ESTIMASI: Journal of Statistics and Its Application
Vol. 4, No. 2, Juli, 2023, Hal. 130-143
DOI: https://doi.org/10.20956/ejsa.v4i2.12250

Pemodelan *Geographically Weighted Logistic Regression* dengan Metode *Ridge*

Reski Amalah[1*], Andi Kresna Jaya[2], dan Nasrah Sirajang[3]
[123]Departemen Statistika, Fakultas MIPA, Universitas Hasanuddin, Kota Makassar, 90245, Indonesia
***Corresponding author, email: amalahreski022@gmail.com**

Abstract

One of the goals of national development is to reduce poverty. Poverty is included in the phenomenon of spatial heterogeneity because it can be shown by the varying economic conditions in each region. The statistical modeling method developed for data analysis takes into account regional factors namely Geographical Weighted Logistic Regression (GWLR). The parameter estimator of the GWLR semiparametric model used in this study was obtained using the Maximum Likelihood Estimation method. In GWLR, the assumption that must be fulfilled is the absence of multicollinearity. One method for dealing with multicollinearity is ridge regression involving the addition of a constant bias (γ). The results obtained were the MSE value of the parameter estimator with the ridge method (707.77) smaller than the GWLR model before using the ridge (715.88). This shows that the ridge method is more effective if there are multicollinearity problems.

***Keywords**: Proverty,* GWLR, *Maximum Likelihood Estimation, ridge, multikolinieritas*

Abstrak

Salah satu sasaran pembangunan nasional adalah menurunkan tingkat kemiskinan. Kemiskinan termasuk dalam fenomena heterogenitas spasial karena dapat ditunjukkan dengan kondisi ekonomi yang bervariasi pada masing-masing wilayah. Metode pemodelan statistika yang dikembangkan untuk analisis data dengan memperhitungkan faktor wilayah yaitu *Geographical Weighted Logistic Regression* (GWLR). Penaksir parameter model semiparametrik GWLR yang digunakan pada penelitian ini diperoleh menggunakan metode *Maximum Likelihood Estimation*. Dalam GWLR, asumsi yang harus dipenuhi yaitu tidak adanya multikolinearitas. Salah satu metode untuk menangani multikolinearitas adalah regresi *ridge* dengan melibatkan penambahan bias konstan (γ). Hasil yang diperoleh adalah nilai MSE dari penduga parameter dengan metode *ridge* (707.77) lebih kecil dibandingkan dengan model GWLR sebelum menggunakan *ridge* (715.88). Hal ini menunjukkan bahwa metode *ridge* lebih efektif digunakan jika terdapat masalah multikolinearitas.

Kata Kunci: Kemiskinan, GWLR, *Maximum Likelihood Estimation, ridge, multikolinieritas*.

1. Pendahuluan

Kemiskinan merupakan gambaran kehidupan di banyak negara berkembang yang mencakup lebih dari satu milyar penduduk dunia. Indonesia adalah negara berkembang, maka masalah kemiskinan merupakan masalah yang penting dan pokok dalam upaya pembangunannya. Berbagai upaya dilakukan oleh pemerintah untuk menanggulangi masalah ini, salah satu upaya yang telah dilakukan oleh Badan Pusat Statistik antara lain adalah melakukan klasifikasi dan penghitungan desa tertinggal. Selain itu untuk mengukur kemiskinan, Badan Pusat Statistik (BPS) menggunakan konsep kemampuan memenuhi kebutuhan dasar (*basic needs approach*). Dengan pendekatan ini, dapat dihitung *Head Count Index* (HCI), yaitu persentase penduduk miskin terhadap total penduduk [1].

Kemiskinan merupakan salah satu fenomena heterogenan spasial. Heterogenitas spasial dapat disebabkan oleh beberapa hal seperti perbedaan dalam geografi, budaya, kebijakan ekonomi yang bervariasi di setiap wilayah. Kemiskinan termasuk dalam heterogenitas spasial karena dapat ditunjukkan dengan kondisi ekonomi yang bervariasi atau berbeda-beda pada masing-masing wilayah. Data kemiskinan dapat berbentuk kategori yaitu miskin dan tidak miskin sehingga untuk menganalisis data kemiskinan dapat digunakan metode statistika yaitu metode *Geographically Weighted Logistic Regression* (GWLR). *Geographically Weighted Logistic Regression* (GWLR) merupakan bentuk lokal dari regresi logistik dengan memperhatikan wilayah yang berasumsi bahwa data berdistribusi Bernoulli. Estimasi parameter model GWLR memerlukan adanya pembobot. Pembobot tersebut dapat dibentuk menggunakan suatu fungsi pembobot yang tergantung pada ukuran berketetanggaan atau biasa disebut *bandwidth* [2]. Penaksir parameter model semiparametrik GWLR yang digunakan pada penelitian ini diperoleh menggunakan metode *maximum likelihood estimation* dengan memberikan pembobot yang berbeda pada setiap wilayah.

