

PEMILIHAN VARIABEL PADA MODEL *GEOGRAPHICALLY WEIGHTED REGRESSION*

Hasbi Yasin

Staf Pengajar Program Studi Statistika FMIPA UNDIP

Abstract

Regression analysis is a statistical analysis that aims to model the relationship between response variable with some predictor variables. Geographically Weighted Regression (GWR) is statistical method used for analyzed the spatial data in local form of regression. One of the problems in GWR is how to choose the significant variables. The number of predictor variables will allow the violation of assumptions about the absence of multicollinearity in the data. Therefore, this needs a method to reduce some of the predictor variables which not significant to the response variable. This paper will discuss how to select significant variables by stepwise method. This method is a combination of forward selection method and the backward elimination method.

Keywords: Geographically Weighted Regression, Backward Elimination, Forward Selection, Stepwise Method

1. Pendahuluan

Sebuah data pengamatan bila melibatkan informasi koordinat lokasi pengambilan data disamping data mengenai peubah-peubah yang sedang diamati maka termasuk dalam kategori data spasial. Analisis terhadap data spasial memerlukan perhatian lebih dibandingkan dengan analisis data nonspasial, khususnya ketika menggunakan regresi. Salah satu hal yang harus mendapat perhatian pada penanganan data spasial adalah kemungkinan munculnya heterogenitas spasial^[1]. Heterogenitas spasial muncul karena kondisi data di lokasi yang satu dengan lokasi yang lain tidak sama, baik dari segi geografis, keadaan sosial-budaya maupun hal-hal lain yang melatarbelakanginya. Salah satu dampak yang ditimbulkan dari munculnya heterogenitas spasial adalah parameter regresi bervariasi secara spasial atau disebut juga terjadi nonstasioneritas spasial pada parameter regresi.

Pada regresi OLS (*Ordinary Least Square*) diasumsikan bahwa nilai duga parameter regresi akan tetap (konstan), artinya parameter regresi bernilai sama untuk setiap titik di dalam wilayah penelitian (parameter global). Bila terjadi heterogenitas spasial pada parameter regresi, maka informasi yang tidak dapat ditangani oleh metode regresi OLS akan ditampung sebagai galat. Bila kasus semacam itu terjadi, regresi OLS menjadi kurang mampu dalam menjelaskan fenomena data yang sebenarnya. Untuk mengantisipasi munculnya heterogenitas spasial pada parameter regresi, regresi OLS dikembangkan menjadi *Geographically Weighted Regression* (GWR). Pada GWR, parameter regresi diasumsikan bervariasi secara spasial. Melalui penggunaan GWR akan dapat diketahui variasi spasial dalam nilai duga parameter, sehingga interpretasi yang berbeda dan berharga dapat diperoleh untuk setiap titik lokasi yang diteliti.

Banyaknya variabel prediktor memungkinkan akan terjadinya pelanggaran asumsi tentang tidak adanya multikolinieritas dalam data^[4]. Oleh karena itu diperlukan suatu metode untuk mereduksi beberapa variabel yang tidak signifikan terhadap responnya. Salah satu metode yang dapat digunakan adalah metode stepwise GWR. Penelitian ini

bertujuan untuk membahas bagaimana cara memilih variabel prediktor yang signifikan dengan menggunakan metode stepwise GWR.

2. Geographically Weighted Regression (GWR)

Model *Geographically Weighted Regression* (GWR) adalah pengembangan dari model regresi dimana setiap parameter dihitung pada setiap lokasi pengamatan, sehingga setiap lokasi pengamatan mempunyai nilai parameter regresi yang berbeda-beda. Model GWR merupakan pengembangan dari model regresi global dimana ide dasarnya diambil dari regresi non parametrik^[7]. Variabel respon y dalam model GWR diprediksi dengan variabel prediktor yang masing-masing koefisien regresinya bergantung pada lokasi dimana data tersebut diamati. Model GWR dapat ditulis sebagai berikut^[3]:

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

dengan:

- y_i : Nilai observasi variabel respon untuk lokasi ke- i
- (u_i, v_i) : Menyatakan koordinat letak geografis (*longitude, latitude*) dari lokasi pengamatan ke- i
- $\beta_k(u_i, v_i)$: Koefisien regresi variabel prediktor ke- k pada lokasi pengamatan ke- i
- x_{ik} : Nilai observasi variabel prediktor ke- k pada lokasi pengamatan ke- i
- ε_i : Error pengamatan ke- i yang diasumsikan identik, independen dan berdistribusi normal dengan mean nol dan varian konstan σ^2

