BAB III

GEOGRAPHICALLY WEIGHTED LASSO

3.1 Variabel Penelitian

Variabel dependen yang digunakan adalah angka Produk Domestik Regional Bruto (PDRB). Variabel penjelas yang digunakan ada 13 variabel sebagai berikut:

X_1 : Jumlah petani

X_2 : Jumlah pertambangan

X_3 : Jumlah industri

X_4 : Jumlah pelanggan air bersih

X_5 : Jumlah perusahaan kontruksi

X_6 : Jumlah hotel dan penginapan

X_7 : Jumlah kendaraan umum

X_8 : Pendapatan asli daerah

X_9 : Jumlah penduduk berkerja

X_{10} : Angka harapan hidup

X_{11} : Harapan lama sekolah

X_{12} : Rata-rata lama sekolah

X_{13} : Pengeluaran perkapita

3.2 Heterogenitas Spasial

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

a. Perumusan Hipotesis

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

(Tidak terdapat heterogenitas spasial)

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

b. Stastistik 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 banyak variabel penjelas, 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 penelitan menggunakan pendekatan GWR.

3.3 *Geographically Weighted Lasso*

Geographically Weighted Lasso merupakan gabungan dua konsep, yaitu konsep lasso yang diterapkan dalam suatu pemodelan dengan konsep GWR yang merupakan suatu metode spasial yang digunakan untuk mengatasi masalah heterogenitas pada suatu metode kuadrat terkecil serta masalah adanya multikolinearitas lokal. Dengan menggunakan GWL diharapkan dugaan koefisien parameter yang diperoleh lebih stabil sehingga hasil prediksi yang didapatkan lebih akurat. Pada dasarnya langkah-langkah untuk memodelkan metode GWL adalah dengan menggabungkan langkah-langkah metode GWR dan algoritma LAR pada lasso. Tahap pertama memodelkan dengan menggunakan metode GWR sehingga mendapatkan model regresi lokal, lalu menganalisis multikolinearitas secara lokal.

3.3.1 Penaksiran Parameter GWR

Penaksiran parameter untuk setiap variabel ke-k pada lokasi pengamatan ke-i dinyatakan sebagai berikut

$$\hat{\beta}(u_{i,}v_i) = \left(X^T W(u_{i,}v_i)X\right)^{-1} X^T W(u_{i,}v_i)Y \quad (3.1)$$

Dimana untuk memperoleh penaksir parameter tersebut digunakan *Weighted Least Square*. Pertama persamaan GWR dapat diubah menjadi berikut

$$\varepsilon_i^2 = Y_i - \beta_0(u_i, v_i) - \sum_{k=1}^{p} \beta_k\,(u_i, v_i)X_{ik} \quad (3.2)$$

Lalu tambahkan pembobot w_{ij} dan minimumkan jumlah kuadrat error-nya

$$\sum_{j=1}^{n} w_{ij}\,\varepsilon_i^2 = \sum_{j=1}^{n} w_{ij}\,[Y_i - \beta_0(u_i, v_i) - \sum_{k=1}^{p} \beta_k\,(u_i, v_i)X_{ik}]^2 \quad (3.3)$$

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

$$W(u_i, v_i) = \begin{bmatrix} w_{i1} & 0 & \cdots & 0 \\ 0 & w_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{in} \end{bmatrix} \quad (3.4)$$

$W(u_i, v_i)$ 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 (3.3) dapat ditulis dalam bentuk matriks sebagai berikut:

$$\varepsilon^T W(u_i, v_i)\varepsilon = Y^T W(u_i, v_i)Y - 2\beta^T(u_i, v_i)X^T W(u_i, v_i)Y + \beta^T W(u_i, v_i)X^T W(u_i, v_i)X\beta(u_i, v_i) \quad (3.5)$$

Minimumkan jumlah kuadrat *error*-nya dengan mendiferensialkan persamaan (3.5) terhadap $\beta^T(u_i, v_i)$ sebagai berikut:

$$\frac{\partial}{\partial\beta^T(u_i, v_i)}\varepsilon^T W(u_i, v_i)\varepsilon = 0 - 2X^T W(u_i, v_i)Y + 2X^T W(u_i, v_i)X\hat{\beta}(u_i, v_i) = 0 \quad (3.6)$$

Sehingga diperoleh,

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

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

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

$$X^T W(u_i, v_i)X\hat{\beta}(u_i, v_i) = X^T W(u_i, v_i) \tag{3.7}$$

Kalikan kedua ruas dari sebelah kiri pada persamaan (3.7) dengan invers dari $X^T W(u_i, v_i)X$ sehingga:

$$[X^T W(u_i, v_i)X]^{-1} X^T W(u_i, v_i)X\hat{\beta}(u_i, v_i) = [X^T W(u_i, v_i)X]^{-1} X^T W(u_i, v_i)Y \tag{3.8}$$

