BAB III

GEOGRAPHICALLY WEIGHTED REGRESSION (GWR)

3.1 Data Spasial

Data spasial adalah data yang berorientasi pada geografis, memiliki sistem koordinat tertentu sebagai dasar referensinya. Terdapat dua bagian penting yang membuat data spasial berbeda dari data lain, yaitu informasi lokasi (spasial) yang berkaitan dengan suatu koordinat baik koordinat geografi (lintang atau *latitude* dan bujur atau *longitude*) dan informasi deskriptif (*attribute*) atau informasi non spasial yang berkaitan dengan lokasi yang memiliki beberapa keterangan yang berkaitan dengannya, contohnya: jenis vegetasi, populasi, luasan, kode pos, dan sebagainya (Hartoyo, Nugroho, Bhirowo, & Khalil, 2010).

Suatu lokasi pengamatan akan memiliki kondisi informasi yang berbeda dengan lokasi pengamatan lainnya. Lokasi pengamatan yang saling berdekatan akan memiliki pengaruh hubungan yang lebih besar daripada lokasi pengamatan yang jauh. Karena data yang diamati adalah data tiap lokasi pengamatan, maka akan mengakibatkan terdapatnya data yang heterogen karena kemungkinan tiap lokasi akan memiliki varians yang berbeda atau perbedaan pengaruh variabel prediktor terhadap variabel respon di setiap lokasi pengamatan.

Salah satu metode statistika untuk mengatasi permasalahan tersebut yang berkaitan dengan regresi, dengan memperhatikan letak geografis atau lokasi pengamatan adalah metode regresi terboboti geografis (*Geographically Weighted Regression).*

3.2 Heterogenitas Spasial

Adanya salah satu aspek spasial yaitu sifat heterogenitas spasial merupakan syarat bisa dilakukan pemodelan data dengan menggunakan pendekatan titik dengan GWR. Untuk mendeteksi ada atau tidaknya heterogenitas spasial dalam model dilakukan uji *Breusch-Pagan* dengan langkah-langkah sebagai berikut :

a. Perumusan Hipotesis

$H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2$

(Tidak terdapat heterogenitas spasial)

H_1 : Minimal ada satu $\sigma_1^2 \neq \sigma^2$, untuk $i \neq j$, dengan $i, j = 1,2, \dots, n$

(Terdapat heterogenitas spasial)

b. Statistik Uji

$$BP = \frac{1}{2} b^T Z (Z^T Z)^{-1} Z^T b$$

dengan

$b = \frac{e_i^2}{\sigma^2} - 1$,

Z : vektor variabel respon y yang berukuran ($n \times 1$),

e_i^2 : kuadrat residual atau *error* untuk pengamatan ke-i dan σ^2 adalah varians

dari e_i

c. Kriteria Pengujian

H_0 ditolak jika $BP > \chi^2_{(\alpha,p)}$, dengan p banyaknya variabel prediktor, dan $\alpha = 5\%$.

d. Kesimpulan

Penafsiran dari H_0 diterima atau ditolak.

- H_0 diterima, kesimpulannya tidak terdapat heterogenitas spasial dalam model.
- H_0 ditolak, kesimpulannya terdapat heterogenitas spasial dalam model. Sehingga aspek spasial terpenuhi dan selanjutnya dapat dilakukan analisis lebih lanjut dalam penelitian menggunakan pendekatan GWR.

3.3 Model *Geographically Weighted Regression* (GWR)

Geographically Weighted Regression (GWR) merupakan pengembangan dari regresi linear yang mana pada GWR berlaku secara lokal. Model ini menghitung parameter pada setiap lokasi pengamatan atau dengan kata lain memperhitungkan lokasi data pengamatan. Sehingga setiap daerah lokasi pengamatan mempunyai nilai parameter regresi yang berbeda-beda, karena dilakukan pembobotan berdasarkan lokasi pengamatan atau daerah tersebut.

