

## PEMODELAN DATA KEMISKINAN DI PROVINSI SUMATERA BARAT DENGAN METODE *GEOGRAPHICALLY WEIGHTED REGRESSION (GWR)*

Ilham Maggri<sup>1</sup>, Dwi Ispriyanti<sup>2</sup>

<sup>1</sup>Alumni Jurusan Statistika FSM UNDIP

<sup>2</sup>Staf Pengajar Jurusan Statistika FSM UNDIP

### Abstract

Counting the number of poor have often been modeled as a function of a global regression, which meant that the regression coefficient value applied to all geographic regions. Though this assumption was not always valid because of the differences in geographic locations most likely causing the spatial heterogeneity. In case of spatial heterogeneity, the regression parameters would vary spatially, so if the global regression model was applied, would produce an average value of those regression parameters which vary spatially. This study uses the method Geographically Weighted Regression (GWR) to analyze data that contains spatial heterogeneity. In GWR model estimation, the model parameters are obtained by using the Weighted Least Square (WLS) which gives a different weighting in each location. This study discusses the factors that influence the level of poverty in the province of West Sumatra. Suitability test of the model results shows that there is no influence of spatial factors on the level of poverty in the province of West Sumatra. The results shows that there are four variables that are assumed to affect the level of poverty in the province of West Sumatra, they are the variable of floor space, the facility to defecate, ability to pay the cost of health center / clinic and education levels of household head. The four variables have a similar effect in every city and county.

**Keywords :** Poverty, Spatial Heterogeneity, Geographically Weighted Regression

### 1. Pendahuluan

Permasalahan kemiskinan merupakan salah satu persoalan mendasar yang terus dihadapi di sejumlah daerah di Indonesia, tidak terkecuali Provinsi Sumatera Barat. Berdasarkan berita resmi Badan Pusat Statistika (BPS) Provinsi Sumatera Barat, jumlah penduduk miskin di Provinsi Sumatera Barat mengalami peningkatan dari 429.25 ribu menjadi 430.02 ribu jiwa dari Maret 2009 ke Maret 2010. Beberapa upaya memutus mata rantai kemiskinan telah dilakukan oleh pemerintah pusat dan daerah, diantaranya dengan pemberian beras miskin (Raskin), Bantuan Langsung Tunai (BLT), pelayanan kesehatan keluarga miskin (Askeskin), Bantuan Operasional Sekolah (BOS) dan pemberian akses yang luas terhadap sumber-sumber pembiayaan Usaha Mikro, Kecil dan Menengah (UMKM)<sup>[12]</sup>. Peranan data kemiskinan menjadi sangat penting dalam keberhasilan pelaksanaan program di atas. Berbagai definisi dan indikator untuk mengukur tingkat kemiskinan dan menghitung jumlah penduduk miskin telah diformulasikan dan dikembangkan, dengan harapan upaya pengentasan kemiskinan akan lebih tepat sasaran.

Penghitungan jumlah penduduk miskin selama ini dilakukan dengan pendekatan pemenuhan kebutuhan dasar yang diterjemahkan dengan pendekatan garis kemiskinan pendapatan dan garis kemiskinan konsumsi<sup>[4,13]</sup>. BPS juga mengembangkan suatu metode mendapatkan suatu kriteria yang secara operasional dapat digunakan untuk menentukan jumlah rumah tangga miskin<sup>[3]</sup>. Metode ini digunakan dalam Pendataan Sosial Ekonomi (PSE) tahun 2005 dengan menggunakan 14 indikator variabel untuk menentukan status kemiskinan. Namun dalam kenyataannya, metode penentuan tingkat kemiskinan menurut pendapat tersebut masih bersifat global, artinya diberlakukan untuk semua lokasi yang

diamati. Padahal kenyataannya kondisi lokasi yang satu tidak selalu sama dengan kondisi yang lain, mungkin karena faktor geografis (*spatial variation*), keadaan sosial budaya maupun hal-hal lain yang melatarbelakangi kondisi lokasi yang diamati, sehingga model penentuan tingkat kemiskinan yang bersifat global tidaklah cocok digunakan karena munculnya heterogenitas spasial. Salah satu dampak yang ditimbulkan karena munculnya heterogenitas spasial adalah parameter regresi bervariasi secara spasial. Pada regresi global diasumsikan bahwa nilai duga parameter regresi akan konstan, artinya parameter regresi sama untuk setiap titik di dalam wilayah penelitian. Bila terjadi heterogenitas spasial pada parameter regresi, maka regresi global menjadi kurang mampu dalam menjelaskan fenomena data yang sebenarnya<sup>[1]</sup>.

Brundson, Fotheringham dan Charlton (1998) mengembangkan sebuah metode untuk menganalisis data spasial yang kemudian diberi nama *Geographically Weighted Regression* (GWR). Pada GWR, parameter regresi diasumsikan bervariasi secara spasial. Melalui GWR akan dapat diketahui variasi spasial dalam nilai duga parameter, sehingga akan dihasilkan nilai parameter untuk setiap titik atau lokasi dimana data tersebut diamati. Dalam GWR digunakan matriks pembobot yang besarnya bergantung pada kedekatan antar lokasi pengamatan. Pada penelitian ini, metode GWR akan diaplikasikan untuk menyelidiki variabel-variabel yang berpengaruh terhadap penentuan tingkat kemiskinan di Provinsi Sumatera Barat dengan memperhatikan variasi spasial dalam mengestimasi parameter modelnya. Variabel yang digunakan penelitian ini hanya menggunakan 9 variabel dan menggunakan fungsi pembobot Kernel Gauss serta penentuan bandwidth optimum dengan menggunakan metode *Cross Validation* (CV). Sedangkan tujuan dari penelitian ini adalah menentukan model tingkat kemiskinan dengan memperhatikan variasi spasial di Provinsi Sumatera Barat menggunakan metode GWR.

