB atas dasar harga konstan ( $X_4$ ), PDRB atas dasar harga berlaku ( $X_5$ ), dan pajak daerah ( $X_6$ ) berpengaruh secara signifikan terhadap PAD.

#### 4.1.2 Uji Asumsi pada Analisis Regresi

Pada taraf signifikansi ( $\alpha$ ) sebesar 10% keputusan menolak  $H_0$  karena nilai  $BP = 12,297 > \chi^2_{(0,01;6)} = 10,644$  dan  $p\text{-value} = 0,056 < 0,1$  sehingga menunjukkan terdapat heterogenitas spasial. Untuk mengatasi masalah heterogenitas spasial dalam regresi, maka digunakan metode GWR.

Kemudian nilai VIF untuk masing-masing variabel bebas dapat dilihat pada Tabel 2.

**Tabel 2.** Nilai VIF

Variabel Bebas	Nilai VIF
$X_1$	5,180
$X_2$	4,334
$X_3$	7,356
$X_4$	57,869
$X_5$	51,754
$X_6$	4,665

Tabel 2 menunjukkan variabel bebas yang mempunyai nilai VIF  $> 10$  yaitu  $X_4$  (PDRB atas harga konstan) dan  $X_5$  (PDRB atas harga berlaku). Artinya pada model regresi terdapat pelanggaran asumsi yang terjadi yaitu adanya multikolinieritas pada variabel  $X_4$  dan  $X_5$ . Untuk menghilangkan multikolinieritas pada model regresi digunakan PCA.

## 4.2 Pemodelan Menggunakan GWR

Langkah awal dalam analisis model GWR adalah menentukan lokasi setiap sampel yang akan digunakan yaitu letak geografis kabupaten dan kota di Jawa Tengah. Selanjutnya menghitung jarak *Euclidean* menggunakan letak geografis tersebut. Kemudian menentukan *bandwidth* optimum menggunakan CV untuk memperoleh matriks pembobot. Berikut adalah *bandwidth* optimum yang diperoleh dari fungsi kernel *fixed* dan *bisquare*.

**Tabel 3.** Pemilihan Fungsi Kernel Berdasarkan AIC dan R<sup>2</sup>

Fungsi Kernel	AIC	R <sup>2</sup>
Adaptive Gaussian	-76,46895	0,983693
<b>Fixed Gaussian</b>	<b>-181,1328</b>	<b>0,997644</b>
Adaptive Bisquare	-135,7016	0,994644
Fixed Bisquare	-87,89393	0,986605
Adaptive Exponential	-87,94297	0,986606
Fixed Exponential	-160,9836	0,996450
Adaptive Tricube	-132,9037	0,994385
Fixed Tricube	-84,81082	0,985813

Berdasarkan Tabel 3, dipilih fungsi kernel *fixed gaussian* pada model GWR sebagai fungsi pembobot karena fungsi kernel *fixed gaussian* pada penelitian ini memiliki nilai AIC terkecil dan nilai R<sup>2</sup> terbesar diantara fungsi kernel lainnya. Nilai *bandwidth* optimum pada fungsi kernel *fixed gaussian* yaitu 0,2054175 dengan CV minimum sebesar 1,559546. Setelah mendapatkan matriks pembobot untuk setiap lokasi pengamatan, selanjutnya melakukan estimasi parameter model GWR. Contoh model GWR pada Kota Tegal adalah sebagai berikut:

$$\hat{y} = 0,39679 - 0,37331Z_1 + 0,0822Z_2 + 0,23168Z_3 + 0,94167Z_4 - 0,19316Z_5 + 0,52972Z_6$$

#### 4.3 Pengujian Multikolinieritas Lokal pada Model GWR

Setiap lokasi pengamatan di kabupaten dan kota di Jawa Tengah memiliki dua variabel atau lebih yang nilai VIF > 10. Sehingga pada model GWR mengenai PAD di Jawa Tengah terjadi multikolinieritas lokal. Untuk menghilangkan multikolinieritas lokal pada model GWR, maka dilakukan reduksi terhadap variabel bebas yang mempengaruhi variabel respon dengan PCA.

