ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 5, Nomor 3, Tahun 2016, Halaman 383-393
Online di: http://ejournal-s1.undip.ac.id/index.php/gaussian

PEMODELAN PENDAPATAN ASLI DAERAH (PAD) DI KABUPATEN DAN KOTA DI JAWA TENGAH MENGGUNAKAN *GEOGRAPHICALLY WEIGHTED RIDGE REGRESSION*

Depy Veronica[1], Hasbi Yasin[2], Tatik Widiharih[3]
[1]Mahasiswa Jurusan Statistika FSM Universitas Diponegoro
[2,3]Staff Pengajar Jurusan Statistika FSM Universitas Diponegoro

ABSTRACT

Linear Regression Analysis is a statistical method for modeling the relation between response variable and predictor variable. Geographically Weighted Regression (GWR) is an expansion of linear regression model if spatial heterogeneity occurred. Local multicollinearity test is required to know the presence of linear correlation between independent variables for each observation location. Geographically Weighted Ridge Regression (GWRR) is a extension of GWR model to solve local multicollinearity problem. Parameter estimation for GWR and GWRR model is done using Weighted Least Square (WLS) method by applying optimum bandwith with Cross Validation (CV) criteria. GWRR model is applied on locally generated recurring revenues (PAD) at district and city of Central Java and its result shows the ability of GWRR model to erase multicollinearity problem. Based on Mean Squared Error (MSE) and Akaike Information Criterion (AIC) value for GWR and GWRR model, it is know that the best model to analyze locally generated recurring revenues (PAD) at district and city of Central Java is GWRR model with the smallest MSE and AIC value.

Keywords : Akaike Information Crietion, Spasial Heterogeneity, Geographically Weighted Ridge Regression, Mean Square Error, Local Multicoliniearity.

1. PENDAHULUAN

Pendapatan Asli Daerah (PAD) secara hukum berdasarkan Undang-Undang Nomor 33 Tahun 2004 tentang Perimbangan Keuangan Antara Pusat dan Daerah Pasal 1 angka 18 bahwa “Pendapatan asli daerah, selanjutnya disebut PAD adalah pendapatan yang diperoleh daerah yang dipungut berdasarkan peraturan daerah sesuai dengan peraturan perundang-undangan. Sumber pendapatan daerah antara lain pajak daerah, retribusi daerah, lain-lain PAD yang sah dan pengelolaan kekayaan daerah yang dipisahkan. Pendapatan asli daerah dapat dipengaruhi oleh faktor-faktor lain seperti pengeluaran pemerintah, jumlah penduduk dan Pendapatan Daerah Regional Bruto (PDRB) [7]. Berdasarkan faktor-faktor tersebut perlu dilakukan analisis faktor mana yang berpengaruh secara signifikan terhadap PAD. Faktor yang diduga mempengaruhi PAD dalam penelitian ini adalah pajak daerah, retribusi, PDRB atas dasar harga berlaku, PDRB atas dasar harga konstan, jumlah penduduk dan belanja daerah.

Dalam menganalisis pendapatan asli daerah di Provinsi Jawa Tengah yang memiliki karakteristik sumber daya dan potensi yang berbeda-beda antar tiap daerah perlu diperhatikan terhadap letak geografis dan faktor lokasi pengamatan. Metode yang dapat digunakan adalah *Geographically Weighted Regression* (GWR) [2].

Selain faktor perbedaan geografis, pendapatan asli daerah juga melibatkan beberapa faktor-faktor yang mempengaruhinya sehingga memungkinkan terjadinya permasalahan multikolinieritas antara variabel prediktor. Salah satu metode yang efektif untuk mengatasi masalah multikolinieritas adalah regresi ridge[5].

Geographically Weighted Ridge Regression (GWRR) merupakan pengembangan dari model GWR untuk menangani adanya kasus multikolinieritas pada data spasial. Dalam penelitian ini, akan dilakukan analisis dalam pemodelan dengan menggunakan metode GWRR pada kasus Pendapatan Asli Daerah (PAD) di Provinsi Jawa Tengah.

2. TINJAUAN PUSTAKA

2.1 Pendapatan Asli Daerah (PAD)

Pendapatan asli daerah merupakan fakor utama dalam pembiayaan daerah, oleh karena itu kemampuan melaksanakan ekonomi diukur dari besarnya kontribusi yang diberikan pendapatan asli daerah terhadap APBD. Semakin tinggi kemampuan daerah dalam menghasilkan PAD, maka semakin besar pula diskresi daerah untuk menggunakan PAD tersebut sesuai dengan aspirasi, kebutuhan, dan prioritas pembangunan daerah [4].