3. Estimasi Parameter Model GWR

Estimasi parameter model GWR dilakukan dengan metode *Weighted Least Squares* (WLS) yaitu dengan memberikan pembobot yang berbeda untuk setiap lokasi dimana data diamati. Pemberian bobot ini sesuai dengan Hukum I Tobler: “*Everything is related to everything else, but near things are more related than distant things*” yang berarti “Segala sesuatu saling berhubungan satu dengan yang lainnya, tetapi sesuatu yang dekat lebih mempunyai pengaruh daripada sesuatu yang jauh”^[8]. Sehingga pada model GWR diasumsikan bahwa daerah yang dekat dengan lokasi pengamatan ke- i mempunyai pengaruh yang besar terhadap estimasi parameternya dari pada daerah yang lebih jauh. Misalkan pembobot untuk setiap lokasi (u_i, v_i) adalah $w_j(u_i, v_i)$, $j = 1, 2, \dots, n$ maka parameter pada lokasi pengamatan (u_i, v_i) diestimasi dengan menambahkan unsur pembobot $w_j(u_i, v_i)$ pada persamaan (1) dan kemudian meminimumkan jumlah kuadrat residual berikut ini :

$$\sum_{j=1}^n w_j(u_i, v_i) \varepsilon_j^2 = \sum_{j=1}^n w_j(u_i, v_i) \left[y_j - \beta_0(u_i, v_i) - \sum_{k=1}^p \beta_k(u_i, v_i) x_{jk} \right]^2$$

Atau dalam bentuk matriks jumlah kuadrat residualnya adalah:

$$\begin{aligned} \boldsymbol{\varepsilon}^T \mathbf{W}(u_i, v_i) \boldsymbol{\varepsilon} = & \mathbf{y}^T \mathbf{W}(u_i, v_i) \mathbf{y} - 2\boldsymbol{\beta}^T(u_i, v_i) \mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{y} \\ & + \boldsymbol{\beta}^T(u_i, v_i) \mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X} \boldsymbol{\beta}(u_i, v_i) \end{aligned} \quad (2)$$

dengan:

$$\boldsymbol{\beta}(u_i, v_i) = \begin{pmatrix} \beta_0(u_i, v_i) \\ \beta_1(u_i, v_i) \\ \vdots \\ \beta_p(u_i, v_i) \end{pmatrix} \text{ dan } \mathbf{W}(u_i, v_i) = \text{diag}(w_1(u_i, v_i), w_2(u_i, v_i), \dots, w_n(u_i, v_i)).$$

Jika persamaan (2) diturunkan terhadap $\boldsymbol{\beta}^T(u_i, v_i)$ dan hasilnya disamakan dengan nol maka diperoleh estimator parameter model GWR:

$$\begin{aligned} \frac{\partial \boldsymbol{\varepsilon}^T \mathbf{W}(u_i, v_i) \boldsymbol{\varepsilon}}{\partial \boldsymbol{\beta}^T(u_i, v_i)} &= 0 - 2\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{y} + 2\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X} \boldsymbol{\beta}(u_i, v_i) = 0 \\ -2\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{y} + 2\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X} \boldsymbol{\beta}(u_i, v_i) &= 0 \\ [\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X} \boldsymbol{\beta}(u_i, v_i) &= [\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{y} \\ \hat{\boldsymbol{\beta}}(u_i, v_i) &= [\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{y} \end{aligned} \quad (3)$$

Misalkan $\mathbf{x}_i^T = (1, x_{i1}, x_{i2}, \dots, x_{ip})$ adalah elemen baris ke- i dari matriks \mathbf{X} . Maka nilai prediksi untuk y pada lokasi pengamatan (u_i, v_i) dapat diperoleh dengan cara berikut:

