

ESTIMASI PARAMETER MODEL *GEOGRAPHICALLY WEIGHTED LOGISTIC REGRESSION*

Fika Dian Lestari, Dadan Kusnandar, Naomi Nessyana Debataraaja

INTISARI

Metode Geographically Weighted Logistic Regression (GWLR) merupakan pengembangan metode regresi logistik dengan mempertimbangkan faktor letak geografis. Faktor letak geografis ini digunakan sebagai pembobot dan menunjukkan sifat lokal pada model GWLR. Metode ini digunakan ketika data memiliki pengaruh heterokedastisitas spasial. Penaksiran parameter model GWLR menggunakan metode Maximum Likelihood Estimation (MLE). Sedangkan pengujian yang dilakukan untuk mengetahui variabel yang signifikan adalah dengan metode uji Wald. Tujuan penelitian ini adalah untuk mengetahui proses estimasi parameter dalam model GWLR dengan pembobot yang digunakan merupakan fungsi kernel. Hasil penelitian menunjukkan model

GWLR untuk jumlah variabel lebih dari dua adalah $g(\mathbf{x}_i) = \beta_0(u_i, v_i) + \sum_{j=1}^p \beta_j(u_i, v_i) x_{ji}$.

Kata Kunci: Regresi logistik, pembobot, faktor letak geografis.

PENDAHULUAN

Analisis regresi merupakan metode statistika yang digunakan untuk menganalisis hubungan antara variabel dependen dan beberapa variabel independen [1]. Metode regresi logistik merupakan salah satu metode statistika yang digunakan untuk mencari hubungan antara variabel dependen yang bersifat kategorik, dengan satu atau lebih variabel independen [2]. Model yang dihasilkan melalui metode regresi logistik yang bersifat global mungkin kurang tepat diaplikasikan pada data yang mengandung unsur heterogenitas spasial. Heterogenitas spasial merupakan keadaan dimana variabel independen yang sama memberikan respon yang tidak sama pada lokasi yang berbeda dalam satu wilayah penelitian. Untuk bisa mengatasi masalah heterogenitas spasial diperlukan data spasial, yaitu data yang memuat informasi geografis suatu daerah, sehingga tidak hanya memuat apa yang diukur [3].

Geographically Weighted Logistic Regression (GWLR) merupakan pengembangan metode regresi logistik dengan mempertimbangkan faktor lokasi. Variabel dependen dalam metode ini bersifat kategorik dengan mempertimbangkan dua kemungkinan atau lebih [2]. Metode GWLR merupakan metode yang tepat digunakan untuk mengatasi heterogenitas spasial, karena model yang dihasilkan metode GWLR ini bersifat lokal. Tujuan dilakukannya penelitian ini adalah untuk mengetahui proses estimasi parameter dalam model *Geographically Weighted Logistic Regression*.

Untuk menganalisis model GWLR hal yang pertama harus dilakukan adalah menentukan nilai u_i dan v_i berdasarkan *longitude* dan *latitude* tiap lokasi penelitian. Kemudian, menghitung jarak *Euclidian* berdasarkan letak geografis. Tahapan selanjutnya yaitu, menentukan nilai *bandwidth* menggunakan metode *Cross Validation (CV)*. Nilai *bandwidth* yang dihasilkan akan berguna untuk menghitung nilai pembobot menggunakan fungsi kernel yang kemudian akan menghasilkan parameter model GWLR, dan tahapan terakhir dalam menganalisis model GWLR adalah melakukan uji signifikansi parameter. Setelah didapatkan model regresi logistik dan GWLR, untuk memilih model terbaik dengan membandingkan nilai AIC terkecil serta presentase ketepatan klasifikasi.