2.2 Faktor-Faktor yang Mempengaruhi Pendapatan Asli Daerah

Pendapatan Asli Daerah dapat dipengaruhi oleh faktor-faktor lain seperti pengeluaran pemerintah, jumlah penduduk dan Pendapatan Daerah Regional Bruto (PDRB) [7]. (i) Pajak daerah adalah pajak negara yang diserahkan kepada daerah untuk dipungut berdasarkan peraturan perundangan-undangan, yang dipergunakan guna membiayai pengeluaran daerah sebagai badan hukum publik[9]. (ii) Retribusi daerah adalah pembayaran kepada negara yang dilakukan kepada mereka yang menggunakan jasa-jasa negara, artinya retribusi daerah sebagai pembayaran atas jasa atau karena mendapat pekerjaan usaha atau milik daerah bagi yang berkepentingan, atau jasa yang diberikan oleh daerah bagi secara langsung maupun tidak langsung[9]. (iii) Belanja Daerah, dengan adanya perbelanjaan tersebut maka akan meningkatkan pengeluaran agregat dan meningkatkan kegiatan ekonomi yang akan berpengaruh terhadap meningkatnya aliran penerimaan pemerintah melalui PAD. (iv) Produk Domestik Regional Brito (PDRB), Dalam perhitungn PDRB dapat dilakukan dengan dua cara yaitu PDRB atas dasar harga berlaku dan PDRB berdasarkan harga konstan. Meningkatnya PDRB maka akan menambah penerimaan pemerintah daerah untuk membiayai program-program pembangunan daerah[7]. (v) Jumlah Penduduk diduga dapat mempengaruhi PAD apabila pertumbuhan penduduk yang tinggi dan diiringi dengan perubahan teknologi akan mendorong tabungan dan juga penggunaan skala ekonomi di dalam produksi[7].

2.3 Regresi Linier Berganda

Metode regresi linier merupakan metode yang memodelkan hubungan antara variabel respon (y) dan variabel prediktor (x). Secara umum model regresi linier dapat dituliskan sebagai berikut [1]:

$$y_i = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ik} + \varepsilon_i \quad ; i = 1,2,\ldots,\mathrm{n}\,; \ \mathrm{k} = 1,2,\ldots,\mathrm{p} \tag{1}$$

Dengan: y_i : Variabel respon pengamatan ke-i
x_{ik} : Variabel bebas ke-k pengamatan ke-i
β_k : Koefisien regresi pada x_k
β_0 : Intersep
ε_i : *Error* random ke-i

Jika ditulis dalam bentuk matriks, persamaan (1) dapat ditulis sebagai berikut:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{2}$$

Metode yang digunakan untuk mengestimasi parameter model regresi adalah dengan meminimumkan jumlah kuadrat error atau yang sering dikenal dengan *Ordinary Least Square* (OLS) [1]. Berdasarkan persamaan (2), dengan metode kuadrat terkecil diperoleh estimasi β, yaitu sebagai berikut :

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

2.4 Uji Heterogenitas Spaial

Heterogenitas spasial terjadi akibat adanya perbedaan karakterisik satu wilayah dengan wilayah lainnya. Pengujian heterogenitas spasial pada model regresi linier berganda dilakukan dengan menggunakan statistik uji *Breusch-Pagan Test*[6].

Hipotesis :

$\mathrm{H_0}$: $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2$ (homoskedastisitas)

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

Taraf signifikansi (α)

Statistik uji :

$$\mathrm{BP} = \frac{1}{2}\mathbf{f}^{\mathbf{T}}\mathbf{Z}(\mathbf{Z}^{\mathbf{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathbf{T}}\mathbf{f} \sim \chi^2(\mathrm{p}) \quad (4)$$

$f_i = (\frac{e_i^2}{\sigma^2} - 1)$, e_i merupakan *least square residual* untuk observasi ke-i. $\mathbf{Z}$ = matriks berukuran n x (p+1) berisi vektor yang sudah dinormal-standarkan untuk setiap observasi, dengan n adalah banyaknya data dan p adalah banyaknya variabel prediktor.Tolak $\mathrm{H_0}$ jika $BP > \chi^2(p)$ Apabila $\mathrm{H_0}$ ditolak perlu diuji dengan menggunakan model GWR.

2.5 *Geographically Weighted Regression*

Geographically Weighted Regression (GWR) adalah pengembangan dari model regresi linier OLS menjadi model regresi terboboti yang memperhatikan efek spasial[2]. Penggunaan metode GWR adalah menentukan model regresi untuk masing-masing titik lokasi $(\mathrm{u_i}, \mathrm{v_i})$ sehingga model-model regresi akan bersifat unik, yaitu model regresi untuk titik yang satu berbeda dengan titik yang lainnya. Secara sistematis model GWR dapat ditulis sebagai berikut :

$$\mathrm{y_i} = \beta_0(\mathrm{u_i}, \mathrm{v_i}) + \sum_{m=1}^{p} \beta_m(\mathrm{u_i}, \mathrm{v_i})\mathrm{x_{im}} + \varepsilon_i \quad (5)$$

Estimasi parameter model GWR adalah dengan menggunakan metode WLS dengan menambahkan pembobot w_{ij} pada persamaan (5) dan meminimumkan jumlah kuadrat *error*-nya. Sehingga bentuk penaksir parameter dari model GWR untuk setiap lokasi[2] adalah :

$$\widehat{\boldsymbol{\beta}}(\mathrm{u_i}, \mathrm{v_i}) = [\mathbf{X}^{\mathbf{T}}\mathbf{W}(\mathrm{u_i}, \mathrm{v_i})\mathbf{X}]^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{W}(\mathrm{u_i}, \mathrm{v_i})\mathbf{Y}$$

Dengan $\mathbf{W}(u_i, v_i)$ merupakan matrik diagonal (n x n) dengan setiap elemen diagonalnya adalah pembobot untuk masing-masing titik lokasi pengamatan (u_i, v_i) yang ditentukan dengan menggunakan fungsi Kernel. Fungsi kernel yang digunakan adalah *Fixed exponential*[2].

$$w_{ij} = \exp\left[-(d_{ij}/h)\right] \quad (6)$$

Nilai h merupakan *bandwidth* yang memiliki nilai *fixed* atau sama untuk setiap lokasi dan d_{ij} adalah jarah *euclidean* antara titik lokasi pengamatan ke-i dengan titik lokasi pengamatan ke-j.