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

Sehingga untuk seluruh pengamatan dapat dituliskan sebagai berikut:

$$\begin{aligned} \hat{\mathbf{y}} &= (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n)^T = \mathbf{L} \mathbf{y} \text{ dan} \\ \hat{\boldsymbol{\varepsilon}} &= (\hat{\varepsilon}_1, \hat{\varepsilon}_2, \dots, \hat{\varepsilon}_n)^T = (\mathbf{I} - \mathbf{L}) \mathbf{y} \end{aligned}$$

dengan \mathbf{I} adalah matriks identitas berukuran $n \times n$ dan

$$\mathbf{L} = \begin{pmatrix} \mathbf{x}_1^T (\mathbf{X}^T \mathbf{W}(u_1, v_1) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(u_1, v_1) \\ \mathbf{x}_2^T (\mathbf{X}^T \mathbf{W}(u_2, v_2) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(u_2, v_2) \\ \vdots \\ \mathbf{x}_n^T (\mathbf{X}^T \mathbf{W}(u_n, v_n) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(u_n, v_n) \end{pmatrix} \quad (4)$$

Estimator $\hat{\boldsymbol{\beta}}(u_i, v_i)$ pada persamaan (3) merupakan estimator tak bias dan konsisten untuk $\boldsymbol{\beta}(u_i, v_i)$.

4. Pembobotan Model GWR

Peran pembobot pada model GWR sangat penting karena nilai pembobot ini mewakili letak data observasi satu dengan lainnya. Skema pembobotan pada GWR dapat menggunakan beberapa metode yang berbeda. Ada beberapa literatur yang bisa digunakan untuk menentukan besarnya pembobot untuk masing-masing lokasi yang berbeda pada model GWR, diantaranya dengan menggunakan fungsi kernel (*kernel function*).

Fungsi kernel digunakan untuk mengestimasi parameter dalam model GWR jika fungsi jarak (w_j) adalah fungsi yang kontinu dan monoton turun^[2]. Pembobot yang terbentuk dengan menggunakan fungsi kernel ini adalah fungsi jarak *Gaussian* (*Gaussian*

Distance Function), fungsi *Exponential*, fungsi *Bisquare*, dan fungsi kernel *Tricube*. Masing-masing fungsi pembobot dapat ditulis sebagai berikut :

1. *Gaussian*^[51]:

$$w_j(u_i, v_i) = \phi(d_{ij}/\sigma h)$$

dimana ϕ adalah densitas normal standar dan σ menunjukkan simpangan baku dari vektor jarak d_{ij} .

2. *Exponential*^[51]:

$$w_j(u_i, v_i) = \sqrt{\exp\left(-\left(d_{ij}/h\right)^2\right)}$$

3. *Bisquare*^[2]:

$$w_j(u_i, v_i) = \begin{cases} \left(1 - \left(d_{ij}/h\right)^2\right)^2, & \text{untuk } d_{ij} \leq h \\ 0, & \text{untuk } d_{ij} > h \end{cases}$$

4. *Tricube*^[2]:

$$w_j(u_i, v_i) = \begin{cases} \left(1 - \left(d_{ij}/h\right)^3\right)^3, & \text{untuk } d_{ij} \leq h \\ 0, & \text{untuk } d_{ij} > h \end{cases}$$

dengan $d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2}$ adalah jarak *eucliden* antara lokasi (u_i, v_i) ke lokasi (u_j, v_j) dan h adalah parameter non negatif yang diketahui dan biasanya disebut parameter penghalus (*bandwidth*).

Ada beberapa metode yang digunakan untuk memilih *bandwidth* optimum, salah satu diantaranya adalah metode *Cross Validation* (CV) yang secara matematis didefinisikan sebagai berikut:

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

dengan $\hat{y}_{\neq i}(h)$ adalah nilai penaksir y_i dimana pengamatan di lokasi (u_i, v_i) dihilangkan dari proses estimasi. Untuk mendapatkan nilai h yang optimal maka diperoleh dari h yang menghasilkan nilai CV yang minimum.

5. Pengujian Hipotesis Model GWR

Pengujian hipotesis pada model GWR terdiri dari pengujian kesesuaian model GWR dan pengujian parameter model. Pengujian kesesuaian model GWR (*goodness of fit*) dilakukan dengan hipotesis sebagai berikut:

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

(tidak ada perbedaan yang signifikan antara model regresi global dan GWR)

$$H_1 : \text{Paling sedikit ada satu } \beta_k(u_i, v_i) \neq \beta_k, k = 0, 1, 2, \dots, p$$

(ada perbedaan yang signifikan antara model regresi global dan GWR).