#### 4.4 Uji Asumsi *Principal Component Analysis*

Pada taraf signifikansi ( $\alpha$ ) sebesar 5% keputusan menerima H<sub>0</sub> karena nilai KMO= 0,744 > 0,5. Sehingga disimpulkan ukuran data cukup untuk dianalisis komponen utama. Nilai MSA untuk masing-masing variabel bebas dapat dilihat pada Tabel 4.

**Tabel 4.** Nilai MSA

Variabel	Nilai MSA
Z <sub>1</sub>	0,663
Z <sub>2</sub>	0,787
Z <sub>3</sub>	0,749
Z <sub>4</sub>	0,726
Z <sub>5</sub>	0,721
Z <sub>6</sub>	0,827

Pada taraf signifikansi ( $\alpha$ ) sebesar 5% keputusan menerima H<sub>0</sub> karena nilai MSA>0,5. Sehingga disimpulkan jumlah penduduk (Z<sub>1</sub>), retribusi daerah (Z<sub>2</sub>), belanja daerah (Z<sub>3</sub>), PDRB atas harga konstan (Z<sub>4</sub>), PDRB atas harga berlaku (Z<sub>5</sub>), dan pajak daerah (Z<sub>6</sub>) memadai untuk dianalisis lebih lanjut dengan PCA.

#### 4.5 Pembentukan Komponen Utama

Variabel X<sub>1</sub>, X<sub>2</sub>, X<sub>3</sub>, X<sub>4</sub>, X<sub>5</sub>, dan X<sub>6</sub> distandarkan terlebih dahulu sehingga terbentuk variabel Z<sub>1</sub>, Z<sub>2</sub>, Z<sub>3</sub>, Z<sub>4</sub>, Z<sub>5</sub>, dan Z<sub>6</sub> yang kemudian dikombinasi linier sehingga menghasilkan variabel baru (komponen utama) yang tidak saling berkorelasi. Pembentukan komponen utama adalah sebagai berikut:

$$PC_1 = -0,337149Z_1 - 0,397705Z_2 - 0,417538Z_3 - 0,447419Z_4 - 0,440097Z_5 - 0,39994Z_6$$



$$\begin{aligned}
PC_2 &= -0,686098Z_1 + 0,369857Z_2 - 0,417575Z_3 + 0,111269Z_4 + 0,062957Z_5 + 0,45087Z_6 \\
PC_3 &= -0,164978Z_1 - 0,508610Z_2 - 0,256359Z_3 + 0,499554Z_4 + 0,56946Z_5 - 0,273015Z_6 \\
PC_4 &= 0,130942Z_1 - 0,636635Z_2 - 0,012633Z_3 - 0,104559Z_4 - 0,0861765Z_5 + 0,747684Z_6 \\
PC_5 &= 0,608529Z_1 + 0,197927Z_2 - 0,765859Z_3 - 0,000861Z_4 + 0,034397Z_5 + 0,052859Z_6 \\
PC_6 &= 0,030479Z_1 - 0,043272Z_2 - 0,019934Z_3 + 0,725911Z_4 - 0,685168Z_5 - 0,019977Z_6
\end{aligned}$$

PC<sub>1</sub> mampu menerangkan total varian data 71,4%. *Scree plot* menunjukkan PC<sub>1</sub> memiliki nilai eigen > 1. Berdasarkan dua alasan tersebut maka PC<sub>1</sub> dipilih untuk menggantikan variabel-variabel yang berpengaruh terhadap PAD di Jawa Tengah.