Dalam pengaplikasian GWLR terkadang ditemukan adanya data multikolinieritas. Multikolinieritas adalah kondisi terdapatnya hubungan linier atau korelasi yang tinggi antara masing-masing variabel prediktor dalam model regresi. Jika terjadi multikolinieritas maka kondisi tersebut dapat menyebabkan hasil dugaan parameter memiliki ragam yang besar sehingga pengujian signifikansi peubah menjadi tidak stabil. Terdapat beberapa metode yang baik digunakan dalam penyelesaian multikolinieritas pada model GWLR tanpa penghapusan variabel prediktor yaitu Analisis Komponen Utama (AKU)*, Partial Least Square* (PLS), *lasso,* dan *ridge* [3].

Beberapa penelitian yang telah menggunakan metode GWLR antara lain dilakukan oleh Widyastuti (2018) yang melakukan pemodelan terhadap faktor-faktor yang berpengaruh terhadap indeks pembangunan manusia di Kalimantan dengan menggunakan metode *Geographically Weighted Logistic Regression* (GWLR). Selain itu, Haeni (2016) juga melakukan estimasi parameter model GWLR pada data yang mengandung multikolinieritas dengan menggunakan *Partial Least Square* (PLS). Sehingga berdasarkan latar belakang tersebut, peneliti tertarik untuk membahas tentang pemodelan *Geographically Weighted Logistic Regression* dengan metode *ridge*.

2. Material dan Metode

2.1 Sumber Data

Penelitian ini menggunakan data sekunder tentang tingkat kemiskinan di Sulawesi Selatan tahun 2016 yang diperoleh dari www.bps.go.id. Data ini terdiri dari 24 kabupaten/kota. Variabel respon bersifat kategori yaitu dengan mengelompokkan kabupaten/kota menjadi miskin atau tidak miskin berdasarkan pada nilai *Head Count Index* (HCI) Provinsi Sulawesi Selatan. Nilai *Head Count Index* (HCI) Provinsi Sulawesi Selatan tahun 2016 sebesar 9.37%, sehingga pengelompokan wilayah yang tergolong miskin atau tidak miskin yaitu nol untuk kabupaten/kota yang tergolong tidak miskin dengan persentase penduduk miskin berada dibawah 9.37% dan satu untuk kabupaten/kota yang tergolong miskin dengan memiliki persentase penduduk miskin diatas 9.37%.

Variabel prediktor terdiri dari persentase rumah tangga memiliki dinding terluas bukan tembok dan kayu, persentase rumah tangga menurut luas lantai rumah <8m^2, persentase penduduk usia 15 tahun ke atas hanya tamat SD, persentase pengeluaran perkapita untuk makanan, angka harapan hidup, persentase penduduk usia 15 tahun ke atas yang tidak bekerja, dan persentase rumah tangga memiliki sumber penerangan bukan listrik.

2.2 Regresi Logistik Biner

Regresi logistik termasuk dalam model linier diperumum atau *Generalized Linier Models* (GLM). Model linier diperumum merupakan pengembangan dari model linier klasik yang mengasumsikan variabel respon tidak harus mengikuti distribusi normal, tetapi harus termasuk dalam distribusi keluarga eksponensial [4]. Hasil pengamatan variabel acak Y respon hanya mempunyai dua kategori yaitu 0 dan 1, sehingga mengikuti distribusi *Bernoulli* dengan distribusi peluang sebagai berikut:

$$P(Y=y)=\pi^{y}(1-\pi)^{1-y}\,;y=0{,}1 \tag{1}$$

Dalam regresi logistik, variabel respon Y dituliskan sebagai $y=\pi(\boldsymbol{x_i})+\varepsilon$, dengan ε mempunyai salah satu dari kemungkinan dua nilai, yaitu apabila nila $y=1$ maka $\varepsilon=1-\pi(\boldsymbol{x_i})$ dengan peluang $\pi(\boldsymbol{x_i})$ dan jika $y=0$ maka $\varepsilon=-\pi(\boldsymbol{x_i})$ dengan peluang $1-\pi(\boldsymbol{x_i})$. Menurut Agresti (2002), Fungsi distribusi peluang untuk Y dengan parameter $\pi(\boldsymbol{x_i})$ dapat dilihat pada persamaan (2).

$$f(y_i)=\left(\pi(\boldsymbol{x_i})\right)^{y_i}\left(1-\pi(\boldsymbol{x_i})\right)^{1-y_i},y=0{,}1 \tag{2}$$

Menurut Hosmer & Lemashow (2000) model regresi logistik biner dapat dilihat pada persamaan (3).

$$\pi(\boldsymbol{x_i})=\frac{\exp\left(\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\cdots+\beta_kx_{ik}+\cdots+\beta_px_{np}\right)}{1+\exp\left(\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\cdots+\beta_kx_{ik}+\cdots+\beta_px_{np}\right)}$$

$$\pi(\boldsymbol{x_i})=\frac{\exp(\boldsymbol{X_i^t\beta})}{1+\exp(\boldsymbol{X_i^t\beta})} \tag{3}$$

2.3 Multikolinieritas

Pada analisis regresi logistik tidak diperkenankan terjadi kasus multikolinieritas. Metode untuk menguji adanya multikolinieritas dapat dilihat pada *tolerance value* atau V*ariance Inflation Factor* (VIF)[5]. Nilai *Variance Inflation Factor* (*VIF*) dapat digunakan sebagai kriteria untuk mendeteksi kasus multikolinieritas pada model regresi linier yang memiliki lebih dari dua variabel prediktor. Nilai *VIF* dapat dinyatakan dalam persamaan (4).

$$VIF_k = \frac{1}{1-R^2{}_k} \tag{4}$$

dengan ${R_k}^2$ adalah koefisien detrminasi antara x_k dengan variabel prediktor lainnya. Apabila *VIF* > 10, maka dapat diindikasikan bahwa terdapat kasus multikolinieritas.