Penentuan statistik uji berdasarkan pada nilai Jumlah Kuadrat Residual (*Residual Sum of Square/RSS*) yang diperoleh masing-masing dibawah H_0 dan H_1 . Dibawah kondisi H_0 , dengan menggunakan metode OLS diperoleh nilai RSS berikut:

$$\begin{aligned} \text{RSS}(H_0) &= \hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}} = (\mathbf{y} - \hat{\mathbf{y}})^T (\mathbf{y} - \hat{\mathbf{y}}) \\ &= \mathbf{y}^T (\mathbf{I} - \mathbf{H}) \mathbf{y} \end{aligned}$$

dengan $\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ yang bersifat idempotent.

Dibawah kondisi H_1 , koefisien regresi yang bervariasi secara spasial pada persamaan (1) ditentukan dengan metode GWR, sehingga diperoleh nilai RSS berikut:

$$\begin{aligned} \text{RSS}(H_1) &= \hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}} = (\mathbf{y} - \hat{\mathbf{y}})^T (\mathbf{y} - \hat{\mathbf{y}}) \\ &= \mathbf{y}^T (\mathbf{I} - \mathbf{L})^T (\mathbf{I} - \mathbf{L}) \mathbf{y} \end{aligned}$$

Dengan menggunakan selisih jumlah kuadrat residual dibawah H_0 dan dibawah H_1 diperoleh^[6]:

$$\begin{aligned} F &= \frac{(RSS(H_0) - RSS(H_1)) / \tau_1}{RSS(H_1) / \delta_1} \\ &= \frac{\mathbf{y}^T [(\mathbf{I} - \mathbf{H}) - (\mathbf{I} - \mathbf{L})^T (\mathbf{I} - \mathbf{L})] \mathbf{y} / \tau_1}{\mathbf{y}^T (\mathbf{I} - \mathbf{L})^T (\mathbf{I} - \mathbf{L}) \mathbf{y} / \delta_1} \end{aligned}$$

Dibawah H_0 F akan mengikuti distribusi F dengan derajat bebas $df_1 = \frac{\tau_1^2}{\tau_2}$ dan

$$df_2 = \left(\frac{\delta_1^2}{\delta_2} \right), \text{ dengan: } \tau_i = \text{tr} \left([(\mathbf{I} - \mathbf{H}) - (\mathbf{I} - \mathbf{L})^T (\mathbf{I} - \mathbf{L})]^i \right), \quad i = 1, 2.$$

Jika diambil taraf signifikansi α maka tolak H_0 jika $F \geq F_{\alpha, df_1, df_2}$.

6. Pemilihan Variabel Prediktor dalam Model GWR

Pada beberapa kasus seringkali dihadapkan pada masalah variabel prediktor yang terlalu banyak dan diduga saling berkorelasi antara variabel satu dengan yang lainnya. Untuk mengatasi hal ini, salah satu metode yang digunakan adalah metode *stepwise*. Metode ini merupakan kombinasi antara metode seleksi maju (*forward selection*) dan metode eliminasi mundur (*backward elimination*)^[6].

6.1 Metode Seleksi Maju (*Forward Selection*)

Metode seleksi maju menganggap bahwa model GWR pertama kali hanya mengandung konstanta. Kemudian satu per satu dari seluruh variabel prediktor diseleksi tingkat signifikansinya. Variabel yang paling signifikan akan menjadi variabel pertama yang digunakan dalam model GWR. Variabel yang tidak signifikan tidak akan digunakan dalam menyusun model GWR. Prosedur pemilihan variabel pada metode *forward* adalah:

1. Bentuk model GWR dengan hanya melibatkan parameter intersep, yaitu:

$$y_i = \beta_0(u_i, v_i) + \varepsilon_i \quad (5)$$

Sehingga estimasi parameternya adalah:

$$\hat{\beta}_0(u_i, v_i) = \frac{\sum_{j=1}^n w_j(u_i, v_i) y_j}{\sum_{j=1}^n w_j(u_i, v_i)}, \quad i = 1, 2, \dots, n$$

