

APLIKASI MULTIVARIATE GEOGRAPHICALLY WEIGHTED REGRESSION MENGGUNAKAN SOFTWARE MATLAB

Puty Andini ¹⁾, Nar Herrhyanto ²⁾, Maman Suherman ³⁾

^{1), 2), 3)} Departemen Pendidikan Matematika FPMIPA UPI

*Surel: putyandini@upi.edu

ABSTRAK: Persamaan regresi dapat diterapkan pada kasus derajat kesehatan masyarakat. Indikator derajat kesehatan yang dalam penulisan ini adalah angka harapan hidup dan persentase gizi buruk balita. Jika persamaan mengandung unsur spasial maka tidak bisa di selesaikan menggunakan regresi global karena akan menyebabkan kesimpulan yang diambil kurang tepat. Keragaman spasial akibat faktor geografis dapat diselesaikan dengan pendekatan titik. Pendekatan titik yang digunakan dalam penulisan ini adalah *Multivariate Geographically Weighted Regression* (MGWR). Menggunakan dua variabel respon, yaitu angka harapan hidup (Y_1) dan persentase gizi buruk balita (Y_2) dan enam variabel prediktor, yaitu angka melek huruf (X_1), rata-rata lama sekolah (X_2) Indeks kesejahteraan rakyat (X_3), persentase Penduduk dengan sumber air minum PDAM/pipa (X_4), persentase penduduk dengan jarak sumber air minum ke tempat penampungan kotoran > 10 meter (X_5), dan pengeluaran perkapita penduduk untuk makan (X_6). Dalam model MGWR penaksiran parameter diperoleh dengan menggunakan *Weighted Least Square* (WLS). Fungsi pembobotan yang digunakan *fixed kernel Gaussian*. Penelitian ini membuat model derajat kesehatan dengan metode MGWR menggunakan *software* Matlab dan membuat peta sebaran kabupaten dan kota di Jawa Barat berdasarkan variabel-variabel yang signifikan terhadap derajat kesehatan. Variabel yang di duga mempengaruhi angka harapan hidup (Y_1) adalah rata-rata lama sekolah (X_2) dan variabel yang di duga mempengaruhi persentase gizi buruk balita (Y_2) adalah Angka Melek Huruf (X_1).

Kata Kunci: Derajat Kesehatan, *Multivariate Geographically Weighted Regression* (MGWR)

ABSTRACT: The regression equation can be applied in cases of public health degree. Indicators of health, which in this paper are the life expectancy and the percentage of malnourished children under five years old. If the equation contains spatial element it can not be resolved using the global regression because it would lead to inaccurate conclusions drawn. Spatial heterogeneity due to geographical factors can be solved approach point. Point approach used in this paper is Multivariate Geographically Weighted Regression (MGWR) with two response variables, namely life expectancy (Y_1) and the percentage of malnourished children under five (Y_2) and six predictor variables, ie the literacy rate (X_1), the mean the duration of schooling (X_2) Index welfare of the people (X_3), the percentage of residents with drinking water taps / plumbing (X_4), the percentage of people with a distance of drinking water source to the septic tank of > 10 meters (X_5), and spending per capita for the population to eat (X_6). In the model parameter estimation MGWR obtained by using Weighted Least Square (WLS). Weighting function used fixed kernel Gaussian. This study makes a model using methods MGWR health status with Matlab Software and create distribution maps districts / cities in West Java based on the variables that significantly influence health status. Variables are assumed to affect life expectancy (Y_1) is the average length of the school (X_2) and the variables that presumably affect the percentage of malnourished children under five (Y_2) is ie the literacy rate (X_1).

Keywords: Health Status, Multivariate Geographically Weighted Regression (MGWR)

1. PENDAHULUAN

Derajat kesehatan masyarakat dapat dilihat dari berbagai indikator yang meliputi indikator angka harapan hidup, angka kematian, angka kesakitan dan status gizi masyarakat (Indonesia D. K., 2008)

Apabila digunakan dua variabel respon yaitu angkat harapan hidup dan status gizi masyarakat di Jawa Barat dan enam variabel prediktor, analisis regresi multivariat dapat digunakan untuk pemodelan angka harapan hidup dan status gizi masyarakat apabila tidak memasukkan unsur spasial. jika unsur spasial digunakan untuk model regresi klasik akan menyebabkan kesimpulan yang kurang tepat karena asumsi residual saling bebas dan asumsi homogenitas tidak terpenuhi. Oleh karena itu, dibutuhkan metode statistika untuk mengatasi penambahan unsur variabel spasial. Matriks pembobotan spasial dengan pendekatan titik sebagai letak geografis suatu wilayah berdasarkan posisi koordinat garis lintang dan garis bujur dapat mencerminkan hubungan antara lokasi yang satu dengan yang lain. Dengan demikian model statistika regresi spasial multivariat dengan pembobotan geografis atau *MULTIVARIATE GEOGRAPHICALLY WEIGHTED REGRESSION* (MGWR) dapat diterapkan.