REGRESI LOGISTIK

Regresi logistik merupakan metode yang dapat digunakan untuk mencari hubungan antara variabel dependen yang bersifat *dichotomous* atau *polychotomous*, dengan satu atau lebih variabel independen [2]. Model regresi logistik berganda digunakan pada jumlah variabel independen yang digunakan lebih dari satu. Adapun model regresi logistik berganda didefinisikan sebagai [1]:

$$\pi(\mathbf{x}) = \frac{\exp\left(\beta_0 + \sum_{k=1}^p \beta_k x_k\right)}{1 + \exp\left(\beta_0 + \sum_{k=1}^p \beta_k x_k\right)} \quad (1)$$

Sehingga, bentuk logit dari persamaan model regresi berganda didefinisikan sebagai [1]:

$$g(\mathbf{x}) = \beta_0 + \sum_{k=1}^p \beta_k x_k \quad (2)$$

Penaksiran parameter dalam metode regresi logistik dilakukan dengan menggunakan metode *Maximum Likelihood Estimation* (MLE) yaitu dengan memaksimalkan fungsi likelihood [1].

$$l(\boldsymbol{\beta}) = \prod_{i=1}^n P(Y_i = y_i) = \prod_{i=1}^n \pi(\mathbf{x}_i)^{y_i} [1 - \pi(\mathbf{x}_i)]^{1-y_i} \quad (3)$$

Fungsi *likelihood* yang diperoleh dengan mengasumsikan pada variabel independen, kemudian untuk mendapatkan nilai maksimum dari fungsi *likelihood* dibentuk fungsi *log-likelihood* [1].

$$L(\boldsymbol{\beta}) = \ln l(\boldsymbol{\beta}) = \sum_{i=1}^n \{y_i \ln[\pi(\mathbf{x}_i)] + (1 - y_i) \ln[1 - \pi(\mathbf{x}_i)]\} \quad (4)$$

Nilai maksimum diperoleh dengan menurunkan fungsi *log-likelihood* terhadap β_k dan menyamakannya dengan nol sehingga terdapat $p+1$ model likelihood. Nilai $\boldsymbol{\beta}$ diperoleh melalui suatu prosedur iteratif [1].

$$\begin{aligned} \sum_{i=1}^n [y_i - \pi(\mathbf{x}_i)] &= 0 \\ \sum_{i=1}^n x_{ki} [y_i - \pi(\mathbf{x}_i)] &= 0 \end{aligned} \quad (5)$$

Pengujian parameter model regresi logistik dilakukan dengan menggunakan statistik uji G dengan hipotesis pada uji serentak sebagai berikut [1]:

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

$$H_1 : \text{Paling tidak ada satu } \beta_k \neq 0$$

Statistik uji:

$$G = -2 \ln \left[\frac{\binom{n_1}{n}^{n_1} \binom{n_0}{n}^{n_0}}{\prod_{i=1}^n \hat{\pi}_i^{y_i} (1 - \hat{\pi}_i)^{1-y_i}} \right] \quad (6)$$

dengan $n_1 = \sum_{i=1}^n y_i$; $n_0 = \sum_{i=1}^n (1 - y_i)$; $n = n_0 + n_1$

Kriteria uji yang digunakan adalah tolak H_0 jika nilai $G > \chi_{(\alpha, k)}^2$ dengan k adalah banyaknya variabel independen. Pengujian parameter secara parsial digunakan untuk mengetahui parameter mana yang berpengaruh signifikan terhadap model dengan menggunakan uji Wald. Hipotesis yang digunakan adalah sebagai berikut [1]:

$$W = \frac{\hat{\beta}_k}{SE(\hat{\beta}_k)} \quad (7)$$

dimana $SE(\hat{\beta}_k) = \sqrt{\text{var}(\hat{\beta}_k)}$

dengan kriteria uji yang digunakan adalah tolak H_0 jika nilai $|W| > Z_{\alpha/2}$.

GEOGRAPHICALLY WEIGHTED REGRESSION (GWR)

Geographically Weighted Regression (GWR) merupakan suatu metode statistika yang digunakan untuk menganalisis heterogenitas spasial. Heterogenitas spasial adalah suatu keadaan dimana variabel independen yang sama memberikan respon yang tidak sama pada lokasi yang berbeda dalam suatu wilayah pengamatan. Dalam model GWR, variabel dependen ditaksir dengan variabel independen yang masing-masing koefisien regresinya tergantung pada lokasi dimana data tersebut diamati [3]. Adapun model GWR dapat dituliskan sebagai berikut:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i) x_{ik} + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (8)$$

Proses pemilihan pembobot (*weight*) merupakan hal yang penting untuk dilakukan dalam pemodelan GWR. Peran pembobot pada model GWR sangat penting, karena nilai pembobot ini akan mewakili letak data observasi satu dengan yang lainnya. Metode yang biasa digunakan dalam proses pembobotan dalam model GWR adalah fungsi *kernel* [3].