Dan jumlah kuadrat residualnya (RSS) adalah:

$$RSS_{(0)} = \mathbf{y}^T [\mathbf{I} - \mathbf{L}_{(0)}]^T [\mathbf{I} - \mathbf{L}_{(0)}] \mathbf{y}$$

dengan:

$$\mathbf{L}_{(0)} = \begin{bmatrix} \frac{w_1(u_1, v_1)}{\sum_{j=1}^n w_j(u_1, v_1)} & \frac{w_2(u_1, v_1)}{\sum_{j=1}^n w_j(u_1, v_1)} & \dots & \frac{w_n(u_1, v_1)}{\sum_{j=1}^n w_j(u_1, v_1)} \\ \frac{w_1(u_2, v_2)}{\sum_{j=1}^n w_j(u_2, v_2)} & \frac{w_2(u_2, v_2)}{\sum_{j=1}^n w_j(u_2, v_2)} & \dots & \frac{w_n(u_2, v_2)}{\sum_{j=1}^n w_j(u_2, v_2)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{w_1(u_n, v_n)}{\sum_{j=1}^n w_j(u_n, v_n)} & \frac{w_2(u_n, v_n)}{\sum_{j=1}^n w_j(u_n, v_n)} & \dots & \frac{w_n(u_n, v_n)}{\sum_{j=1}^n w_j(u_n, v_n)} \end{bmatrix}$$

2. Misalkan x_1, x_2, \dots, x_p adalah semua kandidat variabel prediktor. Untuk setiap x_k dibentuk model GWR berikut:

$$y_i = \beta_0(u_i, v_i) + \beta_k(u_i, v_i)x_{ik} + \varepsilon_i \tag{6}$$

dengan estimasi parameternya seperti pada persamaan (1) tetapi dengan matriks \mathbf{X} berukuran $n \times 2$, dengan elemen 1 pada kolom pertama dan n nilai observasi x_k pada kolom kedua. Hitung $RSS_{(k)}$:

$$RSS_{(k)} = \mathbf{y}^T [\mathbf{I} - \mathbf{L}_{(k)}]^T [\mathbf{I} - \mathbf{L}_{(k)}] \mathbf{y}$$

dengan \mathbf{L}_k sama seperti pada persamaan (4) dengan matriks \mathbf{X} dan vektor $\mathbf{x}_i^T = (x_{i1}, x_{i2}, \dots, x_{ip})$ yang disesuaikan.

3. Untuk setiap $k = 1, 2, \dots, p$ hitung statistik $R_{(k)}$

$$R_{(k)} = \frac{RSS_{(k)} / \delta_{1(k)}}{RSS_{(0)} / \delta_{1(0)}}, \quad k = 1, 2, \dots, p$$

dengan:

$$\delta_{i(k)} = tr \left\{ \left[\mathbf{I} - \mathbf{L}_{(k)} \right]^T \left[\mathbf{I} - \mathbf{L}_{(k)} \right] \right\}^i \quad \text{dan} \quad \delta_{i(0)} = tr \left\{ \left[\mathbf{I} - \mathbf{L}_{(0)} \right]^T \left[\mathbf{I} - \mathbf{L}_{(0)} \right] \right\}^i, \quad i = 1, 2.$$

Jika penambahan variabel x_k pada model (6) tidak dapat mengurangi jumlah kuadrat residual secara signifikan maka dapat diasumsikan bahwa model ini sama dengan model (5). Dengan asumsi ini maka $R_{(k)}$ dapat didekati dengan distribusi

$$F \left(\left(\frac{\delta_{1(k)}^2}{\delta_{2(k)}^2} \right), \left(\frac{\delta_{1(0)}^2}{\delta_{2(0)}^2} \right) \right).$$

Misalkan ξ_k adalah variabel random berdistribusi $F \left(\left(\frac{\delta_{1(k)}^2}{\delta_{2(k)}^2} \right), \left(\frac{\delta_{1(0)}^2}{\delta_{2(0)}^2} \right) \right)$. Untuk setiap

$k = 1, 2, \dots, p$ maka dapat dihitung p -value :

$$p_k = P \left(\xi_k \leq R_{(k)} \right), \quad k = 1, 2, \dots, p$$

Anggap bahwa:

$$p_{k_0} = \min_{1 \leq k \leq p} \{ p_k \}.$$

Untuk suatu level signifikansi α , jika $p_{k_0} < \alpha$, maka untuk pertama kali variabel x_{k_0} dimasukkan dalam model karena p -value terkecil yang menunjukkan bahwa variabel x_{k_0} paling signifikan untuk mengurangi RSS model. Jika $p_{k_0} \geq \alpha$ maka tidak ada satu pun variabel prediktor yang akan dimasukkan dalam model GWR.