## 2. Bahan dan Metode

### 2.1. Regresi Linier

Suatu metode yang sering digunakan untuk menyatakan hubungan antara variabel respon dengan variabel prediktor adalah metode regresi. Model regresi linier untuk  $p$  variabel prediktor dan jumlah pengamatan sebanyak  $n$  dapat ditulis:

$$y_i = \beta_0 + \sum_{k=1}^p \beta_k x_{ik} + \varepsilon_i \tag{1}$$

dimana  $i = 1, 2, \dots, n$  dan error-nya diasumsikan identik, independen dan berdistribusi normal dengan mean nol dan varians konstan<sup>[8]</sup>.

Model regresi (1) dapat ditulis dalam bentuk matriks sebagai berikut:

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

dengan

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{12} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \text{ dan } \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

dan penaksir parameternya adalah  $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$  dengan:

- $\hat{\boldsymbol{\beta}}$  : vektor dari parameter yang ditaksir berukuran  $(p+1) \times 1$
- $\mathbf{X}$  : matriks variabel bebas berukuran  $n \times (p+1)$
- $\mathbf{Y}$  : vektor observasi dari variabel respon berukuran  $(n \times 1)$

Model (1) disebut juga model regresi global (*global regression*) karena parameter model regresi tersebut berlaku untuk semua lokasi penelitian<sup>[9]</sup>. Dengan kata lain hubungan antara variabel respon dengan variabel prediktornya diasumsikan konstan untuk semua lokasi dimana data tersebut diamati.

Sedangkan penaksir  $\sigma^2$  didapat berdasarkan penaksir  $\hat{\beta}$ . Berdasarkan asumsi regresi dan menurut model (1), maka nilai  $\sigma^2$  sama untuk setiap  $y_i, i = 1, 2, \dots, n$ , yaitu  $\sigma^2 = E[y_i - \mathbf{x}_i^T \boldsymbol{\beta}]^2$ , sehingga penaksir  $\sigma^2$  adalah  $s^2 = \frac{SSE}{n-p-1}$  dengan

$$SSE = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{Y}^T (\mathbf{I} - \mathbf{H})\mathbf{Y} \text{ dan } \mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$$

Untuk menentukan model regresi terbaik digunakan Statistik Cp, yaitu<sup>[8]</sup>:

$$Cp = \frac{SSE}{s^2} - (n - 2p) \tag{3}$$

Model yang terbaik diperoleh jika statistik Cp mendekati banyaknya parameternya ( $p$ ).

Untuk mengetahui apakah model regresi sesuai dengan data yang digunakan dilakukan uji kesesuaian model secara serentak, dengan hipotesis sebagai berikut:

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

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

Statistik uji dalam pengujian tersebut adalah  $F = \frac{MSR}{MSE}$ . Jika diberikan tingkat signifikansi

sebesar  $\alpha$ , maka  $H_0$  ditolak jika  $F_{hitung} > F_{\alpha; v_1, v_2}$ , dengan  $v_1 = p$  dan  $v_2 = (n - p - 1)$

Sedangkan untuk mengetahui variabel mana saja yang secara statistik signifikan mempengaruhi variabel respon dilakukan uji signifikansi secara parsial, dengan hipotesis adalah sebagai berikut:

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0, \text{ dengan } k = 1, 2, \dots, p$$

Statistik uji yang digunakan adalah  $t = \frac{\hat{\beta}_k}{s(\hat{\beta}_k)}$ ,  $k = 1, 2, \dots, p$ . Jika diberikan tingkat

signifikansi sebesar  $\alpha$ , maka  $H_0$  ditolak jika  $|t_{hitung}| > t_{\alpha/2, df}$  dimana  $df = n - p - 1$ .

Penggunaan model regresi didasarkan kepada asumsi klasik dimana dengan terpenuhinya asumsi-asumsi tersebut penaksiran kuadrat terkecil dari koefisien regresi menjadi yang penaksir yang tidak bias. Pengujian asumsi klasik yang penting adalah uji normalitas residual, uji heterokedasitas, uji multikolinieritas dan uji independensi error<sup>[10]</sup>.

#### 1. Normalitas Residual

Asumsi distribusi normal setiap errornya dibutuhkan sebagai kesahihan dalam pengujian hipotesis model. Pengujian terhadap asumsi kenormalan yang digunakan dalam penelitian ini adalah uji Kolmogorov-Smirnov. Bentuk hipotesis yang digunakan adalah:

$$H_0 : \text{distribusi error mengikuti distribusi normal}$$

$$H_1 : \text{distribusi error tidak mengikuti distribusi normal}$$

Statistik ujinya adalah  $D = \max|F_0(X_i) - S_n(X_i)|$ ,  $i = 1, 2, \dots, n$  dengan  $F_0(X_i)$  merupakan fungsi distribusi frekuensi kumulatif relatif dari teoritis dibawah  $H_0$ .  $S_n(X_i)$  merupakan distribusi frekuensi kumulatif amatan sebanyak sampel. Jika diberikan tingkat

signifikansi sebesar  $\alpha$ , maka diambil keputusan dengan menerima  $H_0$  atau dengan kata lain error mengikuti distribusi normal jika  $D < D(\alpha)$ , dimana  $D(\alpha)$  merupakan nilai kritis yang diperoleh dari tabel Kolmogorov-Smirnov atau jika  $\text{sig} > \alpha$ .