#### 4.6 Pemodelan Menggunakan RPCA

Setelah didapatkan variabel PC<sub>1</sub>, maka dilakukan regresi terhadap variabel respon PAD sehingga diperoleh model RPCA sebagai berikut:

$$\hat{y} = 0,00000005377 - 0,4166PC_1$$

Pada uji signifikansi RPCA dengan taraf signifikansi ( $\alpha$ ) sebesar 5% keputusan menolak H<sub>0</sub> karena nilai  $F_{hitung} = 197 > F_{(0,05;1;68)} = 3,981$  dan  $p\text{-value} = 0,000 < 0,05$ . Sehingga disimpulkan ada hubungan antara variabel respon PAD dengan PC<sub>1</sub> secara bersama-sama. Sedangkan pada uji koefisien RPCA secara individu, keputusan menolak H<sub>0</sub> karena nilai  $|t_{hitung}| = |-14,04| > t_{(0,025;68)} = 1,6675$  dan  $p\text{-value} = 0,000 < 0,05$ . Sehingga disimpulkan terdapat pengaruh antara variabel respon PAD dengan PC<sub>1</sub> secara individu.

#### 4.7 Pemodelan Menggunakan GWRPCA

Setelah didapatkan PC<sub>1</sub>, maka dilakukan uji heterogenitas spasial untuk mengetahui adanya perbedaan karakteristik satu wilayah dengan wilayah lainnya menggunakan uji *Breusch-Pagan* (BP). Pada taraf signifikansi ( $\alpha$ ) sebesar 5% keputusan menolak H<sub>0</sub> karena nilai  $BP = 58,335 > \chi^2_{(0,05;1)} = 3,841$  dan  $p\text{-value} = 0,000 < 0,05$ , sehingga menunjukkan terdapat heterogenitas spasial. Untuk mengatasi masalah heterogenitas spasial dalam model RPCA, maka digunakan metode GWRPCA.

Matriks pembobot yang digunakan yaitu pembobot fungsi kernel *fixed gaussian* dengan *bandwidth* optimum sebesar 0,08321242 dan CV minimum 3,009035. Selanjutnya menghitung matriks pembobot setiap lokasi pengamatan untuk mengestimasi parameter model GWRPCA. Contoh model GWRPCA pada Kota Tegal sebagai berikut:

$$\hat{y} = -0,11779 - 0,08812PC_1$$

Jika model tersebut dibawa ke dalam model variabel yang distandarkan, maka akan menjadi model GWRPCA sebagai berikut:

$$\hat{y} = -0,11779 + 0,0297Z_1 + 0,03506Z_2 + 0,03681Z_3 + 0,03944Z_4 + 0,0388Z_5 + 0,03526Z_6$$

##### 4.7.1 Pengujian Kesesuaian Model GWRPCA

Pada taraf signifikansi ( $\alpha$ ) sebesar 5% keputusan menolak H<sub>0</sub> karena nilai  $F_1 = 36,0441 \geq F_{(0,05;68;22,044)} = 1,8776$  dan  $p\text{-value} = 0,000 < 0,05$ . Sehingga disimpulkan ada perbedaan antara model RPCA dengan GWRPCA. Oleh karena itu model GWRPCA lebih layak untuk menggambarkan jumlah PAD di Jawa Tengah.

##### 4.7.2 Pengujian Parsial Parameter Model GWRPCA

Pada taraf signifikansi ( $\alpha$ ) sebesar 5%, nilai  $F_3$  untuk *intercept* = 10,541 dan untuk PC<sub>1</sub> = 2,223. Keputusan menolak H<sub>0</sub> karena  $p\text{-value}$  untuk *intercept* = 0,000 dan PC<sub>1</sub> = 0,032 < 0,05. Sehingga disimpulkan ada perbedaan pengaruh yang signifikan dari variabel PC<sub>1</sub> antar lokasi amatan. Sebagai contoh pada lokasi Kota Tegal, diketahui nilai  $|t_{hitung}|$  untuk *intercept* = |-1,90455| dan PC<sub>1</sub> = 1,96952. Keputusan adalah menolak H<sub>0</sub> karena  $p\text{-value}$  untuk *intercept* = 0,031 dan PC<sub>1</sub> = 0,027 < 0,05. Sehingga disimpulkan variabel PC<sub>1</sub> pada lokasi Kota Tegal signifikan mempengaruhi variabel respon PAD di Jawa Tengah.