2.4 Heterogenitas Spasial

Pengujian heterogenitas spasial dilakukan untuk mengetahui apakah terdapat karakteristik atau keunikan sendiri di setiap lokasi pengamatan. Adanya heterogenitas spasial dapat menghasilkan parameter regresi yang berbeda-beda di setiap lokasi pengamatan. Heterogenitas spasial dapat diuji menggunakan rumus *Breusch-Pagan* (BP) dengan Hipotetsis sebagai berikut:

H_0 : ${\sigma_1}^2 = {\sigma_2}^2 = \cdots = {\sigma_n}^2 = \sigma^2$(terjadi homoskedastisitas)

H_1: Minimal ada satu ${\sigma_1}^2 \neq \sigma^2$ (terjadi heteroskedastisitas)

Statistik uji:

$$BP = \left(\frac{1}{2}\right)\boldsymbol{f^T Z}(\boldsymbol{Z^T Z})^{-1}\boldsymbol{Z^T f} \sim \chi^2{}_{(\alpha;k-1)} \tag{5}$$

elemen vektor $\boldsymbol{f}$ adalah

$$f_i = \left(\frac{e_i^2}{\sigma^2}\right) - 1 \; ; e_i = \; y_i - \bar{y}$$

dengan matriks $\boldsymbol{f}$ berukuran $(n \times 1)$, e_i adalah *residual* untuk pengamatan ke-*i*, σ^2 adalah varians *residual*, dan $\mathbf{Z}$ merupakan matriks berukuran $(n(p+1))$ yang berisi vektor standar (z) untuk setiap lokasi [6].

2.5 Geographically Weighted Regression

Geographically Weighted Regression (GWR) merupakan salah satu model yang dimunculkan dari metode pendekatan titik yaitu pendekatan berdasarkan posisi koordinat garis lintang (*latitude*) dan garis bujur (*longitude*). Model *Geographically Weighted Regression* (GWR) adalah pengembangan dari model regresi dimana parameter dihitung pada setiap lokasi pengamatan, sehingga setiap lokasi pengamatan mempunyai nilai parameter yang berbeda-beda. Dengan adanya aspek spasial di dalam GWR, salah satu efek spasial yang terjadi antar wilayah yaitu heterogenitas spasial [7]. Model GWR dapat dinyatakan pada persamaan (6).

$$y_i = \beta_o(u_i, v_i) + \textstyle\sum_{k=1}^{p} \beta_k(u_i, v_i)x_{ik} + \varepsilon_i \tag{6}$$

dengan nilai y_i adalah nilai observasi variabel respon pada pengamatan ke-i, $\beta_o(u_i, v_i)$ adalah nilai konstanta/*intercept* pada pengamatan ke-i, $\beta_k(u_i, v_i)$ adalah koefisien regresi variabel prediktor ke-k pada pengamatan ke-i, x_{ik} adalah nilai observasi variabel prediktor ke-k pada pengamatan ke-i, dan ε_i adalah residual ke-i.

Model GWR membutuhkan sebuah *bandwidth*. Fungsi dari *bandwidth* adalah untuk menentukan bobot dari suatu lokasi terhadap lokasi lain yang digunakan sebagai pusat. Ada beberapa metode yang digunakan untuk memilih *bandwidth* optimum, salah satunya adalah validasi silang atau c*ross validation* (CV) dengan rumus:

$$CV(h) = \sum_{i=1}^{n} \left(y_i - \hat{y}_{\neq i}(h)\right)^2 \tag{7}$$

dengan $\hat{y}_{\neq i}(h)$ yaitu nilai penaksir y_i (*fitting value*) dengan pengamatan di lokasi (u_i, v_i) dihilangkan dalam proses penaksiran, y_i adalah pengamatan ke-*I* dan n adalah jumlah sampel. Untuk mendapatkan nilai h yang optimal, maka diperoleh dari h yang menghasilkan nilai CV yang minimum [8]. Sebelum pembobot ditentukan, harus dihitung dahulu d_{ij} yang merupakan jarak lokasi (u_i, v_i) menggunakan jarak *euclidian* pada persamaan (8)

$$d_{ij} = \sqrt{\left(u_i - u_j\right)^2 + \left(v_i - v_j\right)^2} \tag{8}$$

Pada penelitian ini, fungsi pembobot spasial yang digunakan adalah *fixed Gaussian kernel.* Fungsi *fixed Gaussian kernel* dapat dilihat pada persamaan (9).

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

dengan w_{ij} merupakan nilai bobot dari pengamatan pada lokasi ke-j untuk penaksiran koefisien pada lokasi ke-i, d_{ij} adalah jarak *euclidean* antara lokasi ke-i dengan lokasi ke-j, dan h adalah *bandwidth*.