2. Heterokedastisitas

Heterokedastisitas terjadi apabila variasi  $\varepsilon_i$  tidak konstan atau berubah-ubah secara sistematis seiring berubahnya nilai variabel prediktor. Uji statistik yang digunakan adalah uji *Glejser*<sup>[10]</sup>. Uji ini diperoleh dengan meregresikan nilai absolut error dari model yang diperoleh sebelumnya dengan semua variabel prediktor yang digunakan. Jika variabel prediktornya signifikan mempengaruhi nilai absolut error, berarti bahwa dalam data terdapat kasus heterokedastisitas. Sebaliknya, jika tidak signifikan berarti bahwa asumsi homokedastisitas terpenuhi.

3. Multikolinieritas

Multikolinieritas berarti adanya hubungan linier yang sempurna atau pasti di antara beberapa atau semua variabel yang menjelaskan model regresi<sup>[10]</sup>. Untuk mendeteksi ada atau tidaknya gejala multikolinieritas dapat diperiksa dengan menggunakan nilai VIF (*Variance Inflation Factor*). Nilai VIF yang semakin besar akan menunjukkan multikolinieritas yang semakin kompleks. Jika nilai VIF  $< 10$ , maka secara signifikan dapat disimpulkan bahwa tidak terdapat multikolinieritas<sup>[14]</sup>.

4. Independensi Error

Asumsi independensi bertujuan untuk mengetahui apakah dalam model regresi terdapat korelasi antara error pada waktu  $t$  dengan error pada waktu  $t-1$ . Salah satu cara untuk mengetahui apakah error berkorelasi atau tidak adalah dengan pengujian statistik Durbin-Watson. Hipotesis yang digunakan adalah:

$H_0$  : error independen (tidak terjadi autokorelasi)

$H_1$  : error tidak independen (terjadi autokorelasi)

Statistik uji yang digunakan adalah:

$$D = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}$$

dengan  $D$  = nilai Durbin-Watson

$e_i$  = error pada waktu ke- $i$

$e_{i-1}$  = error pada waktu  $i-1$

Jika diberikan tingkat signifikansi sebesar  $\alpha$ , maka dari tabel Durbin-Watson akan diperoleh nilai kritis batas bawah  $D_{\alpha,L}$  dan nilai kritis batas atas  $D_{\alpha,U}$ . Keputusan diambil dengan menerima  $H_0$  atau dengan kata lain error independen jika  $D > D_{\alpha,U}$ . Dan apabila  $D_{\alpha,L} < D < D_{\alpha,U}$ , maka tidak bisa ditarik kesimpulan<sup>[12]</sup>.

**2.2 Geographically Weighted Regression (GWR)**

*Geographically Weighted Regression* (GWR) adalah metode statistika yang digunakan untuk menganalisis heterogenitas spasial<sup>[9]</sup>. Heterogenitas spasial adalah apabila satu variabel bebas yang sama memberikan respon yang tidak sama pada lokasi yang berbeda dalam satu wilayah penelitian<sup>[1]</sup>. Model GWR menghasilkan penaksir parameter model yang bersifat lokal untuk setiap titik atau lokasi dimana data tersebut diamati.

Dalam model GWR, variabel respon  $y$  ditaksir dengan variabel prediktor yang masing-masing koefisien regresinya tergantung pada lokasi dimana data tersebut diamati.

Model GWR dapat ditulis sebagai berikut:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i)x_{ik} + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (4)$$

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

$x_{ik}$  : nilai observasi variabel prediktor  $k$  pada pengamatan ke- $i$

$\beta$  : koefisien regresi

$u_i$  : koordinat spasial *longitude* untuk pengamatan ke- $i$

$v_i$  : koordinat spasial *latitude* untuk pengamatan ke- $i$

$\varepsilon_i$  : error ke- $i$

Dalam pengujian hipotesis ada beberapa asumsi yang digunakan dalam model GWR ini, yaitu<sup>[12]</sup>:

- i. Bentuk error,  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$  diasumsikan independen identik dan mengikuti distribusi normal dengan mean nol dan varian konstan,  $\varepsilon_i \sim IIDN(0, \sigma^2)$
- ii. Misalkan  $\hat{y}_i$  adalah penaksir dari  $y_i$  dilokasi ke- $i$ , maka untuk semua lokasi ( $i = 1, 2, \dots, n$ ),  $\hat{y}_i$  adalah penaksir yang tak bias untuk  $E(y_i)$  atau dapat ditulis  $E(\hat{y}_i) = E(y_i)$  untuk semua  $i$ .

Metode penaksiran parameter pada model GWR adalah dengan metode *Weighted Least Square* (WLS) yaitu dengan memberikan pembobot yang berbeda untuk setiap lokasi dimana data tersebut dikumpulkan. Misalkan pembobot untuk setiap lokasi ke- $i$  adalah  $w_j(u_i, v_i)$   $j = 1, 2, \dots, n$ , maka parameter lokasi  $(u_i, v_i)$  diestimasi dengan menambahkan unsur pembobot pada Persamaan (4) dan kemudian meminimumkan jumlah kuadrat error berikut ini<sup>[6]</sup>:

$$\sum_{j=1}^n w_j(u_i, v_i) \varepsilon_j^2 = \sum_{j=1}^n w_j(u_i, v_i) (y_j - \beta_0(u_i, v_i) - \beta_1(u_i, v_i)x_{j1} - \beta_2(u_i, v_i)x_{j2} - \dots - \beta_p(u_i, v_i)x_{jp})^2$$

$$\text{Misalkan } \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{p1} & x_{p2} & \dots & x_{np} \end{pmatrix}, \quad \mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \boldsymbol{\beta}_1(u_i, v_i) = \begin{pmatrix} \beta_0(u_i, v_i) \\ \beta_1(u_i, v_i) \\ \vdots \\ \beta_p(u_i, v_i) \end{pmatrix}$$

$$\mathbf{W}(u_i, v_i) = \text{diag}[w_1(u_i, v_i), w_2(u_i, v_i), \dots, w_n(u_i, v_i)] \text{ dan } \boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$$

maka:

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

Jika persamaan ini dideferensialkan terhadap matrik  $\boldsymbol{\beta}^T(u_i, v_i)$  dan hasilnya disamakan dengan nol maka diperoleh penaksir parameter dari model GWR untuk setiap lokasi yaitu<sup>[9]</sup>:  $\hat{\boldsymbol{\beta}}(u_i, v_i) = (\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{Y}$

Karena terdapat  $n$  lokasi sampel maka penaksir ini merupakan penaksir setiap baris dari matrik lokal parameter seluruh lokasi penelitian. Matriknya adalah:

$$\beta = \begin{pmatrix} \beta_0(u_1, v_1) & \beta_1(u_1, v_1) & \beta_2(u_1, v_1) & \cdots & \beta_p(u_1, v_1) \\ \beta_0(u_2, v_2) & \beta_1(u_2, v_2) & \beta_2(u_2, v_2) & \cdots & \beta_p(u_2, v_2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \beta_0(u_n, v_n) & \beta_1(u_n, v_n) & \beta_2(u_n, v_n) & \cdots & \beta_p(u_n, v_n) \end{pmatrix}$$

**2.2.1 Koordinat Spasial**

Variabel koordinat spasial *longitude* dan *latitude* merupakan variabel yang digunakan dalam pembobotan dalam pembentukan model GWR. *Longitude* adalah garis membujur yang menghubungkan antara sisi utara dan sisi selatan bumi (kutub) yang digunakan untuk menugukur sisi barat-timur koordinat suatu titik di belahan bumi. Sedangkan *latitude* adalah garis melintang diantara kutub utara dan kutub selatan yang menghubungkan antara sisi timur dan barat bagian bumi yang dijadikan ukuran dalam menugukur sisi utara-selatan koordinat suatu titik di belahan bumi.

**2.2.2 Pemilihan Bandwidth**

*Bandwidth* dapat dianalogikan sebagai radius suatu lingkaran, sehingga sebuah titik yang berada didalam radius lingkaran masih dianggap memiliki pengaruh<sup>[2]</sup>. Nilai *bandwidth* yang optimal menunjukkan seberapa banyak pengamatan yang berpengaruh signifikan dalam pembentukan model GWR<sup>[11]</sup>.

Ada beberapa metode yang digunakan untuk memilih *bandwidth* optimum, salah satunya adalah *Cross Validation (CV)*<sup>[9]</sup>:

$$CV = \sum_{i=1}^n (y_i - \hat{y}_{\neq i}(h))^2 \tag{5}$$

dengan  $\hat{y}_{\neq i}(h)$ : nilai penaksir  $y_i$  (*fitting value*) dimana pengamatan di lokasi  $(u_i, v_i)$  dihilangkan dalam proses penaksiran

**2.2.3 Pemilihan Pembobot (Weight)**

Proses pembobotan dalam penaksiran parameter model GWR mengikuti *Tobler's First Law of Geography*, yaitu data yang lebih dekat dengan lokasi  $i$  akan mempunyai pengaruh yang lebih kuat dalam memprediksi parameter di lokasi ke  $i$  dibandingkan data yang lebih jauh. Ada beberapa literatur yang bisa digunakan untuk menentukan besarnya pembobot untuk masing-masing lokasi yang berbeda pada model GWR, diantaranya:

1. Fungsi Invers Jarak (*Inverse Distance Function*)

Misalkan  $1/d_{ij}$  adalah fungsi invers jarak yang mewakili pembobot antar lokasi

$$(u_i, v_i) \text{ dan lokasi } (u_j, v_j) \text{ dimana } d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2} \tag{6}$$

adalah jarak Euclidian antara lokasi  $(u_i, v_i)$  dan lokasi  $(u_j, v_j)$ .

2. Fungsi Kernel (*Kernel Function*)

Fungsi kernel digunakan untuk mengestimasi parameter dalam model GWR. Pembobot yang terbentuk dengan menggunakan fungsi kernel ini adalah fungsi jarak Gauss (*Gaussian Distance Function*), fungsi *Bisquare*, dan fungsi kernel *tricube*. Fungsi pembobotnya masing-masing dapat ditulis sebagai berikut<sup>[7]</sup>:

- a. Fungsi Kernel Gauss:  $w_j(u_i, v_i) = \exp(-(d_{ij}/h)^2)$  (7)

- b. *Bisquare*:  $w_j(u_i, v_i) = \begin{cases} (1 - (d_{ij}/h)^2)^2, & \text{untuk } d_{ij} \leq h \\ 0, & \text{untuk } d_{ij} > h \end{cases}$  (8)

$$c. \text{ Tricube: } w_j(u_i, v_i) = \begin{cases} \left(1 - (d_{ij}/h)^3\right)^3, & \text{untuk } d_{ij} \leq h \\ 0, & \text{untuk } d_{ij} > h \end{cases} \quad (9)$$

dengan  $d_{ij}$  : jarak antar lokasi  $(u_i, v_i)$  ke lokasi  $(u_j, v_j)$

$h$  : adalah parameter non-negatif yang diketahui dan biasanya disebut parameter penghalus (*bandwidth*).

Jika pembobot yang digunakan adalah fungsi kernel maka pemilihan *bandwidth* sangatlah penting karena *bandwidth* merupakan pengontrol keseimbangan antara kesesuaian kurva terhadap data dan kemulusan data.

#### 2.2.4. Uji Hipotesis Model GWR

##### 1. Pengujian Kesesuaian Model (*Goodness of Fit*)

Pengujian ini dilakukan dengan hipotesis sebagai berikut:

$H_0 : \beta_k(u_i, v_i) = \beta_k, k = 1, 2, \dots, p$  (tidak ada perbedaan yang signifikan antara model regresi global dengan GWR), untuk setiap  $i, i = 1, \dots, n$

$H_0$  : Paling sedikit ada satu  $\beta_k(u_i, v_i)$  yang berhubungan dengan lokasi  $(u_i, v_i)$  (ada perbedaan yang signifikan antara model regresi global dan GWR).