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2}$$

Pada penelitian ini digunakan metode *Cross Validation* (CV) untuk menentukan nilai *bandwidth* yang optimum dengan rumus sebagai berikut[10]:

$$\mathrm{CV} = \sum_{i=1}^{n}[y_i - \hat{y}_{\neq i}(h)]^2 \quad (7)$$

dengan $\hat{y}_{\neq i}(h)$ adalah nilai penaksir y_i dimana pengamatan di lokasi i dihilangkan dari proses penaksiran. Untuk mendapatkan nilai h yang optimum maka diambil dari nilai yang memiliki CV minimum. Percobaan untuk mendapatkan nilai *bandwidth* yang optimum digunakan algoritma *golden section search* dan nilai h berada diantara d_{ij} yang terkecil dan d_{ij} yang terbesar.

Pengujian kesesuaian model (*goodness of fit*) diperlukan untuk mengetahui apakah model GWR lebih sesuai digunakan jika dibandingkan dengan model regresi berganda. Pengujian ini dilakukan dengan hipotesis sebagai beriku[3] :

$\mathrm{H_0}:\ \beta_m(u_i, v_i) = \beta_m$ (tidak ada perbedaan yang signifikan antara model Regresi dengan model GWR)

$\mathrm{H_1}$: Paling sedikit ada satu $\beta_m(u_i, v_i) \neq \beta_m$; $m = 1,2,3,\ldots,p$ (ada perbedaan yang signifikan antara model Regresi dengan model GWR)

Statistik Uji :

$$F_2 = \frac{\mathrm{SSE}(\mathrm{H_0}) - \mathrm{SSE}(\mathrm{H_1})/\tau_1}{\mathrm{SSE}(\mathrm{H_0})/(n-p-1)}$$

dengan

$\mathrm{SSE}(\mathrm{H_0}) = \mathbf{Y^T}(\mathbf{I}-\mathbf{H})\mathbf{Y}$ dimana $\mathbf{H} = \mathbf{X}(\mathbf{X^T X})^{-1}\mathbf{X^T}$; $\mathrm{SSE}(\mathrm{H_1}) = \mathbf{Y^T}(\mathbf{I}-\mathbf{L})^{\mathrm{T}}(\mathbf{I}-\mathbf{L})\mathbf{Y}$

$df_1 = \frac{\tau_1^2}{\tau_2}$ dimana $\tau_i = \mathrm{tr}([(\mathbf{I}-\mathbf{H}) - (\mathbf{I}-\mathbf{L})^{\mathrm{T}}(\mathbf{I}-\mathbf{L})]^{\mathrm{i}}), i = 1,2$

$\mathbf{L}=[\mathrm{x_i^T}[\mathbf{X^T W}(u_i, v_i)\mathbf{X}]^{-1}\mathbf{X^T W}(u_i, v_i)]$ dan $df_2 = n - p - 1$

H_0 ditolak apabila $F_2 > F_{\alpha,df_1,df_2}$.

Uji hipotesis yang selanjutnya dilakukan apabila diperoleh kesimpulan bahwa model GWR lebih baik dibandingkan dengan model regresi linier berganda adalah uji pengaruh lokasi secara parsial untuk megetahui apakah parameter bersifat lokal atau bersifat global[6]. Pengujian ini dilakukan dengan Hipotesis sebgai berikut :

$\mathrm{H_0}:\ \beta_{1m}(u_1, v_1) = \beta_{2m}(u_2, v_2) = \cdots = \beta_{nm}(u_n, v_n)$ untuk m = 0,1,2,…,p dan i = 1,2,…,n (tidak ada perbedaan pengaruh yang signifikan dari variabel prediktor $\mathrm{X_m}$ antara satu lokasi dengan lokasi lainnya.

$\mathrm{H_1}$: Paling sedikit ada satu $\beta_{im}(u_i, v_i) \neq 0$ (ada perbedaan pengaruh yang signifikan dari variabel prediktor $\mathrm{X_m}$ antara satu lokasi dengan lokasi lainnya.

Statistik Uji:

$$F_3 = \frac{V_m^2/tr\left(\frac{1}{n}\boldsymbol{B}_m^T\left[\boldsymbol{I}-\frac{1}{n}\boldsymbol{J}\right]\boldsymbol{B}_m\right)}{SSE(H_1)/\delta_1}$$

dengan:

$\boldsymbol{B}_m = \boldsymbol{e}_m^T[\boldsymbol{X}^T\boldsymbol{W}(u_i, v_i)\boldsymbol{X}]^{-1}\boldsymbol{X}^T\boldsymbol{W}(u_i, v_i)$; $\boldsymbol{\beta}_m(u_i, v_i) = [\hat{\beta}_1(u_i, v_i) \ \ldots \hat{\beta}_P(u_i, v_i)]^T$

$$V_m^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\beta}_m(u_i, v_i) - \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_m(u_i, v_i)\right)^2 = \frac{1}{n}\boldsymbol{\beta}_m^T\left[\mathbf{I}-\frac{1}{n}\boldsymbol{J}\right]\boldsymbol{\beta}_m$$

J adalah matriks berukuran n x n yang semua elemennya adalah 1 dan e_m adalah vektor kolom berukuran (p+1) yang bernilai satu untuk elemen ke-m dan nol untuk lainnya.