2.6 *Geographically Weighted Logistic Regression*

Geographically Weighted Logistic Regression (GWLR) merupakan salah satu metode regresi logistik yang dapat mempertimbangkan faktor spasial sehingga akan dihasilkan nilai parameter bagi masing-masing titik atau lokasi dimana data tersebut diamati. Metode ini dikembangkan dari metode GWR yang digunakan untuk memprediksi atau menduga model dari kumpulan data yang memiliki peubah biner melalui model logistik [9]. Pembobot $\boldsymbol{w}_{ij}$ diberikan pada masing-masing pengamatan. Sehingga, model yang terbentuk sesuai dengan model logistik pada persamaan (3) dan dapat dilihat pada persamaan (10):

$$\pi(\boldsymbol{x_i}) = \frac{\exp\left(\beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i)x_{ik}\right)}{1 + \exp\left(\beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i)x_{ik}\right)}$$

$$\pi(\boldsymbol{x_i}) = \frac{\exp\left(\boldsymbol{X}_{\boldsymbol{i}}^t \boldsymbol{\beta}(u_i, v_i)\right)}{1 + \exp\left(\boldsymbol{X}_{\boldsymbol{i}}^t \boldsymbol{\beta}(u_i, v_i)\right)} \tag{10}$$

Pengujian parameter model GWLR digunakan untuk mengetahui parameter yang berpengaruh signifikan terhadap model. Pengujian dilakukan secara serentak dan secara

parsial. Pengujian secara serentak menggunakan uji rasio *likelihood* dengan hipotesisnya sebagai berikut:
$H_0: \beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \cdots = \beta_k(u_i, v_i) = 0$
$H_1: paling\ tidak\ terdapat\ satu\ \beta_k(u_i, v_i) \neq 0$
Statistik Uji:

$$G^2(u_i, v_i) = 2\left\{\sum_{i=1}^{n}\left[y_{1i}\ln\hat{\pi}(\boldsymbol{x_i}) + y_{0i}\ln\left(1 - \hat{\pi}(\boldsymbol{x_i})\right)\right]\right\} - \sum_{i=1}^{n}\left[n_{1i}\ln(n_{1i}) + n_{0i}\ln(n_{0i}) + n\ln(n)\right]$$

Statistik uji G^2 mendekati distribusi *chi-square*. Kriteria pengujiannya adalah tolak H_0 jika nilai $G^2 > \chi^2_{\alpha;p}$ dengan p adalah jumlah prediktor dan nilai $\chi^2_{\alpha;p}$ dapat diperoleh dari table *Chi-square.* Selanjutnya dilakukan parameter model GWLR secara parsial dengan hipotesis parsial dengan hipotesis sebagai berikut:
$H_0: \beta_k(u_i, v_i) = 0$
$H_1: paling\ tidak\ terdapat\ satu\ \beta_k(u_i, v_i) \neq 0, k = 1,2, \dots, p\ ; i = 1,2, \dots, n$

Statistik Uji:

$$W_{hit} = \left(\frac{\hat{\beta}_k(u_i, v_i)}{SE\left(\hat{\beta}_k(u_i, v_i)\right)}\right)^2$$

dengan $SE\left(\hat{\beta}_k(u_i, v_i)\right) = \sqrt{var\ \hat{\beta}_k(u_i, v_i)}$ dan kriteria ujinya yaitu tolak H_0 jika $|W_{hit}| > \chi^2_{\alpha;p}$.

2.7 Metode *Ridge*

Penduga parameter regresi *ridge* menggunakan metode *Least Square* (LS) yaitu dengan menambahkan bilangan positif kecil (γ) pada diagonal matriks $\boldsymbol{X^T X}$, sehingga bias yang terjadi dapat dikendalikan. Penduga regresi *ridge* dapat dituliskan pada persamaan (10).

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{RR}} = (\boldsymbol{X}^T\boldsymbol{X} + \gamma\boldsymbol{I})^{-\mathbf{1}}\boldsymbol{X^T Y} \tag{11}$$

Kemudian penduga *ridge* pada model regresi logistik untuk menangani masalah multikolinieritas diperoleh dengan metode Lagrange untuk meminimumkan fungsi *Weighted Sum of Square Error* (WSSE) berikut:

$$\boldsymbol{WSSE} = (\boldsymbol{Y} - \boldsymbol{X\beta})^{\boldsymbol{T}}\boldsymbol{W}(\boldsymbol{Y} - \boldsymbol{X\beta})$$

dengan fungsi kendala $0 = \gamma\boldsymbol{\beta^T\beta}$, sehingga penduga parameter dengan metode *ridge* pada regresi logistik dapat ditulis pada persamaan (12)

$$\widehat{\boldsymbol{\beta}}_{LRR} = (\boldsymbol{X}^T\boldsymbol{MX} + \gamma\boldsymbol{I})^{-1}\boldsymbol{X}^T\boldsymbol{MX}\widehat{\boldsymbol{\beta}}_{MLE} \quad (12)$$

$\boldsymbol{W}$ pada WSSE kemudian dinotasikan dengan $\boldsymbol{M}$ dan $\widehat{\boldsymbol{\beta}}_{MLE}$ diperoleh diperoleh dari maksimum *likelihood* sehingga mengahasilkan persamaan (12). Untuk nilai tetapan γ, Dorugade dan Kashid (2010) mengajukan nilai k persamaan (13)

$$\gamma = max\left(0, \frac{p\hat{\sigma}^2}{\widehat{\boldsymbol{\beta}}^T\widehat{\boldsymbol{\beta}}} - \left[\frac{1}{n(VIF)_{max}}\right]^2\right) \quad (13)$$

dengan $\hat{\sigma}^2 = \frac{(\boldsymbol{Y}-\boldsymbol{X\beta})^T(\boldsymbol{Y}-\boldsymbol{X\beta})}{n-p-1}$