Statistik uji yang digunakan adalah:  $F^* = \frac{SSE(H_0)/df_1}{SSE(H_1)/df_2}$

dengan  $SSE(H_0) = \mathbf{Y}^T (\mathbf{I} - \mathbf{H}) \mathbf{Y}$  dimana  $\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$   $df_1 = n - p - 1$

$SSE(H_1) = \mathbf{Y}^T (\mathbf{I} - \mathbf{S})^T (\mathbf{I} - \mathbf{S}) \mathbf{Y}$ ,  $df_2 = (n - 2tr(\mathbf{S}) + tr(\mathbf{S}^T \mathbf{S}))$

S adalah matriks proyeksi dari model GWR, yaitu matriks yang memproyeksikan nilai  $y$  menjadi  $\hat{y}$  pada lokasi  $(u_i, v_i)$ ,

$$\mathbf{S} = \begin{bmatrix} x_1^T [\mathbf{W}(u_i, v_i) \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i) \\ x_2^T [\mathbf{W}(u_i, v_i) \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i) \\ \vdots \\ x_n^T [\mathbf{W}(u_i, v_i) \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i) \end{bmatrix}$$

adalah matriks  $n \times n$  dan  $\mathbf{I}$  adalah matrik identitas ordo  $n$  dan  $H_0$  ditolak jika nilai

$$F^* > F_{\alpha; df_1, df_2}.$$

##### 2. Pengujian Parameter Model

Pengujian ini dilakukan dengan menguji parameter secara parsial. Pengujian ini dilakukan untuk mengetahui parameter mana saja yang signifikan mempengaruhi variabel responnya. Bentuk hipotesisnya adalah sebagai berikut:

$H_0 : \beta_k(u_i, v_i) = 0$

$H_1 : \beta_k(u_i, v_i) \neq 0; k = 1, 2, \dots, p$

Statistik uji yang digunakan adalah:  $T = \frac{\hat{\beta}_k(u_i, v_i)}{\hat{\sigma} \sqrt{g_{kk}}}$

dengan  $g_{kk}$  adalah elemen diagonal ke- $k$  dari matriks  $\mathbf{G} \mathbf{G}^T$ , dan  $H_0$  ditolak jika

$$|T_{hit}| > t_{\alpha/2; df_2}$$

### 2.3. Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah variabel yang tersedia dalam kuesioner Survei Sosial Ekonomi Nasional (SUSENAS) yang telah dilakukan oleh BPS Provinsi Sumatera Barat tahun 2009 yang berjumlah 9 variabel yang mencirikan rumah tangga miskin per-Kota/Kabupaten. Variabel respon (Y) dan variabel prediktor (X) adalah sebagai berikut:

- $Y$  : Banyaknya rumah tangga yang berada di bawah garis kemiskinan
- $X_1$  : Banyaknya rumah tangga yang luas lantai bangunan tempat tinggal  $\leq 8m^2$
- $X_2$  : Banyaknya rumah tangga yang jenis lantai bangunan tempat tinggal terbuat dari bambu/ tanah/ kayu berkualitas rendah
- $X_3$  : Banyaknya rumah tangga yang jenis dinding tempat tinggal terluasnya terbuat dari bambu/ kayu berkualitas rendah
- $X_4$  : Banyaknya rumah tangga yang tidak punya fasilitas tempat buang air besar atau bersifat umum
- $X_5$  : Banyaknya rumah tangga yang sumber penerangan utama bukan listrik
- $X_6$  : Banyaknya rumah tangga yang tidak sanggup membayar biaya pengobatan di Puskesmas/poliklinik
- $X_7$  : Banyaknya rumah tangga yang sumber air minumnya dari sumur/ mata air tak terlindung/ sungai
- $X_8$  : Banyaknya rumah tangga yang bahan bakar untuk memasaknya dari kayu/minyak tanah
- $X_9$  : Banyaknya rumah tangga yang pendidikan tertinggi kepala rumah tangganya SD kebawah

Selain itu juga digunakan data mengenai koordinat spasial (longitude dan latitude) tiap Kota/Kabupaten yang diperlukan sebagai variabel untuk menentukan pembobot dalam metode GWR. Dalam penelitian ini, observasi yang digunakan adalah Kota/Kabupaten di Provinsi Sumatera Barat yang berjumlah 19 Kota/Kabupaten.

### 2.4 Metode Analisis

Tahapan analisis yang digunakan untuk mencapai tujuan penelitian adalah:

1. Menganalisis model regresi global untuk penentuan tingkat kemiskinan di provinsi Sumatera Barat dengan langkah-langkah sebagai berikut:
  - a. Memodelkan variabel respon (Y) dengan variabel prediktor (X)
  - b. Menguji kesesuaian model regresi linier secara serentak
  - c. Menguji parameter model secara parsial
  - d. Menguji asumsi-asumsi yang harus dipenuhi dalam regresi
2. Menganalisis model GWR untuk penentuan tingkat kemiskinan di Provinsi Sumatera Barat dengan langkah-langkah sebagai berikut:
  - a. Menentukan nilai *bandwidth* untuk tiap daerah yang optimum berdasarkan nilai CV
  - b. Menentukan jarak euclidian antar lokasi pengamatan berdasarkan posisi geografis. Jarak euclidian antara lokasi  $i$  yang terletak pada koordinat  $(u_i, v_i)$  terhadap lokasi  $j$  yang terletak pada koordinat  $(u_j, v_j)$
  - c. Menentukan pembobot dengan menggunakan fungsi *Kernel Gaussian*
  - d. Menaksir parameter model GWR dengan menggunakan metode *Weight Least Square (WLS)*
  - e. Menguji kesesuaian model GWR (*goodness of fit*)
  - f. Menguji parameter model