Model untuk *Geographically Wighted Regression* (GWR) sebagai berikut (Fotheringham, Brunsdon, & Charlton, 2002) :

$$Y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i) X_{ik} + \varepsilon_i \tag{3.1}$$

dimana

Y_i	: Nilai variabel respon pada titik lokasi pengamatan ke-*i*
$\beta_0(u_i, v_i)$	: Konstanta/*intercept* GWR
$\beta_k(u_i, v_i)$	: Koefisien regresi ke-*k* pada titik lokasi pengamatan ke-*i*
u_i, v_i	: Titik koordinat lintang dan bujur pada lokasi pengamatan ke-*i*
X_{ik}	: Nilai variabel prediktor ke-*k* pada titik lokasi pengamatan ke-*i*
ε_i	: *Error* pada titik lokasi ke-i yang diasumsikan independen, identik, dan berdistribusi normal dengan rata-rata nol dan varians σ^2

3.4 Penaksir Parameter GWR

Seperti dijelaskan sebelumnya, metode GWR merupakan pengembangan dari regresi linear klasik, dimana koefisien pada GWR berlaku secara lokal, berbeda untuk setiap lokasi pengamatan. Sehingga jika pada regresi linear klasik, penaksir parameternya menggunakan *Ordinary Least Square* (OLS), maka pada GWR penaksir parameternya menggunakan *Weighted Least Square* (WLS), yang mana setiap lokasi pengamatan diboboti secara berbeda. Dengan kata lain variabel prediktor dan variabel responnya bergantung pada titik lokasi pengamatan.

Penaksir parameter GWR $\beta_k(u_i, v_i)$ untuk setiap variabel ke-k pada lokasi pengamatan ke-i, dinyatakan sebagai berikut :

$$\hat{\beta}(u_i, v_i) = [X^T W(u_i, v_i) X]^{-1} X^T W(u_i, v_i) Y \tag{3.2}$$

Dimana untuk memperoleh penaksir parameter tersebut digunakan *Weighted Least Square*. Pertama persamaan (3.1) dapat diubah menjadi berikut :

$$\varepsilon_i^2 = Y_i - \beta_0(u_i, v_i) - \sum_{k=1}^{p} \beta_k(u_i, v_i) X_{ik} \tag{3.3}$$

Lalu, tambahkan pembobot w_{ij} pada persamaan (3.3) dan minimumkan jumlah kuadrat *error-nya*, sebagai berikut :

$$\sum_{j=1}^{n} w_{ij}\varepsilon_i^2 = \sum_{j=1}^{n} w_{ij}\left[Y_i - \beta_0(u_i,v_i) - \sum_{k=1}^{p}\beta_k(u_i,v_i)X_{ik}\right]^2 \tag{3.4}$$

Dimana w_{ij} merupakan pembobot untuk setiap titik lokasi pengamatan (u_i, v_i), dengan $j = 1,2,\dots,n$. Sehingga matriks pembobot pada titik lokasi pengamatan ke-i adalah $W(u_i, v_i)$ dinyatakan sebagai berikut :

$$W(u_i,v_i) = \begin{bmatrix} w_{i1} & 0 & \dots & 0 \\ 0 & w_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{in} \end{bmatrix} \tag{3.5}$$

$W(u_i, v_i)$ merupakan matriks diagonal berukuran $(n \times n)$ dengan setiap elemen diagonalnya adalah pembobot untuk masing-masing titik lokasi pengamatan (u_i, v_i) atau w_{ij}.

Persamaan (3.4) dapat ditulis dalam bentuk matriks sebagai berikut :

$$\varepsilon^T W(u_i,v_i)\varepsilon = Y^T W(u_i,v_i)Y - 2\beta^T(u_i,v_i)X^T W(u_i,v_i)Y + \beta^T(u_i,v_i)X^T W(u_i,v_i)X\beta\,(u_i,v_i) \tag{3.6}$$

Minimumkan jumlah kuadrat *error*-nya dengan mendiferensialkan persamaan (3.6) terhadap $\beta^T(u_i, v_i)$ sebagai berikut :

$$\frac{\partial}{\partial\beta^T(u_i,v_i)}\varepsilon^T W(u_i,v_i)\varepsilon = 0 - 2X^T W(u_i,v_i)Y + 2X^T W(u_i,v_i)X\hat{\beta}(u_i,v_i) = 0 \tag{3.7}$$