3. Hasil dan Diskusi

3.1 Uji Multikolinieritas

Metode yang digunakan untuk mengidentifikasi adanya multikolinieritas pada model *Geographically Weighted Logistic Regression* (GWLR) yaitu dengan melihat nilai *Variance Inflation Factors* (VIF). Jika nilai VIF variabel prediktor lebih dari 10 maka telah terjadi multikolinieritas, sedangkan apabila nilai VIF variabel prediktor kurang dari 10 maka dapat dikatakan bahwa tidak terjadi multikolinieritas pada data.

Tabel 1. Hasil pengujian asumsi multikolinieritas

Variabel	VIF	Keterangan
x_1	**24,0590**	Multikolineritas
x_2	6,3627	Tidak Multikolineritas
x_3	3,5567	Tidak Multikolineritas
x_4	2,3370	Tidak Multikolineritas
x_5	6,7663	Tidak Multikolineritas
x_6	**21,3678**	Multikolineritas
x_7	**13,7666**	Multikolineritas

Pada Tabel 1 di atas dapat disimpulkan bahwa data yang digunakan mengandung multikolinieritas dikarenakan nilai VIF pada setiap variabel kurang dari 10.

3.2 Uji Heterogenitas Spasial

Pengujian heterogenitas spasial menggunakan uji BP. Hipotesis yang digunakan adalah sebagai berikut:

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

H_1: Minimal ada satu ${\sigma_1}^2 \neq \sigma^2$

Statistik uji:

$$BP = \left(\frac{1}{2}\right)\boldsymbol{f}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-\mathbf{1}}\boldsymbol{Z}^T\boldsymbol{f}$$

$= 16{,}12$

diperoleh nilai BP = 16,12 > ${\chi^2}_{(6;0{,}05)}$= 12,592 sehingga kriteria keputusan adalah tolak H_0, artinya bahwa terjadi heterogenitas spasial. Oleh karena itu, pemodelan regresi yang digunakan dengan memperhatikan lokasi adalah GWR namun karena data dalam bentuk kategori maka dapat digunakan pemodelan GWLR.

3.3 Penaksir Parameter Model *Geographical Weighted Logistic Regression*

Analisis *Geographically Weigthed Logistic Regression* (GWLR) menggunakan metode MLE dilakukan terlebih dahulu sebelum melakukan pendugaan parameter dengan metode *ridge*.

Perhitungan MLE bergantung pada bentuk distribusi probabilitas populasi yang mendasari variabel diteliti. Digunakan distribusi Bernoulli dengan probabilitas sukses = $\pi(\boldsymbol{x_i})$ dan gagal = $1 - \pi(\boldsymbol{x_i})$ sehingga fungsi kepadatan peluang (fkp) yaitu:

$$f(y_i) = \left(\pi(\boldsymbol{x_i})\right)^{y_i}\left(1 - \pi(\boldsymbol{x_i})\right)^{1-y_i}, y = 0\,,1$$

dari persamaan diatas diperoleh fungsi *likelihood*:

$$l(\beta) = \prod_{i=1}^{n} f(y_i) = \prod_{i=1}^{n} \left(\pi(\boldsymbol{x_i})\right)^{y_i}\left(1 - \pi(\boldsymbol{x_i})\right)^{1-y_i}$$

$$l(\beta) = \prod_{i=1}^{n} \left(\frac{\exp\left(\sum_{k=0}^{p} \beta_k(u_i, v_i)x_{ik}\right)}{1 + \exp\left(\sum_{k=0}^{p} \beta_k(u_i, v_i)x_{ik}\right)}\right)^{y_i} \left(1 - \frac{\exp\left(\sum_{k=0}^{p} \beta_k(u_i, v_i)x_{ik}\right)}{1 + \exp\left(\sum_{k=0}^{p} \beta_k(u_i, v_i)x_{ik}\right)}\right)^{1-y_i}$$

$$= \exp\left\{\sum_{k=0}^{p} \left(\sum_{i=1}^{n} y_i x_{ik}\right) \beta_k(u_i, v_i)\right\} \left\{\prod_{i=1}^{n} \left(1 + \exp\left(\sum_{k=0}^{p} \beta_k(u_i, v_i)x_{ik}\right)^{-1}\right)\right\}$$

Berdasarkan fungsi *likelihood* yang telah didapatkan, maka estimasi parameter model *Geographically Weighted Logistic Regression (*GWLR) dengan menggunakan *Maximum Likelihood Estimator* (MLE), untuk fungsi *likelihood* variabel respon berdistribusi *Bernoulli* sebagai berikut:

$$l\left(\beta(u_i, v_i)\right) = \frac{\exp\left(\sum_{k=1}^{p}\left(\sum_{i=1}^{n} y_i x_{ik}\right) \beta_k(u_i, v_i)\right)}{\prod_{i=1}^{n}\left(1 + \exp\left(\sum_{k=0}^{p} \beta_k(u_i, v_i)x_{ik}\right)\right)}$$