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

Uji parsial signifikansi parameter model dilakukan untuk mengetahui parameter mana saja yang berpengaruh terhadap variabel responnya untuk setiap lokasi pengamatan. Pengujian hipotesis sebagai berikut[6] :

$\mathrm{H_0}:\ \beta_m(u_i, v_i) = 0$

$\mathrm{H_1}$: Paling sedikit ada satu $\beta_m(u_i, v_i) \neq 0$ dimana m=1,2,…,p dan i=1,2,…,n

sehingga statistik uji yang digunakan adalah :

$$T = \frac{\hat{\beta}_m(u_i, v_i)}{\hat{\sigma}\sqrt{c_{mm}}}$$

Dengan c_{mm} adalah elemen diagonal ke-m dari matrik $\mathbf{C_i C_i^T}$ yaitu $\mathbf{C} = (\mathbf{X}^{\mathrm{T}}\mathbf{W}(\mathrm{u_i}, \mathrm{v_i})\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{W}(\mathrm{u_i}, \mathrm{v_i})$. H_0 ditolak apabila $|T_{hit}| \geq t_{\frac{\alpha}{2};df}$. dengan derajat bebas $df = \frac{\delta_1^2}{\delta_2}$ dan $\hat{\sigma} = \frac{RSS(H_1)}{\delta_1}$.

Dikarenakan dalam penelitian ini digunakan banyak variabel sehingga diduga bahwa terdapat hubungan linier antar variabel bebasnya maka perlu dilakukan pengujian mltikolinieritas. untuk mendeteksi adanya multikolinieritas dapat diidetifikasi melalui nilai *Variance Inflantion Factor* (VIF)[10]. VIF untuk model GWR adalah sebagai berikut :

$$VIF = \frac{1}{1 - R_m^2(u_i, v_i)}$$

Dengan $R_m^2(u_i, v_i)$ adalah koefisien determinasi antara variabel prediktor x_m dengan variabel prediktor lainnya untuk setiap lokasi (u_i, v_i). Sedangkan untuk mencari VIF pada model GWRR yang sudah dilakukan *centering* dan *scaling* seperti pada persamaan 8, maka nilai VIF merupakan nilai dari elemen diagonal utama *inverse* dari $[\mathbf{X}^{*\mathbf{T}}\mathbf{W}\mathbf{X}^* + \mathrm{k}\mathbf{I}]$ dengan k merupakan koefisien ridge optimum[8]**.** Adanya multikolinieritas dilihat dari nilai VIF yang lebih besar dari 10. Apabila terjadi multikolinieritas pada model GWR, maka analisis dapat dilanjutkan dengan menggunakan model GWRR.

2.6 *Geographically Weighted Ridge Regression*

Langkah awal dalam pemodelan GWRR adalah melakukan *centering* dan *scaling* yang menghasilkan matriks **X*** yang berukuran n x p dan **y*** adalah vektor variabel respon yang berukuran n x 1[8].

$$\mathbf{X}^* = \begin{bmatrix} \left(\frac{x_{11}-\bar{x}_1}{S_1\sqrt{n-1}}\right) & \left(\frac{x_{21}-\bar{x}_2}{S_2\sqrt{n-1}}\right) & \cdots & \left(\frac{x_{p1}-\bar{x}_p}{S_p\sqrt{n-1}}\right) \\ \left(\frac{x_{12}-\bar{x}_1}{S_1\sqrt{n-1}}\right) & \left(\frac{x_{22}-\bar{x}_2}{S_2\sqrt{n-1}}\right) & \cdots & \left(\frac{x_{p2}-\bar{x}_p}{S_p\sqrt{n-1}}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \left(\frac{x_{1n}-\bar{x}_1}{S_1\sqrt{n-1}}\right) & \left(\frac{x_{2n}-\bar{x}_2}{S_2\sqrt{n-1}}\right) & \cdots & \left(\frac{x_{pn}-\bar{x}_p}{S_p\sqrt{n-1}}\right) \end{bmatrix}; \ \boldsymbol{y}^* = \begin{bmatrix} \left(\frac{y_1-\bar{y}}{S_y\sqrt{n-1}}\right) \\ \left(\frac{y_2-\bar{y}}{S_y\sqrt{n-1}}\right) \\ \vdots \\ \left(\frac{y_n-\bar{y}}{S_y\sqrt{n-1}}\right) \end{bmatrix} \quad (8)$$

dimana $\bar{\mathrm{x}}_{\mathrm{m}} = \frac{\sum_{i=1}^{n} x_{mi}}{n}$ yang merupakan rata-rata dari masing-masing variabel prediktor ke-m, dengan m = 1, 2, …, p dan i =1, 2, …, n

$$\mathrm{S_m} = \sqrt{\sum_{i=1}^{n} \frac{(x_{mi}-\bar{x}_m)^2}{n-1}}; \qquad \mathrm{S_y} = \sqrt{\sum_{i=1}^{n} \frac{(y_i-\bar{y})^2}{n-1}}$$

$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$ yang merupakan rata-rata dari masing-masing variabel respon ke-i, dengan i =1, 2, …, n

Estimasi parameter untuk metode GWRR adalah :

$$\hat{\boldsymbol{\beta}}(\mathrm{u_i}, \mathrm{v_i}, \mathrm{k}) = (\mathbf{X}^{*\mathbf{T}}\mathbf{W}(\mathrm{u_i}, \mathrm{v_i})\mathbf{X}^* + \mathrm{k}\mathbf{I})^{-1}\mathbf{X}^{*\mathbf{T}}\mathbf{W}(\mathrm{u_i}, \mathrm{v_i})\mathbf{y}^*$$

Dalam menentukan matriks pembobot digunakan nilai bandwidth optimum dengan kriteria CV seperti pada persamaan (7) dan koefisien k yaitu nilai yang terletak antara 0<k<1 yang diperoleh berdasarkan kriteria CV[10].