Sehingga diperoleh,

$$0 - 2X^T W(u_i,v_i)Y + 2X^T W(u_i,v_i)X\hat{\beta}(u_i,v_i) = 0$$

$$-2X^T W(u_i,v_i)Y + 2X^T W(u_i,v_i)X\hat{\beta}(u_i,v_i) = 0$$

$$2X^T W(u_i,v_i)X\hat{\beta}(u_i,v_i) = 2X^T W(u_i,v_i)Y$$

$$X^T W(u_i,v_i)X\hat{\beta}(u_i,v_i) = X^T W(u_i,v_i) \tag{3.8}$$

Kalikan kedua ruas dari sebelah kiri pada persamaan (3.8) dengan invers dari $X^T W(u_i, v_i)X$ sehingga :

$$[X^T W(u_i,v_i)X]^{-1}X^T W(u_i,v_i)X\hat{\beta}(u_i,v_i) = [X^T W(u_i,v_i)X]^{-1}X^T W(u_i,v_i)Y \tag{3.9}$$

Karena $[X^T W(u_i,v_i)X]^{-1}X^T W(u_i,v_i)X = I$, maka :

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

atau

$$\hat{\beta}(u_i,v_i) = [X^T W(u_i,v_i)X]^{-1}X^T W(u_i,v_i)Y \tag{3.10}$$

(Maulani, 2013).

3.5 Pembobotan dan Pemilihan *Bandwidth* Optimum

Pembobotan dalam model regresi terboboti geografis merupakan aspek penting. Berdasarkan hukum Tobler I, "Segala sesuatu saling berhubungan satu dengan yang lainnya, tetapi sesuatu yang dekat lebih mempunyai pengaruh dari pada sesuatu yang jauh" (hukum Tobler I dalam Miller, 2004). Sehingga titik terdekat dari data yang diamati pada titik lokasi pengamatan ke-*i* akan memiliki pengaruh yang lebih besar dalam menaksir parameter $\beta_k(u_i,v_i)$. Maka, semakin dekat jarak antara lokasi pengamatan dengan titik i akan semakin besar juga bobotnya.

Fungsi pembobot kernel *bisquare* terdiri dari (Chasco, Garcia, & Vicens, 2007) :

1. *Fixed Bisquare* :

$$w_{ij} = \begin{cases} \left(1 - \left(\dfrac{d_{ij}}{h}\right)^2\right)^2, & \text{untuk } d_{ij} < h \\ 0, & \text{lainnya} \end{cases}$$

2. *Adaptive Bisquare* :

$$w_{ij} = \begin{cases} \left(1 - \left(\dfrac{d_{ij}}{h_i}\right)^2\right)^2, & \text{untuk } d_{ij} < h_i \\ 0, & \text{lainnya} \end{cases}$$

Dimana $d_{ij} = \sqrt{\left(u_i - u_j\right)^2 + \left(v_i - v_j\right)^2}$, merupakan fungsi jarak *euclidean* antara titik lokasi pengamatan ke-*i* dengan titik lokasi pengamatan ke-*j*.

u_i : *Longitude* pada lokasi ke-*i*

u_j : *Longitude* pada lokasi ke-j

v_i : *Latitude* pada lokasi ke-i

v_j : *Latitude* pada lokasi ke-j

Dan h merupakan $bandwidth$. Secara teoritis, *bandwidth* merupakan lingkaran dengan radius h dari titik lokasi pengamatan ke-i. Lokasi pengamatan terdekat dengan lokasi ke-i akan memiliki pengaruh yang lebih besar dalam penaksiran parameter. Pengamatan yang terletak dalam radius h akan diboboti sesuai dengan fungsi pembobotan yang digunakan. Sedangkan pengamatan yang terletak di luar radius h akan diboboti nol sehingga tidak mempengaruhi penaksiran parameter.