a. *Fixed Gaussian Kernel*

Metode pembobotan *Fixed Gaussian Kernel* dirumuskan sebagai berikut [3]:

$$w_i(u_i, v_i) = \exp \left[-\frac{1}{2} \left(\frac{d_{ij}}{h} \right)^2 \right] \quad (9)$$

b. *Addaptive Gaussian Kernel*

Metode pembobotan *Addaptive Gaussian Kernel* dirumuskan sebagai berikut [3]:

$$w_i(u_i, v_i) = \exp \left[-\frac{1}{2} \left(\frac{d_{ij}}{h_{i(q)}} \right)^2 \right] \quad (10)$$

c. *Fixed Bisquare Kernel*

Metode pembobotan *Fixed Bisquare Kernel* dirumuskan sebagai berikut [3]:

$$w_i(u_i, v_i) = \begin{cases} \left[1 - \left(\frac{d_{ij}}{h} \right)^2 \right]^2 & , \text{ jika } d_{ij} \leq h \\ 0 & , \text{ jika } d_{ij} > h \end{cases} \quad (11)$$

d. *Addaptive Bisquare Kernel*

Metode pembobotan *Addaptive Bisquare Kernel* dirumuskan sebagai berikut [3]:

$$w_i(u_i, v_i) = \begin{cases} \left[1 - \left(\frac{d_{ij}}{h} \right)^2 \right]^2 & , \text{ jika } d_{ij} \leq h \\ 0 & , \text{ jika } d_{ij} > h \end{cases} \quad (12)$$

Dimana h merupakan nilai *bandwidth* untuk pembobot *Fixed Gaussian Kernel* dimana pada metode ini *bandwidth* yang dihasilkan bersifat tunggal untuk semua lokasi. Sedangkan $h_{i(q)}$ merupakan nilai *bandwidth* untuk pembobot *Addaptive Gaussian Kernel*, dimana pada metode pembobotan ini setiap lokasi pengamatan memiliki *bandwidth* masing-masing. Untuk mencari nilai d_{ij} (Jarak Euclidian) dapat digunakan rumus sebagai berikut:

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2} \quad (13)$$

Kriteria untuk menentukan nilai *bandwidth optimum* dapat diperoleh dengan pendekatan *least square* menggunakan metode *Cross Validation* (CV) yang dirumuskan sebagai berikut [3]:

$$CV(h) = \sum_{i=1}^n (y_i - \hat{y}_{\neq i}(h))^2 \quad (14)$$

dengan $\hat{y}_{\neq i}(h)$ adalah nilai penaksir y_i dimana pengamatan lokasi i dihilangkan dari proses penaksiran.

GEOGRAPHICALLY WEIGHTED LOGISTIC REGRESSION (GWLR)

Geographically Weighted Logistic Regression (GWLR) adalah perpaduan antara GWR dan regresi logistik. GWLR merupakan pengembangan model regresi logistik dengan mempertimbangkan faktor lokasi, dimana variabel dependen dalam penelitian ini bersifat kategorik [5]. Pada model GWLR, lokasi geografis dimasukkan kedalam model melalui fungsi pembobot. Dalam model GWLR model variabel dependen diprediksi dengan variabel independen yang masing-masing koefisien regresinya bergantung pada lokasi dimana data tersebut diamati. Model GWLR dapat ditulis sebagai berikut [5]:

$$\pi(\mathbf{x}_i) = \frac{\exp\left(\beta_0(u_i, v_i) + \sum_{j=1}^p \beta_j(u_i, v_i) x_{ji}\right)}{1 + \exp\left(\beta_0(u_i, v_i) + \sum_{j=0}^p \beta_j(u_i, v_i) x_{ji}\right)} \quad (15)$$

Sedangkan bentuk transformasi logit dari model GWLR adalah sebagai berikut [5]:

$$g(\mathbf{x}_i) = \ln\left[\frac{\pi(\mathbf{x}_i)}{1 - \pi(\mathbf{x}_i)}\right] = \beta_0(u_i, v_i) + \sum_{j=1}^p \beta_j(u_i, v_i) x_{ji} \quad (16)$$

Penaksiran parameter dalam metode GWLR dilakukan dengan menggunakan metode *Maximum Likelihood Estimation* (MLE) yaitu dengan memaksimumkan fungsi likelihood.