### 3. Analisis Dan Pembahasan

#### 3.1. Model Regresi Global

Sebagai langkah awal untuk analisis model GWR terlebih dahulu dibentuk model regresi global, yaitu model regresi terbaik antara tingkat kemiskinan dengan faktor-faktor yang mempengaruhinya. Pembentukan model regresi terbaik menggunakan *best subset model*. Metode ini dilakukan dengan memasukkan satu persatu variabel prediktor kedalam model dengan mempertimbangkan nilai  $R^2$ ,  $R^2 - adj$ , statistik Cp dan nilai standar errornya ( $s$ ). Model yang terbaik diperoleh jika penambahan variabel dapat meningkatkan nilai  $R^2$ , mengurangi nilai  $s$  serta statistik Cp mendekati banyaknya parameternya ( $p$ ). *Best subset model* dapat menampilkan model-model terbaik dari kombinasi-kombinasi variabel prediktor yang dimasukkan, sehingga dapat dipilih model terbaik. Dengan menggunakan metode ini diperoleh model dengan delapan variabel prediktor  $X_1, X_2, X_3, X_4, X_6, X_7, X_8$  dan  $X_9$  dengan nilai  $R^2$  sebesar 97.6%, nilai  $R^2 - adj$  sebesar 95.7%, nilai  $s$  sebesar 578.99 serta nilai statistik Cp yang sama dengan banyaknya parameter yaitu 8.0. Namun setelah diuji ternyata tidak memenuhi asumsi. Oleh karena itu, dari beberapa model kombinasi variabel prediktor, model akhir yang diduga adalah dengan memasukkan sebanyak empat variabel prediktor  $X_1, X_4, X_6$  dan  $X_9$  yang menghasilkan nilai  $R^2$  sebesar 96.5%, nilai  $R^2 - adj$  sebesar 95.5%, nilai  $s$  sebesar 592.43 serta nilai statistik Cp yang mendekati banyaknya parameter yaitu 4.2. Model awal regresi yang terbentuk adalah:

$$\hat{Y} = -1 - 5.11X_1 + 0.0649X_4 + 0.0290X_6 + 0.198X_9$$

Untuk mengetahui apakah model tersebut dapat mewakili kondisi sebenarnya atau tidak, maka dilakukan uji kesesuaian model dengan menggunakan nilai F dengan hipotesis:

$$H_0 : \beta_1 = \beta_4 = \beta_6 = \beta_9 = 0$$

$$H_1 : \text{Paling sedikit ada satu } \beta_k \neq 0, k = 1, 4, 6, 9$$

Diperoleh nilai  $F_{hitung} = 96.74 > F_{0.05;7,11} = 3.01$ . Maka dapat diambil keputusan bahwa  $H_0$  ditolak sehingga dapat disimpulkan bahwa model regresi sesuai untuk menggambarkan hubungan antara variabel respon dengan variabel prediktor.

Pengujian signifikansi parameter dilakukan secara parsial, untuk mengetahui variabel mana saja yang secara statistik signifikan mempengaruhi variabel respon dengan hipotesis:

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0, \text{ dengan } k = 1, 4, 6, 9$$

Diperoleh hasil sebagai berikut:

**Tabel 1.** Uji Parameter Model

Prediktor	Koefisien	T	P-value	Kesimpulan
Intercept	-0.9	0.00	0.998	Tidak signifikan
$X_1$	-5.11	-2.15	0.049	Signifikan
$X_4$	0.0649	4.33	0.010	Signifikan
$X_6$	0.0290	3.15	0.007	Signifikan
$X_9$	0.1980	3.74	0.002	Signifikan
S = 592.43 R-Sq = 96.5% R-Sq(adj) = 95.5%				

Sehingga model yang terbentuk adalah model yang tanpa memasukkan  $\beta_0$  (*intercept*) karena tidak signifikan, modelnya adalah:

$$\hat{Y} = -5.11X_1 + 0.0649X_4 + 0.0290X_6 + 0.198X_9$$

Langkah selanjutnya adalah melakukan uji asumsi. Dari uji kenormalan dengan Kolmogorov-Smirnov diperoleh nilai *P-Value* melebihi 0.150 lebih besar dari  $\alpha = 0.05$ , sehingga diperoleh kesimpulan bahwa error model regresi yang terbentuk mengikuti distribusi normal. Sedangkan asumsi homokedastisitas terpenuhi, karena dari hasil pengujian dengan uji *Glejser* menunjukkan bahwa nilai P semua parameter lebih besar dari tingkat signifikansi  $\alpha = 0.05$ . Asumsi non-multikolinieritas juga terpenuhi, karena hasil pengujian menunjukkan bahwa nilai VIF dari semua parameter kurang dari 10. Asumsi yang terakhir yaitu tidak terjadi autokorelasi juga terpenuhi, karena dengan menggunakan nilai statistik Durbin Watson diperoleh nilai  $D = 1.96571 > D_{\alpha,U} = 1.85$ .

### 3.2 Model GWR

Model GWR dibangun berdasarkan empat variabel pada model regresi global. Aplikasi pemodelan GWR pada penelitian ini menggunakan paket program R 2.11.0. Sebagai tahap awal dalam pembentukan model GWR adalah menghitung *bandwidth* dengan menggunakan metode *Cross Validation* (CV) seperti pada Persamaan (5). Nilai *bandwidth* untuk tiap lokasi digunakan untuk membentuk matriks pembobot untuk setiap daerah ke-*i*. Misalkan matrik pembobot di lokasi  $(u_1, v_1)$  adalah  $W(u_1, v_1)$ , maka langkah awal mendapatkan matrik pembobot ini adalah dengan mencari jarak euclidian lokasi  $(u_1, v_1)$  ke semua lokasi penelitian dengan menggunakan Persamaan (6). Setelah itu menghitung pembobot dengan fungsi kernel Gauss dengan menggunakan Persamaan (7). Jarak euclide dan pembobot di lokasi  $(u_1, v_1)$  dapat dilihat pada Tabel 2.