2. METODOLOGI

A. Regresi Multivariat

Model regresi multivariat merupakan model regresi linier dengan variabel respon lebih dari satu dan satu atau lebih variabel prediktor (Rencher, 2002).

Pemodelan regresi linier multivariat untuk variabel respon berjumlah q dan variabel prediktor berjumlah p adalah:

$$\begin{aligned} Y_{i1} &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_{i1} \\ Y_{i2} &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_{i2} \\ &\vdots \\ Y_{iq} &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_{iq} \end{aligned} \quad (1)$$

dengan

Y_{iq} : nilai amatan ke- i untuk respon ke- q ,

x_{ip} : nilai amatan ke- i untuk prediktor ke- p ,

β_{0q} : konstanta/Intersep ,

β_p : parameter model regresi multivariat,

ε_{it} : error ($i = 1, 2, \dots, n$) yang diasumsikan $\varepsilon_i \sim (0, \sigma^2)$ dan $cov(\varepsilon_i, \varepsilon_j) = 0$ dengan $i \neq j$.

Secara umum persamaan (1) dapat ditunjukkan dengan matriks dalam persamaan

$$Y_{(n \times q)} = X_{(n \times (p+1))} B_{((p+1) \times q)} + \varepsilon_{(n \times q)} \quad (2)$$

Metode penaksiran untuk menaksir nilai koefisien regresi adalah metode kuadrat terkecil dengan meminimumkan jumlah kuadrat residual (Rencher, 2002).

B. Geographically Weighted Regression (GWR)

Geographically Weighted Regression (GWR) merupakan salah satu kontribusi untuk pemodelan data yang mengandung unsur heterogenitas spasial. Pada metode Geographically Weighted Regression (GWR) parameter dapat ditaksir dimana saja di daerah penelitian (Charlton & Fotheringham, 2009). GWR merupakan pengembangan dari model regresi linier sederhana yang menggabungkan hubungan spasial local kedalam model regresi (Fotheringham, Brunson, & Charlton, 2002). Regresi linier sederhana memiliki parameter konstan pada setiap lokasi pengamatan, sedangkan GWR memiliki taksiran parameter pada setiap lokasi atau bersifat lokal sehingga model regresi untuk titik yang satu berbeda dengan yang lainnya. Model GWR dapat ditulis sebagai berikut:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i) x_{ik} + \varepsilon_i ; i=1,2,3,\dots,n \quad (3)$$

Setiap parameter dihitung pada setiap titik lokasi geografis. Penaksiran parameter model GWR menggunakan metode *Weighted Least Squares* (WLS) yaitu dengan memberikan pembobotan yang berbeda untuk setiap lokasi

pengamatan sebelum diminimalkan (Fotheringham, Brunson, & Charlton, 2002). Penaksir parameter GWR dengan menggunakan metode WLS dapat ditulis sebagai berikut:

$$\hat{\beta}(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) y \quad (4)$$

Menurut Fotheringham, et al (2002), fungsi pembobot yang digunakan pada model GWR diantaranya sebagai berikut :

1) Fixed kernel

Fungsi Fixed Kernel memiliki *bandwidth* yang sama pada setiap titik lokasi pengamatan. Terdapat dua jenis Fungsi Fixed Kernel yaitu:

a. Fungsi kernel Gaussian

$$w_i = e^{-1.5 \left(\frac{d_i}{b} \right)^2} \quad (5)$$

b. Fungsi kernel Bi-Square

$$w_i = \begin{cases} \left[1 - \left(\frac{d_i}{b} \right)^2 \right]^2, & \text{jika } d_i < b \\ 0, & \text{jika } d_i \geq b \end{cases} \quad (6)$$

2) Adaptive Kernel

Fungsi Adaptive Kernel memiliki *bandwidth* yang berbeda pada setiap lokasi pengamatan.