$$\begin{aligned} l(\boldsymbol{\beta}(u_i, v_i)) &= \prod_{i=1}^n P(Y_i = y_i) = \prod_{i=1}^n \pi(\mathbf{x}_i)^{y_i} [1 - \pi(\mathbf{x}_i)]^{1-y_i} \\ &= \left\{ \exp \sum_{i=1}^n y_i \left(\sum_{j=1}^p \beta_j(u_i, v_i) x_{ji} \right) \right\} \left\{ \prod_{i=1}^n 1 + \exp \left(\sum_{j=1}^p \beta_j(u_i, v_i) x_{ji} \right) \right\}^{-1} \end{aligned} \quad (17)$$

Setelah didapatkan bentuk *likelihood* kemudian dilakukan operasi logaritma, sehingga bentuk *ln likelihood* nya menjadi :

$$\begin{aligned} L(\boldsymbol{\beta}(u_i, v_i)) &= \ln l(\boldsymbol{\beta}(u_i, v_i)) = \ln \prod_{i=1}^n P(Y_i = y_i) = \ln \prod_{i=1}^n \pi(\mathbf{x}_i)^{y_i} [1 - \pi(\mathbf{x}_i)]^{1-y_i} \\ &= \sum_{i=1}^n y_i \left(\sum_{j=1}^p \beta_j(u_i, v_i) x_{ji} \right) - \sum_{i=1}^n \ln \left(1 + \exp \left(\sum_{j=1}^p \beta_j(u_i, v_i) x_{ji} \right) \right) \end{aligned} \quad (18)$$

Pada model GWLR faktor pembobot yang digunakan adalah faktor letak geografis. Untuk setiap wilayah yang menunjukkan sifat lokal pada model GWLR mempunyai nilai yang berbeda-beda sehingga untuk mendapat model GWLR, pada fungsi *ln likelihood* nya diberikan pembobot ($L^*(\boldsymbol{\beta}(u_i, v_i))$).

$$\begin{aligned} L^*(\boldsymbol{\beta}(u_i, v_i)) &= \sum_{j=0}^p \left(\sum_{m=1}^n w_m(u_i, v_i) y_m x_{jm} \right) \beta_j(u_i, v_i) - \\ &\quad \sum_{m=1}^n w_m(u_i, v_i) \ln \left\{ 1 + \exp \left(\sum_{j=0}^p \beta_j(u_i, v_i) x_{jm} \right) \right\} \end{aligned} \quad (19)$$

Nilai parameter $\boldsymbol{\beta}$ dari turunan fungsi *log-likelihood* yang dihasilkan melalui suatu prosedur iteratif [6]. Pengujian parameter model GWLR secara parsial dilakukan untuk mengetahui variabel independen mana saja yang berpengaruh secara signifikan terhadap variabel dependennya dengan hipotesis pada uji parsial adalah sebagai berikut [7]:

$$H_0 : \beta_j(u_i, v_i) = 0 \quad j = 1, 2, \dots, p$$

$$H_1 : \beta_j(u_i, v_i) \neq 0 \quad j = 1, 2, \dots, p$$

Statistik uji :

$$W = \frac{\hat{\beta}_j(u_i, v_i)}{SE \hat{\beta}_j(u_i, v_i)} \quad (20)$$

Dengan kriteria uji yang digunakan adalah tolak H_0 jika nilai $|W| > Z_{\alpha/2}$.

PEMILIHAN MODEL TERBAIK

Kriteria yang digunakan untuk menentukan model terbaik dalam metode *Geographically Weighted Logistic Regression* (GWLR) adalah berdasarkan nilai *Akaike's Information Criterion* (AIC) dan evaluasi hasil klasifikasi [3].

Akaike Information Criterion (AIC)

Ada beberapa metode yang digunakan untuk memilih model terbaik, salah satu yang paling sering digunakan adalah *Akaike's Information Criterion* (AIC) yang didefinisikan sebagai berikut:

$$AIC(h) = D(h) + 2K(h) \tag{21}$$

Dimana $D(h)$ merupakan nilai devians model dengan *bandwidth* (h) dan $K(h)$ merupakan jumlah parameter dalam model *bandwidth* (h). $AIC(h)$ digunakan karena kompleksitas dari model yaitu perbedaan dari nilai pengamatan dengan nilai dugaan, $AIC(h)$ juga tergantung pada peubah dalam model serta nilai *bandwidth*. Model terbaik adalah model dengan nilai AIC terkecil [3].