**Tabel 2.** Jarak Euclide dan Pembobot di lokasi  $(u_1, v_1)$

Kota/Kabupaten	Jarak Euclid	Kernel Gauss
Kep. Mentawai	0	1
Pesisir Selatan	1.380818	0.39327
Solok	2.232655	0.01438
Sijunjung	2.130989	0.00941
Tanah Datar	1.595882	0.40889
Padang Pariaman	1.283431	0.49854
Agan	2.258672	0.09421
Lima Puluh Kota	2.374475	0.05002
Pasaman	1.962684	0.11973
Solok Selatan	2.449945	0.04690
Dharmasraya	2.526114	0.29173
Pasaman Barat	1.926381	0.24544
Padang	1.810153	0.02106
Solok	2.616759	0.03831
Sawahlunto	2.345556	0.02110
Padang Panjang	2.316970	0.03755
Bukittinggi	2.329972	0.07171
Payakumbuh	2.558782	0.06374
Pariaman	2.168017	0.10995

Berdasarkan Tabel 2, maka matriks pembobot yang dibentuk dengan fungsi Kernel Gauss di lokasi  $(u_1, v_1)$  adalah:

$$\begin{aligned} \mathbf{W}(u_1, v_1) &= \text{diag}(w_1(u_1, v_1), w_2(u_1, v_1), \dots, w_{19}(u_1, v_1)) \\ &= \text{diag}(1.00000 \quad 0.39327 \quad 0.01438 \quad 0.00941 \quad 0.40889 \\ &\quad 0.49854 \quad 0.09421 \quad 0.05002 \quad 0.11973 \quad 0.04690 \\ &\quad 0.29173 \quad 0.24544 \quad 0.02106 \quad 0.03831 \quad 0.02110 \\ &\quad 0.03755 \quad 0.07171 \quad 0.06374 \quad 0.10995) \end{aligned}$$

Matriks pembobot ini digunakan untuk menaksir parameter di lokasi  $(u_1, v_1)$ . Sedangkan untuk menaksir parameter di lokasi  $(u_2, v_2)$  perlu dicari terlebih dahulu matrik pembobot  $\mathbf{W}(u_2, v_2)$  dengan cara yang sama seperti langkah sebelumnya, demikian seterusnya sampai dengan matrik pembobot pengamatan terakhir  $\mathbf{W}(u_{19}, v_{19})$ . Misalkan  $\hat{\beta}(u_1, v_1)$  adalah penaksir parameter di lokasi  $(u_1, v_1)$ , maka nilai penaksir parameter tersebut adalah:

$$\hat{\beta}(u_1, v_1) = (\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{Y}$$

$$= \left( \begin{bmatrix} 1 & 37 & \dots & 4641 \\ 1 & 13 & \dots & 13805 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 34 & \dots & 2371 \end{bmatrix}^T \mathbf{W}(u_1, v_1) \begin{bmatrix} 1 & 37 & \dots & 4641 \\ 1 & 13 & \dots & 13805 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 34 & \dots & 2371 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 37 & \dots & 4641 \\ 1 & 13 & \dots & 13805 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 34 & \dots & 2371 \end{bmatrix}^T \mathbf{W}(u_1, v_1) \begin{bmatrix} 2553 \\ 7432 \\ \vdots \\ 584 \end{bmatrix}$$

Sehingga diperoleh nilai taksiran parameter di semua lokasi  $(u_i, v_i)$ ,  $i = 1, 2, \dots, 19$  dan diperoleh model seperti terlihat pada Tabel 4.

Setelah diperoleh hasil estimasi dengan menggunakan model regresi global dan model GWR, selanjutnya dilihat ada atau tidaknya perbedaan yang signifikan antara model regresi global dengan model GWR atau dengan kata lain menguji signifikansi faktor spasial pada data kemiskinan di Provinsi Sumatera Barat dan diperoleh hasil sebagai berikut:

**Tabel 3.** Anova Fotheringham

	SS	df	F	P-value
<b>Model Global</b>	4913673	15	1.0084	0.4779
<b>Model GWR</b>	4766625	14.66		

Tabel 3 menunjukkan bahwa  $Pvalue = 0.4799 > \alpha = 0.05$ , maka dapat diambil keputusan tidak ada perbedaan signifikan antara model regresi global dengan model GWR pada data kemiskinan di Provinsi Sumatera Barat. Dengan kata lain, faktor spasial tidak berpengaruh secara signifikan terhadap tingkat kemiskinan di Provinsi Sumatera Barat.

Pengujian parameter model dilakukan untuk mengetahui faktor-faktor yang berpengaruh terhadap tingkat kemiskinan di setiap lokasi  $(u_i, v_i)$ . Dengan menggunakan nilai  $\alpha = 0.05$ , maka nilai  $t_{(0.025;15)} = 2.131$ . Karena nilai  $|t_{hitung}|$  setiap parameter di setiap lokasi lebih besar dari  $t_{(0.025;15)}$ , maka dapat diketahui model terbaik yang terbentuk pada masing-masing Kota/Kabupaten di Provinsi Sumatera Barat. Adapun modelnya dapat dilihat pada Tabel 4.