Sehingga pemilihan $bandwidth$ optimum sangat penting untuk dilakukan. Salah satu metode untuk menentukan $bandwidth$ optimum adalah validasi silang atau $cross\ validation$ (CV). $Bandwidth$ optimum adalah $bandwidth$ yang menghasilkan nilai CV minimum. Persamaan matematis *Cross Validation* (CV) (Fotheringham, Brunsdon, & Charlton, 2002) :

$$CV = \sum_{i=1}^{n} (y_i - \hat{y}_{\neq i}(h))^2 \tag{3.11}$$

Dimana $\hat{y}_{\neq i}(h)$ adalah penaksir y_i dimana pengamatan di lokasi ke-i dihilangkan dari proses penaksiran.

Pada fungsi pembobot *fixed*, $bandwidth$ (h) berlaku untuk setiap lokasi pengamatan. Namun pada fungsi pembobot $adaptive$, *bandwidth* (h_i) berbeda sesuai lokasi pengamatan ke-i.

3.6 Uji Keberartian Model GWR

Uji keberartian model GWR dilakukan untuk mengetahui apakah faktor geografis berpengaruh signifikan atau tidak terhadap model, dengan langkah-langkah sebagai berikut :

a. Hipotesis pengujian

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

(Tidak ada pengaruh yang signifikan antara faktor geografis terhadap model GWR)

H_1 : Minimal ada satu $\beta_k(u_i, v_i)$ yang berbeda untuk $k = 0,1,2, \dots, p$; dan $i = 1,2, \dots, n$

(Ada pengaruh yang signifikan antara faktor geografis terhadap model GWR)

b. Besaran-Besaran yang Diperlukan

- $JK(S)_{OLS} = Y^T(I - S_0)^T(I - S_0)Y$

  dengan $S_0 = X(X^TX)^{-1}X^T$

- $JK(S)_{GWR} = Y^T(I - S_1)^T(I - S_1)Y$

  dengan $S_1 = \begin{bmatrix} x_1^T[X'W(u_1, v_1)X]^{-1}X^TW(u_1, v_1) \\ x_2^T[X'W(u_2, v_2)X]^{-1}X^TW(u_2, v_2) \\ \vdots \\ x_n^T[X'W(u_n, v_n)X]^{-1}X^TW(u_n, v_n) \end{bmatrix}$

c. Statistik Uji

$$F = \frac{(JK(S)_{OLS} - JK(S)_{GWR})/v_1}{JK(S)_{GWR}/\delta_1}$$

d. Kriteria Pengujian

Dengan mengambil taraf nyata α, maka :

H_0 ditolak jika $F \geq F_{\alpha;dk_1,dk_2}$, dimana $dk_1 = \frac{v_1^2}{v_2}$, dengan $v_1 = n - p - 1 - \delta_1$, $v_2 = n - p - 1 - 2\delta_1 + \delta_2$, dimana $d_k = tr[(I - S)^T(I - S)]^k, k = 1,2,$ dengan S adalah matriks *hat* dari model yang mentransformasi vektor $\hat{Y}$ dari nilai Y pengamatan dan dk penyebut $dk_2 = \frac{\delta_1^2}{\delta_2}$ (Brundson *et al*, 1999 dalam Prasetyo, 2012).

e. Kesimpulan

Penafsiran dari H_0 diterima atau ditolak.

- H_0 ditolak, kesimpulannya ada perbedaan yang signifikan antara faktor geografis terhadap model.
- H_0 diterima, kesimpulannya tidak ada pengaruh yang signifikan antara faktor geografis terhadap model.