Dengan d_i jarak antara lokasi (u_i, v_i) dan h adalah parameter penghalus (*bandwith*) dengan rumus :

$$d_i = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2} \quad (9)$$

Ada beberapa metode yang digunakan untuk memilih *bandwidth* optimum, *Cross Validation* (CV) atau metode validasi silang merupakan salah satu metode yang digunakan untuk menentukan *bandwidth* optimum. Metode *Cross Validation* dapat ditulis sebagai berikut:

$$C = \sum_{i=1}^n [y_i - \hat{y}_{i \neq i}(b)]^2 \quad (10)$$

Dengan $\hat{y}_{i \neq i}$ merupakan nilai penaksir y_i dimana pengamatan pada titik lokasi ke- i dihilangkan dari proses penaksiran. Nilai *bandwidth* optimum diperoleh saat nilai CV minimum (Fotheringham, Brunson, & Charlton, 2002).

C. *Multivariate Geographically Weighted Regression (MGWR)*

Metode *Multivariate Geographically Weighted Regression* (MGWR) merupakan pengembangan dari model linier spasial dengan penaksir bersifat lokal untuk setiap lokasi pengamatan. Asumsi yang digunakan adalah vektor error (ϵ) berdistribusi normal multivariat dengan mean vektor nol dan matriks varian-kovarian . Apabila terdapat variabel respon Y_1, Y_2, \dots, Y_q dan variabel prediktor X_1, X_2, \dots, X_p pada lokasi ke- i , maka model *Multivariate Geographically Weighted Regression* (MGWR) dapat dinyatakan sebagai berikut:

$$\begin{aligned} Y_{i1} &= \beta_{01}(u_i, v_i) + \beta_{11}(u_i, v_i)X_{i1} + \dots + \beta_{p1}(u_i, v_i)X_{ip} + \epsilon_{i1} \\ Y_{i2} &= \beta_{02}(u_i, v_i) + \beta_{12}(u_i, v_i)X_{i1} + \dots + \beta_{p2}(u_i, v_i)X_{ip} + \epsilon_{i2} \\ &\vdots \\ Y_{iq} &= \beta_{0q}(u_i, v_i) + \beta_{1q}(u_i, v_i)X_{i1} + \dots + \beta_{pq}(u_i, v_i)X_{ip} + \epsilon_{iq} \end{aligned} \quad (11)$$

Dimana

- Y_{ii} : nilai amatan ke- i untuk variabel respon ke- q ,
- X_{ii} : nilai amatan ke- i untuk prediktor ke p ,
- u_i, v_i : koordinat titik lokasi pengamatan ke- i (*longitude, latitude*)
- $\beta_{u,q}(u_i, v_i)$: konstanta/Intercept MGWR
- $\beta_r(u_i, v_i)$: parameter model MGWR
- ϵ_{ii} : error pada titik lokasi ke- i untuk prediktor ke- q

Penaksiran koefisien regresi menggunakan metode kuadrat terkecil terboboti atau *Weighted Least Square* (WLS). Penaksiran tersebut adalah sebagai berikut:

$$\hat{\beta}(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) Y \quad (12)$$

Pengujian keberartian model MGWR dilakukan untuk menentukan model lebih baik secara signifikan dalam memodelkan data. Perumusan hipotesisnya sebagai berikut:

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

H_1 : paling sedikit ada satu $\beta_k(u_i, v_i) \neq \beta_k$; $k = 1, 2, \dots, p$

Atau

H_0 : tidak ada pengaruh faktor geografis pada model

H_1 : ada pengaruh faktor geografis pada model

Statistik uji

$$F_{nli} = \frac{|(Y^T(I-M)Y) - (Y^T(I-S)Y)|}{\frac{n-E-r-1}{|Y^T(I-S)^T(I-S)Y|} \begin{pmatrix} r_1 \\ r_2 \end{pmatrix}} \quad (13)$$

Dimana

$$M = X(X^T X)^{-1} X^T$$

$$S_{(n \times n)} = \begin{bmatrix} X_1^T (X^T W(u_1, v_1) X)^{-1} X^T W(u_1, v_1) \\ X_2^T (X^T W(u_2, v_2) X)^{-1} X^T W(u_2, v_2) \\ \vdots \\ X_n^T (X^T W(u_n, v_n) X)^{-1} X^T W(u_n, v_n) \end{bmatrix}$$

$$r_h = \text{tr}([(I - S)(I - S)]^h) ; h: 1, 2$$

Kriteria pengujian, tolak H_0 jika $F_{h,1} > F_{\left(\alpha, n-p-r_1, \left(\frac{r_1}{r_2}\right)\right)}$