Pengujian kesesuaian model (*goodness of fit*) dilakukan dengan membandingkan model GWRR dengan model GWR dengan hipotesis adalah sebagai berikut[8] :

$\mathrm{H_0}$: $\beta_m(\mathrm{u_i}, \mathrm{v_i}, \mathrm{k}) = \beta_m(\mathrm{u_i}, \mathrm{v_i})$ (tidak ada perbedaan yang signifikan antara model GWRR dengan model GWR)

$\mathrm{H_1}$: Paling sedikit ada satu $\beta_m(\mathrm{u_i}, \mathrm{v_i}, \mathrm{k}) \neq \beta_m(\mathrm{u_i}, \mathrm{v_i})$; untuk m=1,2,…p dan i=1,2,…,n (ada perbedaan yang signifikan antara model GWRR dengan model GWR)

Sehingga diperoleh statistik uji sebagai berikut :

$F(1) = \frac{\mathbf{y}^{*\mathbf{T}}[(\mathbf{I}-\mathbf{S})^{\mathrm{T}}(\mathbf{I}-\mathbf{S})-(\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})]\mathbf{y}^{*}/\mathrm{v}_1}{\mathbf{y}^{*\mathbf{T}}(\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})\mathbf{y}^{*}/\mathrm{u}_1}$

Untuk $\mathbf{S}=[\mathrm{x}_i^{\mathrm{T}}[\mathbf{X}^{\mathrm{T}}\mathbf{W}(\mathrm{u_i},\mathrm{v_i})\mathbf{X}]^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{W}(\mathrm{u_i},\mathrm{v_i})]$; $\mathbf{G} = [\mathrm{x}_i^{\mathrm{T}}[\mathbf{X}^{\mathrm{T}}\mathbf{W}(\mathrm{u_i},\mathrm{v_i})\mathbf{X}+\mathbf{kI}]^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{W}(\mathrm{u_i},\mathrm{v_i})]$

$\mathrm{v_i} = \mathbf{tr}\left([(\left(\mathbf{I}-\mathbf{S}\right)^{\mathrm{T}}(\mathbf{I}-\mathbf{S}) - \left(\mathbf{I}-\mathbf{G}\right)^{\mathrm{T}}(\mathbf{I}-\mathbf{G}))]^{\mathrm{i}}\right)$; $\mathrm{u_i} = \mathbf{tr}\,([(\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})]^{\mathrm{i}})$, i = 1, 2

Dengan derajat bebas $df_1 = (\frac{v_1^2}{v_2})$ dan $df_2 = (\frac{u_1^2}{u_2})$. Tolak $\mathrm{H_0}$ jika F(1) ≥ F_{α,df_1,df_2}.

Pengujian serentak dilakukan untuk mengetahui signifikansi variabel-variabel prediktor pada model GWRR secara serentak, hipotesisnya adalah sebagai berikut[8] :

$\mathrm{H_0}$: $\beta_1\,(\mathrm{u_i},\mathrm{v_i},\mathrm{k}) = \beta_2\,(\mathrm{u_i},\mathrm{v_i},\mathrm{k}) = \cdots = \beta_p\,(\mathrm{u_i},\mathrm{v_i},\mathrm{k}) = 0$

$\mathrm{H_1}$: Paling sedikit ada satu $\beta_m\,(\mathrm{u_i},\mathrm{v_i},\mathrm{k}) \neq 0$; dengan m=1,2,…,p dan i=1,2,…,n

Statistik Uji :

$F(2) = \frac{\mathbf{y}^{*\mathbf{T}}\mathbf{y}^{*}/\mathrm{n}}{\mathbf{y}^{*\mathbf{T}}(\mathbf{I}-\mathbf{G})^{\mathrm{T}}(\mathbf{I}-\mathbf{G})\mathbf{y}^{*}/\mathrm{u}_1}$

Dengan derajat bebas $df_1 = \mathrm{n}$ dan $df_2 = (\frac{u_1^2}{u_2})$. Tolak $\mathrm{H_0}$ jika F(2) ≥ F_{α,df_1,df_2}.