#### 4.8 Pemilihan Model Terbaik

Model GWRPCA dengan fungsi kernel *fixed gaussian* lebih baik digunakan untuk PAD di Jawa Tengah karena mempunyai nilai  $R^2$  terbesar dan AIC terkecil.

**Tabel 5.** Pemilihan Model Terbaik

Model	$R^2$	AIC	Model	$R^2$	AIC
Regresi PCA	0,7434	108,4194	GWRPCA	0,9929	-105,3498

Koefisien determinasi global GWRPCA pada pemodelan PAD di Jawa Tengah yaitu 0,9929. Artinya 99,29% variasi total dalam PAD dijelaskan oleh jumlah penduduk, retribusi daerah, belanja daerah, PDRB atas harga konstan, PDRB atas harga berlaku, dan pajak daerah. Sisanya 0,0071% dijelaskan oleh faktor lain. Koefisien determinasi lokal terbesar terdapat pada Kota Semarang dan koefisien determinasi lokal terkecil terdapat pada Kabupaten Tegal.

#### 5. PENUTUP

Berdasarkan hasil dan pembahasan bab sebelumnya, diperoleh kesimpulan yaitu:

1. Pada pemodelan PAD di Jawa Tengah terjadi multikolinieritas pada model regresi dan multikolinieritas lokal pada model GWR.
2. Variabel-variabel yang diduga berpengaruh terhadap PAD di Jawa Tengah yaitu jumlah penduduk, retribusi daerah, belanja daerah, PDRB atas Harga Konstan, PDRB atas harga berlaku, dan pajak daerah dapat diwakili oleh  $PC_1$  dengan total varian data 71,4%.
3. Model GWRPCA dengan pembobot fungsi kernel *fixed gaussian* lebih baik digunakan untuk PAD di Jawa Tengah karena mempunyai nilai  $R^2$  terbesar dan AIC terkecil.

#### DAFTAR PUSTAKA

- [1] Fotheringham, A. S., Brundson, C. and Charlthorn, M.. 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, UK.
- [2] Gujarati, D. N. 2006. *Dasar-dasar Ekonometri*. Edisi 3 Jilid 2. Jakarta: Erlangga.
- [3] Johnson, R. A. and Winchen, D. W. 2007. *Applied Multivariate Statistical Analysis*. Sixth Edition. New Jersey: Prentice Hall.
- [4] Kawedar, W. 2008. *Akuntansi Sektor Publik: Pendekatan Penganggaran Daerah dan Akuntansi Keuangan Daerah*. Buku 1. Semarang: Badan Penerbit UNDIP dan Widya Karya.
- [5] Montgomery, D. C. 2001. *Designs and Analysis of Experiments*. Canada: John Wiley and Sons.
- [6] Pratnyaningrum, N., Yasin, H., dan Hoyyi, A. 2015. *Pemodelan Persentase Balita Gizi Buruk di Jawa Tengah dengan Pendekatan Geographically Weighted Regression Principal Components Analysis (GWRPCA)*. Jurnal Gaussian Vol.4, No.2, Hal.171-180.
- [7] Purhadi, dan Yasin, H. 2012. *Mixed Geographically Weighted Regression Model Case Study: The Percentage of Poor Households in Mojokerto 2008*. European Journal of Scientific Research, Vol. 69, Issue 2, Hal. 188-196.
- [8] Simamora, B. 2005. *Analisis Multivariat Pemasaran*. Jakarta: PT Gramedia Pustaka Utama.
- [9] Tazliqoh, A. Z., Rahmawati, R., dan Safitri, D. 2015. *Perbandingan Regresi Komponen Utama dengan Regresi Ridge pada Analisis Faktor-Faktor Pendapatan Asli Daerah (PAD) Provinsi Jawa Tengah*. Jurnal Gaussian Vol. 4, No. 1, Hal. 1-10.