4. Pada kondisi $p_{k_0} < \alpha$, maka tambahkan setiap variabel x_k ($k \neq k_0$) pada model:

$$y_i = \beta_0(u_i, v_i) + \beta_{k_0}(u_i, v_i)x_{ik_0} + \beta_k(u_i, v_i)x_{ik} + \varepsilon_i \quad (7)$$

Kemudian model dikalibrasi dan dihitung RSS model:

$$RSS_{(k_0, k)} = \mathbf{y}^T [\mathbf{I} - \mathbf{L}_{(k_0, k)}]^{-T} [\mathbf{I} - \mathbf{L}_{(k_0, k)}] \mathbf{y}, \quad k \neq k_0.$$

Sehingga:

$$R_{(k_0, k)} = \frac{RSS_{(k_0, k)} / \delta_{1(k_0, k)}}{RSS_0 / \delta_{1(0)}}, \quad k \neq k_0$$

dengan $\mathbf{L}_{k_0, k}$ dan $\delta_{i, k_0, k}$ identik dengan langkah sebelumnya.

Kemudian dihitung p -value model sebagai berikut:

$$p_k = P\left(\xi_k \leq R_{(k_0, k)}\right), \quad k \neq k_0$$

dengan ξ_k ($k \neq k_0$) berdistribusi $F\left(\left(\frac{\delta_{1(k_0, k)}^2}{\delta_{2(k_0, k)}}\right), \left(\frac{\delta_{1(k_0)}^2}{\delta_{2(k_0)}}\right)\right)$.

Misalkan $p_{k_1} = \min_{k \neq k_0} \{ p_k \}$, jika $p_{k_1} < \alpha$, maka variabel x_{k_1} dimasukkan dalam model pada persamaan (7). Jika $p_{k_1} \geq \alpha$ maka model pada persamaan (7) adalah model GWR final.

5. Ulangi langkah 4 sampai tidak ada satu pun variabel prediktor yang dapat dimasukkan dalam model sehingga didapatkan model GWR final.

6.2 Metode Eliminasi Mundur (*Backward Elimination*)

Metode eliminasi mundur awalnya menganggap bahwa seluruh variabel prediktor digunakan dalam menyusun model GWR. Kemudian satu per satu dari seluruh variabel prediktor diseleksi tingkat signifikansinya. Variabel yang pengaruhnya paling tidak signifikan akan menjadi variabel pertama yang dikeluarkan dari dalam model GWR. Variabel yang signifikan tidak akan dikeluarkan dari model GWR. Prosedur pemilihan variabel pada metode *backward* adalah:

1. Membentuk model GWR dengan memasukkan semua kandidat variabel prediktor:

$$y_i = \beta_0(u_i, v_i) + \sum_{m=1}^p \beta_m(u_i, v_i)x_{im} + \varepsilon_i \quad (8)$$

Sehingga jumlah kuadrat residualnya (RSS) adalah:

$$RSS_{(1,2,\dots,p)} = \mathbf{y}^T [\mathbf{I} - \mathbf{L}_{(1,2,\dots,p)}]^{-T} [\mathbf{I} - \mathbf{L}_{(1,2,\dots,p)}] \mathbf{y}$$

2. Misalkan x_1, x_2, \dots, x_p adalah semua kandidat variabel prediktor. Untuk setiap x_k $k = 1, 2, \dots, p$ dibentuk model GWR dengan mengeliminasi variabel prediktor x_k berikut:

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

dengan Jumlah kuadrat residualnya (RSS) adalah:

$$RSS_{(1, \dots, k-1, k+1, \dots, p)} = \mathbf{y}^T \left[\mathbf{I} - \mathbf{L}_{(1, \dots, k-1, k+1, \dots, p)} \right]^T \left[\mathbf{I} - \mathbf{L}_{(1, \dots, k-1, k+1, \dots, p)} \right] \mathbf{y}$$

Sehingga diperoleh $R_{(k)}$:

$$R_{(k)} = \frac{RSS_{(1, 2, \dots, p)} / \delta_{1(1, 2, \dots, p)}}{RSS_{(1, \dots, k-1, k+1, \dots, p)} / \delta_{1(1, \dots, k-1, k+1, \dots, p)}}$$

dengan:

$$\delta_{i(1, 2, \dots, p)} = \text{tr} \left\{ \left[\mathbf{I} - \mathbf{L}_{(1, 2, \dots, p)} \right]^T \left[\mathbf{I} - \mathbf{L}_{(1, 2, \dots, p)} \right] \right\}^i$$

$$\delta_{i(1, \dots, k-1, k+1, \dots, p)} = \text{tr} \left\{ \left[\mathbf{I} - \mathbf{L}_{(1, \dots, k-1, k+1, \dots, p)} \right]^T \left[\mathbf{I} - \mathbf{L}_{(1, \dots, k-1, k+1, \dots, p)} \right] \right\}$$

dan $i = 1, 2$.