Pengujian parsial ditunjukan untuk mengetahui masing-masing variabel prediktor yang berpengaruh signifikan terhadap variabel respon pada model GWRR. Digunakan hipotesis sebagai berikut[8]:

$\mathrm{H_0}$: $\beta_m\,(\mathrm{u_i},\mathrm{v_i},\mathrm{k}) = 0$ (variabel lokal x_m pada lokasi ke-i tidak signifikan berpengaruh pada variabel respon)

$\mathrm{H_1}$: $\beta_m\,(\mathrm{u_i},\mathrm{v_i},\mathrm{k}) \neq 0$ (variabel lokal x_m pada lokasi ke-i signifikan berpengaruh pada variabel respon) untuk m = 1,2,…,p dan i=1,2,…n

Statistik uji :

$\mathrm{T_{hit}} = \frac{\hat{\beta}_{(\mathrm{u_i},\mathrm{v_i},\mathrm{k})}}{\hat{\sigma}\sqrt{\mathrm{C_{mm}}}}$

Dengan C_{mm} adalah elemen diagonal ke-j dari matrik $\mathbf{CC^T}$. Untuk $\mathbf{C} = (\mathbf{X}^{*\mathbf{T}}\mathbf{W}(\mathrm{u_i},\mathrm{v_i})\mathbf{X}^{*} + \mathrm{k}\mathbf{I})^{-1}\mathbf{X}^{*\mathbf{T}}\mathbf{W}(\mathrm{u_i},\mathrm{v_i})$ dan $\hat{\sigma}$ adalah akar dari $\hat{\sigma}^2 = \frac{\mathrm{SSE}}{\mathrm{n}-2\mathrm{tr}(\mathbf{G})+\mathrm{tr}(\mathbf{G^TG})}$ jika tingkat signifikansi α, maka diambil keputusan tolak H_0 jika $|T_{hit}| \geq t_{\frac{\alpha}{2},df}$ dimana df = $(\frac{u_1^2}{u_2})$.

2.7 Pemilihan Model Terbaik

Model yang dibandingkan adalahmodel regresi linier berganda, model GWR dan GWRR dengan melihat nilai MSE dan AIC yang terkecil.

MSE $= \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

AIC = $2\,n\ln(\hat{\sigma}) + n\ln(2\pi) + n\left\{\frac{n+tr(\mathbf{L})}{n-2-tr(\mathbf{L})}\right\}$

3. METODE PENELITIAN

Data yang digunakan bersifat sekunder bersumber dari BPS Jateng. Data Pendapatan Asli Daerah (PAD) dan faktor-faktor yang akan digunakan dalam penelitian,meliputi seluruh kabupaten dan kota yang ada di Jawa Tengah pada tahun 2011 - 2013.Variabel 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). Software statistik yang digunakan adalah SPSS.16, R, MATLAB, dan Microsoft Excel 2013.

Langkah-lagkah penelitian sebagai berikut :

1. Membentuk model regresi linier berganda
2. Menganalisa adanya heterogenitas spasial pada model regresi linier berganda
3. Membentuk model GWR

4. menganalisa adanya multikolinieritas lokal pada model GWR
5. Membentuk model GWRR
6. Memilih model terbaik dengan AIC dan MSE

4. HASIL DAN PEMBAHASAN

Untuk mengetahui adanya pengaruh antara variabel prediktor dan variabel respon maka pengujian yang dilakukan adalah dengan menggunakan metode regresi linier berganda dan diperoleh model sebagai berikut :

$\hat{y} = 0.101 - 0.291\,X_1 + 0.734X_2 + 0.058X_3 + 13.845X_4 - 2.761X_5 + 0.666X_6$

Dikarenakan data yang digunakan adalah data spasial, maka diduga terdapat heterogenitas spasial sehingga dilakukan pengujian terhadap varian residual dengan menggunakan *Breusch-Pagan* dan diperoleh hasil statistik BP (294.748) > $X^2_{(0.95,6)}$ (12.592). Maka dapat disimpulkan bahwa terdapat kasus heterogenitas spasial atau terdapat karakteristik yang berbeda terhadap data pendapatan asli daerah di Provinsi Jawa Tengah antar tiap Kabupaten dan Kota. Sehingga analisis selanjutnya dapat digunakan metode *Geographically Weighted Regression* (GWR) sebagai penanganan dari adanya masalah heterogenitas spasial.

4.2 *Geographically Weighted Regression*

Langkah awal untuk mendapatkan model GWR adalah menentukan nilai *bandwidth* optimum dengan menggunakan kriteria CV. Penentuan *bandwidth* menggunakan fungsi pembobot *fixed exponential kernel* dan diperoleh nilai *bandwidth* sebesar 3.064697 dengan nilai CV minimum adalah 37.97278. Selanjutnya nilai *bandwidth* yang diperoleh digunakan untuk menentukan matriks pembobot dalam melakukan estimasi parameter dengan menggunakan metode *Weighted Least Square* (WLS).

Pengujian kesesuaian model (*goodness of fit*) antara model regresi linier berganda dan GWR dan diperoleh hasil sebagai berikut :

Tabel 1. Uji Kesesuaian Model GWR

F2 Stat	df1	df2	F(2)tabel
6.1759	14.7309	98	0.46713

Berdasarkan Tabel 10, menunjukan bahwa nilai F(2) statistik sebesar 6.1759 > F(2)tabel sebesar 0.46713, dengan menggunakan taraf signifikansi α=5% maka dapat disimpulkan bahwa H_0ditolak yang artinya ada perbedaan yang signifikan antara model regresi linier berganda dengan model GWR.

Apabila dalam pengujian kesesuaian model diperoleh kesimpulan bahwa terdpat perbedaan yang signifikan antara model GWR dan regresi linier berganda maka pengujian yang selanjutnya dilakukan adalah uji pengaruh lokasi secara parsial untuk mengetahui apakah ada perbedaan pengaruh yang signifikan dari variabel prediktor X_m antara satu lokasi dengan lokasi yang lain.