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

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

(Maulani, 2013)

3.3.2 Pembobotan *fixed exponential kernel*

Pembobotan merupakan kedekatan titik lokasi ke-i dengan titik lokasi pengamatan lainnya. Fungsi tersebut dinyatakan dalam diagonal matriks dimana elemen-elemen diagonalnya merupakan sebuah fungsi dari setiap titik lokasi pengamatan. Matriks diagonal pembobotan dapat ditulis sebagai berikut :

$$w(u_i, v_i) = \begin{bmatrix} w_{(f)i1} & 0 & \cdots & 0 \\ 0 & w_{(f)i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{(f)in} \end{bmatrix}$$

Dengan :

$w(u_i, v_i)$: matriks pembobot untuk lokasi pengamatan ke-i.

$w_{(f)in}$: nilai dari *fixed exponential kernel* untuk data pada titik ke- n dalam pengujian model di lokasi pengamatan ke-i.

Dengan nilai *bandwidth* menggunakan *fungsi exponential kernel* sebagai berikut :

$$w_f(u_i, v_i) = e^{\left(\frac{-d_{ij}}{h}\right)}$$

Dengan :

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

u_i : koordinat *lattuide* (lintang) pada lokasi ke-i.

v_i : koordinat *longitude* (bujur) pada lokasi ke-i.

h : *bandwidth* pada semua lokasi.

Bandwidth merupakan lingkaran dengan radius h dari titik lokasi pengamtan ke-i. Lokasi pengamtan terdekat dengan lokasi ke-i akan memiliki pengaruh yang lebih besar dalam penaksiran parameter. Pengamatan yang terletak dalam radius akan diboboti sesuai dengan fungsi pembobotan yang digunakan. Sedangkan pengamatan yang terletak di luar radius h akan diboboti nol sehigga tidak mempengaruhi penaksiran paramter. Sehingga pemilihan *bandwidth* optimum sangat pentik untuk dilakukan.

Metode untuk menentukan *bandwidth* optimum adalah validasi silang atau *cross validation* (CV). *Bandwidth* optimum adalah *bandwidth* yang menghasilan nilai CV minimum. Persamaan matematik CV :

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

(Fotheringham & Brunsdon, 2002)

3.3.3 Algoritma *Least Angle Regression* (LAR)

Algoritma LAR merupakan sebuah algoritma untuk menghasilkan model linear yang ditemukan Efron dkk. Pada tahun 2002. Algoritma LAR memberikan jalan yang efisien dalam menyelesaikan regresi LASSO (Tibshirani, 2011). Berikut algoritma LAR :

1. Bakukan variabel penjelas sehingga memilik nilai tengah nol dan varins satu. Mulai dengan residual $r = y - \hat{y}, \beta_0, \beta_1, \beta_2, \ldots, \beta_p = 0$
2. Cari variabel penjelas x_j yang paling berkorelasi dengan r.

3. Ubah nilai β_j dari 0 bergerak menuju koefisien kuadrat terkeci $\langle x_j, r\rangle$ sampai kompetitor lain x_k memiliki korelasi sebesar korelasi x_j dengan sisaaan sekarang.
4. Ubah nilai β_j dan β_k bergerak dalam arah yang didefinisikan oleh koefisien kuadrat terkecil bersama dari sisaan sekarang dalam $\langle x_j, x_k\rangle$ sampai kompetitor x_l lain memiliki korealasi dengan sisaaan sekarang dengan besaran yang sama.
5. Teruskan cara ini sampai semua p variabel penjelas telahmasuk. Setelah min (N-1,p) langkah, solusi model untuk OLS diperoleh.

3.4 Langkah-langkah analisis

1. Mendeskripsikan seluruh variabel
2. Uji asumsi klasik
3. Memodelkan dengan menggunakan regresi linier berganda dengan metode kuadrat terkecil
4. Uji asumsi keheterogenan spasial menggunakan uji *Breusch-Pagan*.
5. Memodelkan dengan menggunakan GWR
   a. Menduga nilai bandwith menggunakan fungsi exponential kernel
   b. Membentuk matriks pembobot
   c. Menduga nilai koefisien dugaan parameter GWR untuk setiap lokasi
6. Mendeteksi multikolinearitas lokal dengan VIF
7. Melakukan pemodelan GWL pada data menggunakan algoritma LAR