Evaluasi Hasil Klasifikasi

Apparent Error Rate (APER) merupakan nilai dari besar kecilnya jumlah observasi yang salah diklasifikasikan oleh model klasifikasi [8]. Tingkat kesalahan klasifikasi disajikan pada Tabel 1 sebagai berikut:

Tabel 1 Kesalahan Klasifikasi

<i>Actual Group Membership</i>	<i>Predicted Group Membership</i>				Total
	Kelompok 1	Kelompok 2	...	Kelompok g	
Kelompok 1	n_{11}	n_{12}	...	n_{1g}	n_1
Kelompok 2	n_{21}	n_{22}	...	n_{2g}	n_2
⋮	⋮	⋮	⋮	⋮	⋮
Kelompok g	n_{g1}	n_{g2}	...	n_{gg}	n_g

Sumber: Johnson dan Wichern, 2007

Tingkat kesalahan dapat dinyatakan dengan

$$APER = \frac{\sum_{i,k=1}^g n_{ik}}{\sum_{i=1}^g n_i}; \quad i, k = 1, 2, \dots, g \tag{22}$$

dimana g adalah banyaknya kelompok dan $i \neq k$.

Sedangkan, untuk tingkat ketepatan klasifikasi (*Actual Error Rate*) dapat dinyatakan dengan

$$AER = 1 - APER \tag{23}$$

KESIMPULAN

Berdasarkan hasil estimasi parameter menggunakan *Maximum Likelihood Estimation* (MLE) model *Geographically Weighted Logistic Regression* untuk lebih dari dua variabel

independen adalah $g(\mathbf{x}_i) = \beta_0(u_i, v_i) + \sum_{j=1}^p \beta_j(u_i, v_i)x_{ji}$.

DAFTAR PUSTAKA

[1]. Hosmer, D.W., dan Lemeshow, S. *Applied Logistic Regression – Second Edition*. John Wiley and Sons. New York, USA; 2000.
 [2]. Rosadi, D. *Analisis Statistika dengan R*. Gadjah Mada University Press. Yogyakarta. 2016

- [3]. Fotheringham, A. S., Brundson, C., dan Charlton, M. *Geographically Weighted Regression*. John Wiley and Sons. Chicester, UK; 2002.
- [4]. Atkinson, P. M., German, S. E., Sear, D. A., dan Clark, M. J. Exploring the Relations Between Riverbank Erosion and Geomorphological Controls Using Geographically Weighted Logistic Regression. *Geographical Analysis*. 2003; 35(1); 58-82.
- [5]. Alburquerque, P. H. M., Medina, F. A. S., dan Silva, A. R. Geographically Weighted Logistic Regression Applied to Scoring Model. *XL ANDAP Congress*. 2016.
- [6]. Faturrahman, M., Puhardi, Sutikno, dan Ratnasari, V. Pemodelan Geographically Weighted Logistic Regression pada Indeks Pembangunan Kesehatan Masyarakat di Provinsi Papua. *Proceeding Seminar Nasional MIPA 2016*. 2016.
- [7]. Desriwendi. Pemodelan Geographically Weighted Logistic Regression (GWLR) dengan Fungsi Pembobot Fixed Gaussian Kernel dan Addaptive Gaussian Kernel (Studi Kasus: Laju Pertumbuhan Penduduk Provinsi Jawa Tengah). *Jurnal Gaussian*. 2015; 4(2); 193-204.
- [8]. Johnson, R. A., and Wichern, W. D. *Applied Multivariate Statistical Analysis 6th ed*. 2007.
- [9]. Caraka, R. E., dan Yasin, H. *Geographically Weighted Regression (GWR) Sebuah Pendekatan Regresi Geografis*. Mobius. Yogyakarta; 2017.

FIKA DIAN LESTARI : Jurusan Matematika FMIPA UNTAN, Pontianak
fkdnlstr@gmail.com

DADAN KUSNANDAR : Jurusan Matematika FMIPA UNTAN, Pontianak
dkusnand@untan.ac.id

NAOMI NESSYANA DEBATARAJA : Jurusan Matematika FMIPA UNTAN, Pontianak
naominessyana@math.untan.ac.id