Selanjutnya dilakukan uji serentak pada model MGWR dengan hipotesis sebagai berikut:

$$H_U: \beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \dots = \beta_p(u_i, v_i) = 0$$

$$H_1: \text{paling sedikit ada satu } \beta_k(u_i, v_i) \neq 0$$

Statisik uji

$$F = \frac{\frac{|Y^T (I - S_U)^T (I - S_U) Y|}{\binom{r_1}{r_2}}}{\frac{|Y^T (I - S)^T (I - S) Y|}{\binom{r_1}{r_2}}} \quad (14)$$

Kriteria pengujian, tolak H_0 jika $F > F_{\left(\alpha, \frac{r_1 - r_2}{r_2}, \frac{r_1}{r_2}\right)}$

Untuk mengetahui parameter mana saja yang signifikan mempengaruhi variabel respon dilakukan pengujian signifikansi parameter sebagai berikut:

$$H_U: \beta_k(u_i, v_i) = 0 ; k = 1, 2, \dots, p; h = 1, 2, \dots, q; i = 1, 2, \dots, n$$

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

Statisik uji

$$t = \frac{\hat{\beta}_{k,h}(u_i, v_i)}{S(\hat{\beta}_{k,h}(u_i, v_i))} \quad (15)$$

Dimana $S(\hat{\beta}_{k,h}(u_i, v_i))$ merupakan akar dari varians $\hat{\beta}_{k,h}(u_i, v_i)$

Kriteria pengujian, tolak H_0 jika $|t| > t_{\left(\frac{\alpha}{2}, (n-p-1)\right)}$

Dengan $n-p-1$ merupakan derajat kebebasan Tabel *Distribusi t-Student*.

Adapun langkah-langkah analisis

- 1) Mendeskripsikan variabel respon dan variabel-variabel prediktor kasus kesehatan di Jawa Barat.
- 2) Menganalisis model *Multivariate Geographically Weighted Regression* (MGWR) sebagai berikut:
 - a. Menentukan koordinat *longitude latitude* tiap kabupaten dan kota di Jawa Barat
 - b. Menghitung jarak *euclidean* antar kabupaten dan kota di Jawa Barat.
 - c. Menentukan bandwidth berdasarkan kriteria CV minimum.

- d. Menghitung matriks pembobotan tiap kabupaten dan kota di Jawa Barat dengan fungsi pembobotan Fixed Kernel Gaussian
 - e. Menaksir parameter *Multivariate Geographically Weighted Regression* (MGWR) dengan menggunakan *bandwidth* optimum
- 3) Menginterpretasi dan menyimpulkan hasil yang diperoleh.

3. HASIL DAN PEMBAHASAN

Untuk melakukan pemodelan terlebih dahulu akan di lakukan uji normal multivariat.

Tests of Normality

| | Kolmogorov-Smirnov ^a | | |
|--------|---------------------------------|----|-------------------|
| | Statistic | df | Sig. |
| RES_Y1 | ,088 | 27 | ,200 [*] |
| RLS_Y2 | ,142 | 27 | ,174 |

^{*}. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

Berdasarkan Tabel diatas dengan menggunakan pengujian *Kolmogorov-Smirnov*, *p-value* residual $> = 0,05$. Dengan demikian dapat disimpulkan bahwa H_0 diterima artinya residual data berdistribusi normal multivariat.

Pada pengujian ini karena satuan data setiap variabel tidak sama, maka data distandardisasi atau ditransformasikan kedalam bentuk Z. Dengan software Matlab diperoleh model MGWR sebagai berikut:

$$Y_1 = 0.0001 - 0.1655X_1 + 0.8566X_2 - 0.1981X_3 + 0.052345X_4$$

$$Y_2 = -0,0003 + 0.4515X_1 - 0.48603X_2 - 0.3759X_3 - 0.4523X_4$$

Setelah mendapatkan model, selanjutnya dilakukan uji kesesuaian model dengan bantuan *software Matlab* diperoleh nilai F_H untuk $Y_1 = 2,9974$ dan F_H untuk $Y_2 = 2,9780$ Karena nilai $F_{hitung} Y_1 < F_{tabel} Y_1$ dan $F_{hitung} Y_2 < F_{tabel} Y_2 >$ Nilai $F(0,10, 1, 22) = 2,95$ maka H_0 ditolak artinya ada pengaruh faktor geografis pada model MGWR. Pengujian serentak model MGWR dengan *software Matlab* diperoleh nilai F hitung adalah 2,4386. Dari Tabel Distribusi F diperoleh $F_{(0,05, 2, 22)} = 2,01$, karena $2,4386 > 2,01$ maka H_0 ditolak artinya terdapat variabel prediktor yang berpengaruh signifikan terhadap variabel respon pada setiap kabupaten dan kota di Jawa Barat.