Tabel 2. Uji Faktor Geografis Pada Setiap Prediktor

Variabel	F3	df1	df2	F tabel
Intersept	1.09854	27.12965	97.223	1.59961
X_1	1.37398	23.97963	97.223	1.63006
X_2	0.40011	11.40447	97.223	1.87359
X_3	1.10023	17.53375	97.223	1.71907
X_4	23.70978	15.31900	97.223	1.76323
X_5	7.90916	17.39541	97.223	1.72156
X_6	44.46392	6.45536	97.223	2.15052

Dengan taraf signifikansi (α) adalah 5%, maka dapat disimpulkan bahwa variabel prediktor X_4, X_5, X_6 ada perbedaan pengaruh yang signifikan dari variabel prediktor X_m antara satu lokasi dengan lokasi yang lainnya. Karena tidak semua variabel prediktor berpengaruh secara lokal maka untuk pengujian parsial parameter model sebaiknya dilakukan dengan menggunakan model *Mixed Geographically Weighted Regression* (MGWR), akan tetapi dalam penelitian ini membatasi masalah tersebut sehingga dalam uji parameter model tetap menggunakan model GWR dengan menganggap bahwa semua variabel prediktor bersifat lokal.

Pengujian yang dilakukan untuk mengetahui variabel yang signifikan terhadap model maka dilakukan uji parsial parameter. Sebagai contoh hasil yang diperoleh untuk Kaupaten Cilacap adalah sebagai berikut :

Tabel 3. Uji Parameter Model GWR pada Kabupaten Cilacap

Variabel	t_{hitung}	Kesimpulan
Intercept	0.90113	Tidak Signifikan
X_1	-1.77849	Tidak Signifikan
X_2	2.16513	Signifikan
X_3	2.45225	Signifikan
X_4	3.94423	Signifikan
X_5	-2.31696	Signifikan
X_6	8.74132	Signifikan

Dari Tabel 3, dapat disimpulkan bahwa variabel 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 Pendapatan Asli Daerah di Kabupaten Cilacap. Karena terdapat beberapa faktor yang mempengaruhi pendapatan asli daerah sehingga memungkinkan terjadinya masalah multikolinieritas, sehingga perlu dilakukan uji multikolinieritas terhadap model.

Adanya multikolinieritas ditunjukan dengan nilai VIF yang lebih besar dari 10. Uji multikolinieritas untuk Kabupaten Ciacap diperoleh hasil sebagai berikut :

Tabel 4. Nilai VIF Kabupaten Cilacap untuk model GWR

Nilai	X_1	X_2	X_3	X_4	X_5	X_6
VIF	3.00532	2.02225	4.06468	25.24504	19.28611	3.02776

Pada Tabel 4 variabel X_4 dan X_5 memiliki nilai VIF lebih dari 10 sehingga dapat disimpulkan terjadi kasus multikolinieritas pada model untuk kabupaten cilacap, hasil yang sama juga ditunjukan untuk lokasi lainnya seperti pada Lampiran 4 dimana nilai VIF untuk variabel X_4 dan X_5 lebih besar dari 10. Karena terdapat multikolinieritas pada model GWR, sehingga untuk analisis selanjutnya dapat dilakukan dengan menggunakan model GWRR.

4.3 *Geographically Weighted Ridge Regression*

Langkah awal untuk mendapatkan model GWRR adalah menentukan nilai *bandwidth* optimum dengan menggunakan kriteria CV. Penentuan *bandwidth* menggunakan fungsi pembobot *fixed exponential kernel* dan diperoleh nilai *bandwidth* sebesar 3.05 dengan nilai CV minimum adalah 0.98918. Selanjutnya menentukan nilai koefisien ridge yang optimum berdasarkan kriteria CV. Nilai k optimum yang dihasilkan adalah 0.15 denganni lai CV sebesar 0.93685. nilai *bandwidth* dan koefisien k digunakan dalam estimasi parameter pada model GWRR. Sebagai contoh model untuk Kabupaten Cilacap adalah sebagai berikut:

$$\hat{y} = -0.05121X_1 + 0.15952X_2 + 0.11574X_3 + 0.20960X_4 + 0.06089X_5 + 0.42789X_6$$

Pengujian kesesuaian model (*goodness of fit*) antara model GWR dan GWRR dan diperoleh hasil sebagai berikut :

Tabel 5. Uji Kesesuaian Model GWRR

F(1) statistik	df1	df2	Ftabel
3.06134	38.0208	103.4755	0.62370

Berdasarkan Tabel 5, menunjukan bahwa nilai F(1) statistik sebesar 3.06134 > Ftabel sebesar 0.62370, dengan menggunakan taraf signifikansi α=5% maka dapat disimpulkan bahwa H_0ditolak yang artinya ada perbedaan yang signifikan antara model GWRdengan model GWRR. Pengujian selanjutnya yang dilakukan adalah pengujian serentak parameter model dan diperoleh hasil sebagai berikut :

Tabel 6. Uji Serentak Parameter Model GWRR

F(2) statistik	df1	df2	Ftabel
1.64665	105	103.4755	0.72363

Berdasarkan Tabel 6, dengan menggunakan taraf signifikansi α=5% diperoleh nilai F(2) statistik sebesar 1.64665 dengan nilai $F_{(0.05;105;103.4755)}$=0.72363. Dari hasil tersebut maka dapat disimpulkan bahwa H_0 ditolak yang artinya variabel prediktor secara serentak berpengaruh terhadap model GWRR.