3. Seperti metode *forward*, hitung p -value dari $R_{(k)}$ sebagai kriteria pemilihan variabel.

$$p_k = P\left(\xi_k \leq R_{(k)}\right)$$

dengan ξ_k adalah variabel random berdistribusi $F \left(\left(\frac{\delta_{1(1, 2, \dots, p)}^2}{\delta_{2(1, 2, \dots, p)}^2} \right), \left(\frac{\delta_{1(1, \dots, k-1, k+1, \dots, p)}^2}{\delta_{2(1, \dots, k-1, k+1, \dots, p)}^2} \right) \right)$.

Misalkan $p_{k_0} = \max_{1 \leq k \leq p} \{p_k\}$, untuk suatu level signifikansi α , jika $p_{k_0} \geq \alpha$, maka variabel x_{k_0} dikeluarkan dari model (8). Jika $p_{k_0} < \alpha$ maka tidak ada satu pun variabel prediktor yang akan dikeluarkan dari model GWR.

4. Ulangi langkah (3) sampai tidak ada satu pun variabel prediktor yang dapat dikeluarkan dari model sehingga didapatkan model GWR final.

6.3 Metode *Stepwise*

Kelemahan dari metode *forward* adalah jika suatu variabel prediktor sudah masuk ke dalam model maka variabel tersebut tidak akan dikeluarkan lagi. Pada kenyataannya jika variabel lain dimasukkan ke dalam model maka kemungkinan variabel yang dimasukkan sebelumnya akan menjadi tidak signifikan dalam mengurangi jumlah kuadrat residualnya. Metode ini tidak dapat mengukur perubahan tingkat signifikansi selama terjadi penambahan variabel prediktor yang lain. Begitu pun sebaliknya pada metode *backward*, jika suatu variabel prediktor sudah dikeluarkan dari model maka variabel tersebut tidak dimasukkan lagi. Oleh karena itu dilakukan kombinasi dari kedua metode ini yang disebut dengan metode *stepwise*.

Prosedur pemilihan variabel pada metode *stepwise* adalah:

1. Bentuk model (5) dan pilih variabel prediktor satu per satu untuk dimasukkan dalam model seperti pada metode *forward*.

2. Lakukan pemeriksaan setiap variabel yang dimasukkan dalam model hasil metode *forward* pada langkah sebelumnya dengan menggunakan metode *backward*. Hal ini dilakukan untuk memeriksa tingkat signifikansi dari variabel terpilih. Jika variabel yang dimasukkan pada langkah sebelumnya menjadi tidak signifikan maka akan dikeluarkan dari model.

Ulangi langkah (1) dan (2) sampai tidak ada lagi variabel yang dapat dimasukkan maupun dikeluarkan dari model.

7. Contoh Aplikasi

Data yang digunakan adalah data spasial tentang tingkat kriminalitas (*crime*) di Columbus sebagai variabel respon kemudian *income* (X_1) dan *housing* (X_2) sebagai prediktornya^[1]. Komputasi model dilakukan dengan menggunakan GUI Matlab R2008a. Pemodelan GWR dengan fungsi pembobot *Gaussian* menunjukkan bahwa dengan menggunakan tingkat signifikansi (α) sebesar 5% maka dapat disimpulkan bahwa model GWR berbeda signifikan dengan model regresi linier global (Tabel 1).

Tabel 1. Uji Kesesuaian Model GWR dengan Pembobot *Gaussian*

Sumber Error	Jumlah Kuadrat	Derajat Bebas	Rata-rata kuadrat	F	p-value
Improvement	5232.4448	30.2665	172.8793	3.4764	0,0009*
GWR	782.4112	15.7335	49.7289		
Regresi	6014.8560	46.0000			

Ket: *) signifikan pada $\alpha = 5\%$

Kemudian untuk melakukan pemilihan variabel digunakan metode seleksi *forward*. Pada tahap pertama dibentuk model GWR tanpa melibatkan variabel prediktor selain konstanta. Hasilnya menunjukkan bahwa jumlah kuadrat residualnya adalah 1931.486 dengan nilai R^2 adalah 85.63%.