Menggunakan *software Arcgis* diperoleh pemetaan Angka Harapan Hidup dan Persentase Gizi Buruk Balita kabupaten dan kota di provinsi Jawa Barat sebagai berikut :



Gambar 1. Pemetaan AHH terhadap variabel prediktor yang signifikan



Gambar 2. Pemetaan PGBB terhadap variabel prediktor yang signifikan

4. KESIMPULAN

Berdasarkan hasil penelitian yang sudah dilakukan yaitu pemodelan derajat kesehatan di Jawa barat dengan Metode *Multivariate Geographically Weighted Regression* (MGWR) diperoleh beberapa kesimpulan sebagai berikut:

1. Model metode *Multivariate Geographically Weighted Regression* (MGWR) dapat di tulis sebagai berikut

$$Y_{i1} = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)X_1 + \epsilon_1$$

$$Y_{i2} = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)X_1 + \epsilon_2$$

Berdasarkan model diatas didapatkan pemodelan untuk Angka Harapan Hidup dan Persentase Gizi Buruk Balita adalah sebagai berikut

$$\widehat{AAH} = 0,0001 + 0,8566 RLS +$$

dan

$$\widehat{PGBB} = 0,0003 + 0,4515 AMH +$$

2. Pemodelan Angka Harapan Hidup diatas dapat diartikan jika terdapat peningkatan Rata-Rata Lama Sekolah satu satuan, maka Angka Harapan Hidup akan mengalami kenaikan sebesar 0,8566. Pemodelan Persentase Gizi Buruk Balita dapat diartikan jika terdapat peningkatan Angka Melek Huruf satu satuan, maka persentase gizi buruk balita akan mengalami kenaikan sebesar 0,4515.

5. DAFTAR PUSTAKA

- [1] Anselin, L. (2001). *Spatial Econometrics*. Blackwell Publishing Ltd.
- [2] Badan Pusat Statistik Provinsi Jawa Barat. 2014, B. (2014). Jawa Barat Dalam Angka.
- [3] Charlton, M., & Fotheringham, A. S. (2009). *Geographically Weighted Regression*.
- [4] Fotheringham, A. S., Brunson, c., & Charlton, M. (2002). *Geographically Weighted Regression* . Jhon wiley & Sons,LTD.
- [5] Indonesia, D. K. (2008). *Profil Kesehatan Indonesia*.
- [6] Indonesia, D. k. (2008). *Profil Kesehatan indonesia* . dinas kesehatan .
- [7] *Local Statistics and Local Model for Spatial Data*. (t.thn.).
- [8] M.Fariz Fadillah Mardianto, N. W. (2013). Model Regresi Multivariat untuk Menentukan Tingkat Kesejahteraan Kabupaten dan Kota di Jawa Timur.
- [9] Purhadi, D. H. (2013). *Faktor-Faktor yang Mempengaruhi Morbiditas Penduduk Jawa Timur dengan MGWR*.
- [10] Nur Faizah., & Euis Heryati (Agustus 1, 2007). Studi Korelasional Antar status Gizi dengan Prestasi Akademik Pada Siswa Sekolah Dasar Cilampeni I Kabupaten Bandung.
- [11] Depkes, R. I (2009). *Profil Kesehatan Indonesia 2008*. Jakarta.
- [12] Rencher, A. C. (2002). *Methods of Multivariate Analysis*. wiley interscience.
- [13] Sindy Saputri, d. i. (2015). Penentuan Model Kemiskinan di Jawa Tengah dengan *Multivariate Geographically Weighted Regression* (MGWR). *Jurnal Gaussian*, vol 5, nomor 2, 161-170.
- [14] Sindy Saputri. (2015). Penentuan Model Kemiskinan di Jawa tengah dengan *Multivariate Geographically Weighted Regression* (Skripsi). Universitas Diponegoro
- [15] Sri Hariri., & Perhandi (2012). Parameter Estimation of Multivariate Geographically Weighted Regression Model Using Matrix Laboratory.
- [16] *Statistik Kesehatan Republik Indonesia tahun 2009*. (2009). Badan Pusat Statistika Republik Indonesia .
- [17] Stat, B. P. (t.thn.). *Penggunaan Metode GWR Untuk Data Sosial dan Ekonomi*. jakarta.
- [18] Sujana. (2003). Teknik Analisis Regresi dan Korelasi. 5.
- [19] Tenaga kesehatan. (t.thn.). *Undang-Undang Republik indonesia No 36 2014*.
- [20] Tobler. (1970). *Tobler's First Law and Spatial Analysis*. Blackwell Publishing.