**Tabel 4.** Model Tiap Kota dan Kabupaten

Kota/Kabupaten	$\hat{Y} = \beta_1 X_1 + \beta_4 X_4 + \beta_6 X_6 + \beta_9 X_9$
Kep. Mentawai	$\hat{Y} = -5.236X_1 + 0.066X_4 + 0.030X_6 + 0.194X_9$
Pesisir Selatan	$\hat{Y} = -5.200X_1 + 0.065X_4 + 0.029X_6 + 0.198X_9$
Solok	$\hat{Y} = -5.114X_1 + 0.064X_4 + 0.029X_6 + 0.199X_9$
Sijunjung	$\hat{Y} = -5.122X_1 + 0.064X_4 + 0.029X_6 + 0.199X_9$
Tanah Datar	$\hat{Y} = -5.230X_1 + 0.065X_4 + 0.029X_6 + 0.199X_9$
Padang Pariaman	$\hat{Y} = -5.213X_1 + 0.065X_4 + 0.029X_6 + 0.198X_9$
Agam	$\hat{Y} = -5.033X_1 + 0.064X_4 + 0.029X_6 + 0.196X_9$
Lima Puluh Kota	$\hat{Y} = -5.045X_1 + 0.064X_4 + 0.029X_6 + 0.196X_9$
Pasaman	$\hat{Y} = -5.060X_1 + 0.064X_4 + 0.029X_6 + 0.196X_9$
Solok Selatan	$\hat{Y} = -5.150X_1 + 0.064X_4 + 0.029X_6 + 0.200X_9$
Dharmasraya	$\hat{Y} = -5.266X_1 + 0.065X_4 + 0.028X_6 + 0.202X_9$
Pasaman Barat	$\hat{Y} = -5.035X_1 + 0.064X_4 + 0.030X_6 + 0.195X_9$
Padang	$\hat{Y} = -5.110X_1 + 0.064X_4 + 0.029X_6 + 0.197X_9$
Solok	$\hat{Y} = -5.082X_1 + 0.064X_4 + 0.029X_6 + 0.199X_9$
Sawahlunto	$\hat{Y} = -5.081X_1 + 0.064X_4 + 0.029X_6 + 0.198X_9$
Padang Panjang	$\hat{Y} = -5.061X_1 + 0.064X_4 + 0.029X_6 + 0.198X_9$
Bukittinggi	$\hat{Y} = -5.039X_1 + 0.064X_4 + 0.029X_6 + 0.197X_9$
Payakumbuh	$\hat{Y} = -5.034X_1 + 0.064X_4 + 0.029X_6 + 0.198X_9$
Pariaman	$\hat{Y} = -5.038X_1 + 0.064X_4 + 0.029X_6 + 0.196X_9$

#### 4. Kesimpulan

Dari hasil analisis dan pembahasan yang telah dilakukan, dapat diperoleh beberapa kesimpulan, yaitu faktor-faktor yang signifikan mempengaruhi tingkat kemiskinan di Provinsi Sumatera Barat adalah luas lantai bangunan tempat tinggal  $\leq 8m^2$ , tidak punya fasilitas tempat buang air besar atau bersifat umum, tidak sanggup membayar biaya pengobatan di Puskesmas/poliklinik dan pendidikan tertinggi kepala rumah tangganya SD kebawah, dimana masing-masing variabel mempunyai pengaruh yang hampir sama untuk setiap daerah, dan model GWR mampu menerangkan keragaman variabel respon sebesar 97.07% dengan jumlah kuadrat error 4766625.

#### DAFTAR PUSTAKA

1. Astutik, S, N.W. Ni Wayan, dan Kurniawan D., *Penggunaan Geographically Weighted Regression Pada Data yang Mengandung Heterokedastisitas Spasial*, Universitas Brawijaya. Malang, 2007.
2. Bailey, T.C. and Gatrell, A.C., *Interactive Spatial Data Analysis*, Prentice Hall, Harlow, 1995.
3. BPS, *Metodologi Penentuan Rumah Tangga Miskin Tahun 2000*, BPS, Jakarta, 2000.
4. BPS, *Analisis dan Penghitungan Tingkat Kemiskinan Tahun 2005*, BPS, Jakarta, 2005.

5. Brundson, C., Fotheringham, A.S. and Charlton, M., *Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity*, 1998.
6. Brundson, C., Fotheringham, A.S., and Charlton, M., , Some Notes on Parametric Significance Test for Geographically Weighted Regression, *J.Reg.Sci.*, 1999, 39: 497-524.
7. Chasco, C., Garcia, I. and Vicens, J., Modelling Spatial Variations in Household Disposable Income with Geographically Weighted Regression, *Munice Personal RePec Arkhiv (MPRA) Working Paper*, 2007, No. 1682.
8. Draper, N. and Smith, H., *Analisis Regresi Terapan*, PT Gramedia Pustaka Utama, Jakarta, 1992.
9. Fotheringham, A.S., Brundson, C. dan Charlton, M., *Geographically Weighted Regression*, John Wiley and Sons, Chichester, UK, 2002.
10. Gujarati, D., *Ekonometrika Dasar*, Erlangga, Jakarta, 1999.
11. Partridge, M.D., Rickman, D.S., Ali, K. and Olfert, M.R., *The Geographic Diversity of U.S. Nonmetropolitan Growth Dynamics A Geographically Weighted Regression Approach*, 2006.
12. Sugiyanto, Analisis Data Spasial Menggunakan Geographically Weighted Regression, *Tesis Program Magister*, Institut Teknologi Sepuluh November, Surabaya, 2008.
13. UNDP, *Kajian Kebutuhan Papua* (Ringkasan Temuan dan Pengaruh terhadap Perumusan Program Bantuan Pembangunan), Kerjasama antara UNDP, BAPPENAS dan BP3D Provinsi Papua, 2005.
14. William, H.H. and Montgomery, D.C., *Probabilita dan Statistika dalam Ilmu Rekayasa Manajemen*, Edisi Kedus, Alih Bahasa: Rudiansyah, Universitas Indonesia Press, Jakarta, 1990.