Pengujian yang selanjutnya dilakukan adalah uji signifikansi parameter model untuk mengetahui variabel yang signifikan berpengaruh terhadap model. Sebagai contoh untuk Kabupaten Cilacap diperoleh hasil sebagai berikut :

Tabel 7. Uji Parameter Model GWRR untuk Kabupaten Cilacap

Variabel	t_{hitung}	Kesimpulan
X_1	-0.72513	Tidak Signifikan
X_2	2.25885	Signifikan
X_3	1.63887	Tidak Signifikan
X_4	2.96808	Signifikan
X_5	0.86228	Tidak Signifikan
X_6	6.05923	Signifikan

Berdasarkan Tabel 7, variabel yang signifikan berpengaruh terhadap pendapatan asli daerah untuk Kabupaten Cilacap adalah Retribusi Daerah (X_2), PDRB Atas Dasar Harga Konstan (X_4) dan Pajak Daerah (X_6). Untuk melihat bahwa model GWRR mampu menangani kasus multikolinieritas pada model GWR, maka dilakukan deteksi multikolinieritas pada model GWRR. Berdasarkan hasil secara kesuluruhan yang diperoleh untuk masing-masing variabel pada setiap lokasi pengamatan diperoleh nilai VIF < 10, sehingga dapat disimpulkan bahwa GWRR mampu menangani maslah multikolinieritas pada model GWR.

4.4 Pemilihan Model Terbaik

Dalam memilih model yang terbaik adalah dengan memilih nilai MSE dan nilai AIC yang terkecil anatara model regresi linier berganda, GWR dan GWRR, diperoleh hasil sebagai berikut :

Tabel 8. Hasil Perbandingan model GWR dan GWRR

Model	MSE	AIC
Regresi Linier	0.1710	1.13600
GWR	0.1377	87.60366
GWRR	0.0058	-247.52051

Berdasarkan Tabel 8, model GWRR memiliki nilai MSE (0.0058) dan AIC (-247.52051) lebih kecil jika dibandingkan dengan model regresi linier dan GWR. Selain itu,dengan melihat nilai VIF secara keseluruhan pada tiap lokasi pengamatan untuk model GWRR tidak ada yang mengandung multikolinieritas..

5. PENUTUP

5.1 Kesimpulan

Berdasarkan hasil dan pembahasan pada bab sebelumnya, dapat diperoleh beberapa kesimpulan, diantaranya :

1. Terdapat perbedaan karakteristik lingkungan dan geografis antar wilayah pengamatan pada model regresi linier, sehingga dapat digunakan model GWR.
2. Terdapat masalah multikolinieritas pada model GWR, sehingga dapat digunakan model GWRR.
3. Model terbak dalam pemodelan Pendapatan Asli Daerah (PAD) Kabupaten/Kota di Provinsi Jawa Tengah adalah model GWRR karena mampu meminimumkan nilai MSE sebesar 0.0058 dan nilai AIC sebesar -247.52051 dan mampu menghilangkan adanya masalah multikolinieritas.

5.2 Saran

Diharapkan untuk penelitian selanjutnya terkait dengan pemodelan dengan menggunakan model GWR agar lebih memperhatikan adanya variabel yang bersifat global dan variabel yang bersifat lokal. Dengan demikian diharapkan dapat memperbaiki adanya kekurangan didalam penelitian ini.

DAFTAR PUSTAKA

[1] Draper, N., and Smith, H. 1992. *Analisis Regresi Terapan*. PT. Gramedia Pustaka Utama. Jakarta.

[2] Foteringham, A. S., Brundson, C. and Charlthon, M.. 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, UK.

[3] Leung, Y., Mei, C.L., & Zhang, W.X. 2000. Statistic Test for Spatial Non-Statonary Based on the Geographically Weighted Regression Model, *Environment and Planning A*, 32 9-32.

[4] Mahmudi, 2010. Manajemen Keuangan Daerah: Buku Seri *Membudayakan Akuntabilitas Publik:* Erlangga, Jakarta, Hal. 18.

[5] Montgomery, D. C., and Peck, E. A. 1991. *Introduction Linier Regression Analysis*. Second Edition. New York: John Wiley & Sons.

[6] Purhadi & 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.

[7] Santoso, P. B., dan Rahayu, R. F. 2005. Analisis Pendapatan Asli Daerah (PAD) dan Faktor-Faktor yang mempengaruhinya dalam Upaya Pelaksanaan Otonomi Daerah di Kabupaten Kediri. *Dinamika Pembangunan*,Vol.2, No.1,Hal.9-18.

[8] Sukmantoro, D. 2014. *Geographically Weighted Ridge Regression Dalam Pemodelan Nilai Tanah. Master Thesis of Statistics* Institut Teknologi Sepuluh Nopember Surabaya.
[9] Sutedi, 2008. *Hukum Pajak dan Retribusi Daerah*, Cetakan Pertama: Ghalia Indonesia, Bogor, Hal 57.
[10] Wheeler, D. 2007. Diagnostic Tools and a Remedial Method for Collinearity in Geographically Weighted Regression. *Environment and Planning A*, 39: 2464-2481.