Kemudian bentuk fungsi log *likelihoodnya* menjadi:

$$\ln l\left(\beta(u_i, v_i)\right) = \sum_{k=0}^{p} \left(\sum_{i=1}^{n} y_i x_{ik}\right) \beta_k(u_i, v_i)$$

$$- \sum_{i=1}^{n} \ln\left\{1 + \exp\left(\sum_{k=0}^{p} \beta_k(u_i, v_i)x_{ik}\right)\right\}$$

Faktor lokasi geografis pda model GWLR merupakan faktor yang ditentukan berdasarkan lokasi geografis sehingga memiliki nilai yang berbeda pada setiap lokasi pengamatan dan menghasilkan estimasi parameter disetiap lokasi. Oleh karena itu, pembobot diberikan pada fungsi ln *likelihood* untuk mendapatkan model GWLR. Pada model GWLR dapat dihitung berdasarkan fungsi ln *likelihood* sebagai berikut:

$$\ln l^*\big(\beta(u_i,v_i)\big) = \sum_{k=0}^{p}\left(\sum_{i=1}^{n} y_i x_{ik} w_i(u_i,v_i)\right)\beta_k(u_i,v_i)$$
$$-\sum_{i=1}^{n} w_i(u_i,v_i)\ln\left\{1+\exp\left(\sum_{k=0}^{p}\beta_k(u_i,v_i)x_{ik}\right)\right\}$$

$$\ln l^*\big(\beta(u_i,v_i)\big) = \textstyle\sum_{i=1}^{n}\Big[\big(\sum_{k=0}^{p} x_{ik}\beta_k(u_i,v_i)\big)y_i w_i(u_i,v_i) - w_i(u_i,v_i)\ln\big\{1+\exp\big(\sum_{k=0}^{p}\beta_k(u_i,v_i)x_{ik}\big)\big\}\Big] \quad (4.2)$$

untuk memperoleh estimasi parameter $\beta(u_i,v_i)$, maka dilakukan dengan melakukan turunan parsial yang telah di dapat pada persamaan (4.2) terhadap $\beta_k(u_i,v_i)$ yang akan ditaksir.

Setelah melakukan penaksiran kemudian disamakan dengan nol sehingga diperoleh turunan pertama sebagai berikut:

$$\frac{\partial \ln l^*\big(\beta(u_i,v_i)\big)}{\partial\beta_k(u_i,v_i)}$$
$$= \frac{\partial}{\partial\beta_k(u_i,v_i)}\sum_{i=1}^{n}\left[\sum_{k=0}^{p} y_i x_{ik} w_i(u_i,v_i)\,\beta_k(u_i,v_i)\right.$$
$$\left.- w_i(u_i,v_i)\ln\left\{1+\exp\left(\sum_{k=0}^{p}\beta_k(u_i,v_i)x_{ik}\right)\right\}\right]$$
$$= \sum_{i=1}^{n} x_{ik} w_i(u_i,v_i)\left(y_i - \frac{\big(\exp\big(\sum_{k=0}^{p}\beta_k(u_i,v_i)x_{ik}\big)\big)}{1+\exp\big(\sum_{k=0}^{p}\beta_k(u_i,v_i)x_{ik}\big)}\right)$$

Untuk memaksimumkan fungsi *log likelihood*, maka persamaan diatas disamakan dengan nol, diperoleh:

$$\frac{\partial \ln l^*\big(\beta(u_i,v_i)\big)}{\partial\beta_k(u_i,v_i)} = \sum_{i=1}^{n} x_{ik} w_i(u_i,v_i)\left(y_i - \frac{\big(\exp\big(\sum_{k=0}^{p}\beta_k(u_i,v_i)x_{ik}\big)\big)}{1+\exp\big(\sum_{k=0}^{p}\beta_k(u_i,v_i)x_{ik}\big)}\right) = 0$$

Berdasarkan turunan parsial, diperoleh fungsi yang bersifat *nonlinier.* maka digunakan pendekatan numerik dengan menggunakan metode *fisher scoring*. Karena dalam proses iterasi pada metode *fisher scoring* membutuhkan turunan kedua, maka turunan kedua dari fungsi *log-likelihood* sebagai berikut:

$$\frac{\partial^2 \ln l^*\big(\beta(u_i,v_i)\big)}{\partial\beta_k^2(u_i,v_i)} = -\sum_{i=1}^{n}\left(\frac{\partial}{\partial\beta_k(u_i,v_i)}\left(\frac{x_{ik} w_i(u_i,v_i)\big(\exp\big(\sum_{k=0}^{p}\beta_k(u_i,v_i)x_{ik}\big)\big)}{1+\exp\big(\sum_{k=0}^{p}\beta_k(u_i,v_i)x_{ik}\big)}\right)\right)$$