Tahap kedua adalah memasukkan setiap variabel prediktor ke dalam model GWR. Hasilnya adalah sebagai berikut:

Tabel 2. Penambahan Variabel pada Tahap Kedua

Penambahan Variabel	R^2	$RSS_{(k)}$	Derajat Bebas	$R_{(k)}$	P-Value (p_k)
<i>income</i> (X_1)	51.79%	6478.148	44.723	2.337	0.983
<i>housing</i> (X_2)	92.83%	963.325	19.585	0.573	0.112*

Ket: *) signifikan pada $\alpha = 0.15$

Berdasarkan Tabel 2, dengan penambahan variabel *housing* (X_2) dalam model GWR dapat menurunkan jumlah kuadrat residual sebesar 968.161 menjadi 963.325 dan mampu menaikkan R^2 menjadi 92.83%. Jika diambil nilai kriteria untuk seleksi *Alpha to entered* (α)=0.15 dapat disimpulkan bahwa variabel *housing* (X_2) merupakan variabel yang paling signifikan. Sehingga variabel *housing* (X_2) akan dimasukkan kedalam model terlebih dulu dibandingkan dengan variabel *income* (X_1).

Tahap ketiga adalah mengestimasi model GWR dengan memasukkan semua variabel prediktor. Hasil estimasi pada tahap ketiga menunjukkan bahwa penambahan variabel *income* (X_1) hanya mampu menurunkan jumlah kuadrat residual menjadi 782.411 dan menaikkan R^2 menjadi 94.18% dengan p-value 0.437. Dengan demikian variabel *income* (X_1) tidak memberikan pengaruh yang signifikan terhadap model, sehingga model GWR terbaiknya adalah:

$$y_i = \beta_0(u_i, v_i) + \beta_2(u_i, v_i)x_{i2} \quad (9)$$

Model (9) menunjukkan bahwa besarnya tingkat kriminalitas di setiap wilayah secara signifikan dipengaruhi oleh kondisi wilayah dan besarnya nilai *housing* (X_2) dimana setiap wilayah mempunyai model yang spesifik.

8. Penutup

Pemilihan variabel signifikan pada model GWR dapat dilakukan dengan metode *forward selection*, *backward elimination* ataupun dengan kombinasi keduanya yaitu dengan menggunakan metode *stepwise*. Pemilihan variabel ini dapat dilakukan untuk mengatasi masalah multikolinieritas pada variabel prediktor. Besarnya nilai kriteria untuk seleksi *Alpha to entered* (α) akan berpengaruh pada proses pemilihan variabel pada model GWR.

DAFTAR PUSTAKA

1. Anselin, L., *Spatial Econometrics: Method and Models*, Kluwer Academic Publishers, the Netherlands, 1988.
2. Chasco, C., Garcia, I., & Vicens, J., Modeling Spastial Variations in Household Disposable Income with Geographically Weighted Regression, *Munich Personal RePEc Arkhive (MPRA)*, Working Papper, 2007, No. 1682.
3. Fotheringham, A.S., Brunson, C., & Charlton, M., *Geographically Weighted Regression*, Jhon Wiley & Sons, Chichester, UK., 2002.
4. Gujarati, D., *Essentials of Econometrics*. Mc Graw-Hill, Inc, New York, 1992.
5. LeSage, J.P., *A Family of Geographically Weighted Regression*, Departement of Economics University of Toledo, 2001.
6. Leung, Y., Mei, C.L., & Zhang, W.X., Statistic Tests for Spatial Non-Stationarity Based on the Geographically Weighted Regression Model, *Environment and Planning A*, 2000, 32: 9-32.
7. Mei, C.L., Wang, N., & Zhang, W.X., "Testing The Importance of The Explanatory Variables in A Mixed Geographically Weighted Regression Model", *Environment and Planning A*, 2006, vol. 38: 587-598.
8. Miller, H.J., 'Tobler's First Law and Spatial Analysis'. *Annals of the Association of America Geographers*, 2004, 94(2): 284-289.