3.7 Pengujian Keberartian Koefisien GWR

Pengujian keberartian koefisien GWR dilakukan untuk mengetahui variabel-variabel prediktor yang berpengaruh signifikan terhadap variabel respon, dengan langkah-langkah sebagai berikut :

a. Perumusan Hipotesis

$H_0 : \beta_k(u_i, v_i) = 0$, untuk setiap $k = 1,2, \dots, p$

$H_1 : \beta_k(u_i, v_i) \neq 0$

b. Besaran-Besaran yang Diperlukan

$$SE\left[\hat{\beta}_k(u_i, v_i)\right] = \sqrt{Var\left[\hat{\beta}_k(u_i, v_i)\right]} = \sqrt{CC^T\sigma^2}$$

dengan

$$C = [X^T W(u_i, v_i) X]^{-1} X^T W(u_i, v_i)$$

dan

$$\sigma^2 = \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{dk_2}$$

c. Statistik Uji

$$t = \frac{\hat{\beta}_k(u_i, v_i)}{SE\left[\hat{\beta}_k(u_i, v_i)\right]}$$

d. Kriteria Pengujian

Dengan mengambil taraf nyata α, maka :

H_0 ditolak, jika $|t| > t_{\left(\frac{\alpha}{2},(n-p-1)\right)}$, dimana n banyaknya pengamatan, dan p banyaknya variabel prediktor.

e. Kesimpulan

Penafsiran dari H_0 diterima atau ditolak.

- Jika H_0 ditolak, maka $\beta_k(u_i, v_i) \neq 0$. Artinya, koefisien regresi lokal $\beta_k(u_i, v_i)$ pada model GWR tersebut berarti.
- Jika H_0 diterima, maka $\beta_k(u_i, v_i) = 0$. Artinya, koefisien regresi lokal $\beta_k(u_i, v_i)$ pada model GWR tersebut tidak berarti.

3.8 Pemilihan Model Terbaik

Dalam pemilihan model dengan pembobotan terbaik dapat dilakukan dengan membandingkan koefisien determinasi lokal (R_i^2) model GWR, dengan rumus sebagai berikut (Fotheringham, Brunsdon, & Charlton, 2002):

$$R_i^2 = \frac{JK(T)_{GWR} - JK(S)_{GWR}}{JK(T)_{GWR}}$$

dimana $JK(T)_{GWR}$ merupakan jumlah kuadrat total model GWR dan dinyatakan sebagai berikut :

$$JK(T)_{GWR} = \sum_{j=1}^{n} w_{ij}\left(Y_j - \bar{Y}\right)^2$$

sedangkan $JK(S)_{GWR}$ merupakan jumlah kuadrat sisa (residual) model GWR, dinyatakan sebagai berikut :

$$JK(S)_{GWR} = \sum_{j=1}^{n} w_{ij}\left(Y_j - \hat{Y}_j\right)^2$$

dimana w_{ij} adalah pembobot dari titik lokasi pengamatan ke-*j* di titik lokasi pengamatan ke-*i*, dengan $i, j = 1,2, \dots, n$. Penggunaan model GWR dengan R_i^2 lebih besar merupakan model yang lebih tepat digunakan.

3.9 Metode Analisis Penentuan Faktor-Faktor

Adapun langkah-langkah analisis yang dilakukan untuk menentukan faktor-faktor yang mempengaruhi Indeks Pembangunan Manusia (IPM) di Provinsi Jawa Barat dengan menggunakan GWR adalah sebagai berikut :

1. Mendeskripsikan variabel respon dan variabel prediktor dalam Indeks Pembangunan Manusia (IPM) di Provinsi Jawa Barat.
2. Menentukan koordinat *latitude* dan *longitude* tiap kota / kabupaten di Jawa Barat.
3. Menghitung jarak *euclidean* spasial antarkota / kabupaten di Jawa Barat.

4. Membandingkan jumlah kuadrat residual (*residual sum of square*) dan koefisien determinasi R^2 dari model GWR dengan pembobot *fixed bisquare*, dan GWR dengan pembobot *adaptive bisquare.*
5. Menentukan *bandwidth* spasial berdasarkan pembobotan terbaiknya.
6. Menghitung matriks pembobot tiap kota / kabupaten di Jawa Barat dengan fungsi pembobotan terbaiknya.
7. Menaksir parameter GWR dengan menggunakan *bandwith* optimum.
8. Menginterpretasi dan menyimpulkan hasil yang diperoleh.