Dengan menggunakan turunan pertama dan kedua yang telah diperoleh, maka proses perhitungan MLE dapat dapat dilanjutkan dengan didekati bentuk dengan metode *fisher scoring* berikut:

$$\boldsymbol{\beta}^{t+1}(u_i,v_i) = \boldsymbol{\beta}^{(t)}(u_i,v_i) + \boldsymbol{I}^{(t)-1}\big(\beta^{(t)}(u_i,v_i)\big)\boldsymbol{S}^{(t)}\big(\beta^{(t)}(u_i,v_i)\big)$$

dengan:

$$\boldsymbol{I}^{(t)}\left(\beta^{(t)}(u_i,v_i)\right)=\begin{bmatrix} h_{00} & h_{01} & \cdots & h_{0p} \\ h_{10} & h_{11} & \cdots & h_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ h_{p0} & h_{p1} & \cdots & h_{pp} \end{bmatrix}$$

$$\boldsymbol{S}^{(t)}(\beta^{(t)}(u_i,v_i))=\begin{bmatrix} \dfrac{\partial \ln l^*\left(\beta(u_i,v_i)\right)}{\partial \beta_0(u_i,v_i)} \\ \dfrac{\partial \ln l^*\left(\beta(u_i,v_i)\right)}{\partial \beta_1(u_i,v_i)} \\ \vdots \\ \dfrac{\partial \ln l^*\left(\beta(u_i,v_i)\right)}{\partial \beta_p(u_i,v_i)} \end{bmatrix}$$

dan $\boldsymbol{H}^{(t)}\left(\beta^{(t)}(u_i,v_i)\right)$ adalah matriks dengan elemen-elemennya adalah:

$$h_{pp}^{(t)}=-E\left(\frac{\partial^2 \ln l^*\left(\beta(u_i,v_i)\right)}{\partial^2 \beta_k^2(u_i,v_i)}\right)=-E(-\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{M}\boldsymbol{X})$$

Untuk setiap iterasi ke-t berlaku:

$$s_p^{(t)}=\frac{\partial \ln l^*\left(\beta(u_i,v_i)\right)}{\partial \beta_k}=\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}(u_i,v_i)(\boldsymbol{Y}-\boldsymbol{\pi}(\boldsymbol{x}))$$

Dengan mengulang prosedur iterasi untuk setiap lokasi ke-i, maka hasil untuk estimasi parameter lokal akan didapatkan. Iterasi akan berhenti pada saat konvergen, yaitu pada saat $\left\|\boldsymbol{\beta}^{(t+1)}(u_i,v_i)-\boldsymbol{\beta}^{(t)}(u_i,v_i)\right\|\le\varepsilon=0{,}001$.

Nilai penduga MLE yang diperoleh dari hasil iterasi *fisher scoring* menghasilkan parameter disetiap lokasi. Tabel 2 merupakan salah satu contoh hasil estimasi parameter dengan metode MLE untuk Kabupaten Jeneponto, Kabupaten Pangkep, Kabupaten Luwu Utara, dan Kabupaten Toraja Utara.

Tabel 2. Hasil Estimasi Parameter $\left(\hat{\beta}_{MLE}\right)$ dengan Metode MLE

Variabel	**Lokasi Pengamatan**			
	Jeneponto	**Pangkep**	**Luwu Utara**	**Toraja Utara**
Intercept	0,3139	0,3229	0,3637	0,3447
x_1	-0,0013	-0,0013	-0,0013	-0,0013
x_2	-0,0017	-0,0017	-0,0015	-0,0015
x_3	0,0005	0,0005	0,0003	0,0004
x_4	0,0015	0,0014	0,0012	0,0013
x_5	-0,0048	-0,0048	-0,0052	-0,0049
x_6	0,0006	0,0005	0,0003	0,0003
x_7	-0,0129	-0,0126	-0,0114	-0,0118

3.4 Penaksir Parameter Model *GWLR* dengan Metode *Ridge*

Setelah memperoleh hasil penduga parameter dengan metode MLE, selanjutnya menghitung nilai tetapan γ. Pemilihan nilai tetapan γ yang optimum diharapkan agar estimasi $\widehat{\boldsymbol{\beta}}_{\boldsymbol{LRR}}$ lebih optimal dari $\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}}$. Metode yang digunakan dalam pemilihan nilai tetapan γ adalah metode Dorugade dan Kashid pada persamaan (13) dengan hasil yang diperoleh untuk nilai *k* disetiap lokasi sehingga diperoleh hasil estimasi parameter dengan metode *ridge* pada model GWLR. Setelah didapatkan model GWLR dengan metode *ridge,* maka dilakukan uji signifikansi parameter. Uji signifikansi parameter yang dilakukan adalah pengujian serentak dengan menggunakan uji rasio *likelihood* dan pengujian parsial dengan menggunakan uji Wald.

Tabel 3. Pengujian Serentak Model GWLR

G^2	$\chi^2_{(0.05;8)}$
274,890	15,507

Berdasarkan Tabel 3 terlihat bahwa nilai statistik uji G^2sebesar 274.890 lebih besar dari nilai $\chi^2_{0.05;7}$ yaitu sebesaar 15.507. Hal ini menunjukkan bahwa tolak H_0 atau minimal ada satu variabel prediktor yang berpengaruh signifikan terhadap persentase penduduk miskin Kabupaten/Kota di Sulawesi Selatan. Pengujian Hipotesis yang terakhir untuk model GWLR adalah pengujian parsial. Pengujian ini bertujuan untuk mengetahui faktor-faktor yang berpengaruh terhadap persentase penduduk miskin di setiap Kabupaten/Kota di provinsi Sulawsei Selatan. Dari hasil pengujian parsial didapatkan bahwa variabel yang signifikan hanya satu yaitu persentase rumah tangga memiliki sumber penerangan bukan listrik untuk seluruh Kabupaten/Kota yang ada di Sulawesi Selatan. Berdasarkan hasil pengujian, maka dapat diperoleh model *Geographically Weighted Logistic Regression* (GWLR) dengan metode *ridge* untuk pemodelan persentase penduduk miskin Kabupaten/Kota di Sulawesi Selatan yaitu:

Kepulauan Selayar:

$$\pi(\boldsymbol{x}) = \frac{\exp(0{,}0138 - 0{,}013x_7)}{1 + \exp(0{,}0138 - 0{,}013x_7)}$$

Kabupaten Bulukumba:

$$\pi(\boldsymbol{x}) = \frac{\exp(0{,}013 - 0.0121x_7)}{1 + \exp(0{,}013 - 0{,}0121x_7)}$$

$$\vdots$$

Kota Palopo:

$$\pi(\boldsymbol{x}) = \frac{\exp(0{,}0121 - 0{,}0111x_7)}{1 + \exp(0{,}0121 - 0{,}0111x_7)}$$

model transformasi logitnya yaitu:

Kepulauan Selayar:

$$g(\boldsymbol{x}) = \ln\left(\frac{\pi(\boldsymbol{x})}{1-\pi(\boldsymbol{x})}\right) = 0.0138 - 0.013x_7$$

Kabupaten Bulukumba: $g(\boldsymbol{x}) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = 0.013 - 0.0121x_7$

Kabupaten Bantaeng:

$$g(\boldsymbol{x}) = \ln\left(\frac{\pi(\boldsymbol{x})}{1-\pi(\boldsymbol{x})}\right) = 0.0131 - 0.0122x_7$$

$$\vdots$$

Kota Palopo:

$$g(\boldsymbol{x}) = \ln\left(\frac{\pi(\boldsymbol{x})}{1-\pi(\boldsymbol{x})}\right) = 0.0121 - 0.0111x_7$$

Berdasarkan model logit tersebut, salah satu contoh item model GWLR di Kepulauan Selayar yang mempunyai probabilitas masuk kategori wilayah yang bermasalah terhadap kemiskinan apabila terjadi peningkatan terhadap persentase rumah tangga yang memiliki sumber penerangan bukan listrik. Artinya, apabila terjadi kenaikan 1 persen terhadap persentase rumah tangga yang memiliki sumber penerangan bukan listrik (x_7) dapat mengakibatkan resiko meningkatnya persentase penduduk miskin yang ada di Kepulauan Selayar sebesar 1.013 atau 1 kali. Begitupula untuk wilayah-wilayah yang ada di Sulawesi Selatan jika terjadi peningkatan persentase rumah tangga yang memiliki sumber penerangan bukan listrik maka akan mempengaruhi peningkatan persentase penduduk miskin di wilayah tersebut.

3.5 Pemilihan Model Terbaik

Salah satu kriteria yang digunakan untuk menentukan model terbaik antara model GWLR dan model GWLR dengan metode *ridge* adalah dengan melihat nilai MSE. Model terbaik adalah model yang memiliki nilai MSE paling kecil. Nilai MSE GWLR dan GWLR dengan metode *ridge* dapat dilihat pada Tabel 4.

Tabel 4. Nilai MSE GWLR dan AIC GWLR dengan metode Ridge

Model	MSE
GWLR metode ridge	707,77
GWLR	715,88

Berdasarkan Tabel 4 dapat diketahui bahwa model GWLR dengan metode *ridge* mempunyai nilai MSE yang kecil dibandingkan model GWLR yang tidak menggunakan metode *ridge*. Sehingga model GWLR dengan metode *ridge* lebih baik dibandingkan model GWLR. Artinya, setelah ditambahkan metode *ridge* kedalam model GWLR, multikolinieritas telah teratasi dengan metode *ridge* sehingga model yang dihasilkan lebih baik dibandingkan dengan model GWLR sebelum diatasi masalah multikolinieritas.

4. Kesimpulan

Kesimpulan dari hasil dan pembahasan adalah bentuk model *Geographically Weigthed Logistic Regression* dengan metode *ridge* menghasilkan model yang berbeda-beda yang terdiri dari 24 model untuk setiap kabupaten/kota di Sulawesi Selatan, Kemudian uji signifikansi parameter yang dilakukan adalah pengujian serentak dengan menggunakan uji rasio *likelihood* dan pengujian parsial dengan menggunakan uji Wald. Hasil dari pengujian adalah variabel yang berpengaruh secara signifikan pada model GWLR dengan metode *ridge* pada umumnya hanya ada satu variabel yang signifikan yaitu variabel x_7 (persentase rumah tangga memiliki sumber penerangan bukan listrik) untuk seluruh Kabupaten/Kota yang ada di Sulawesi Selatan sehingga terbentuk model di masing-masing wilayah. Berikut adalah model yang terbentuk pada wilayah Makassar yang menjadi salah satu item dari model GWLR dengan metode *ridge:*

$$\pi(\boldsymbol{x}) = \frac{\exp(0{,}0134 - 0{,}0125x_7)}{1 + \exp(0{,}0134 - 0{,}0125x_7)}$$

Daftar Pustaka

[1] Badan Pusat Statistik. *Data dan Informasi Kemiskinan Sulawesi Selatan Tahun 2017.* Makassar: BPS. 2017.

[2] Shara,Y. *Pemodelan Geographically Weighted Regression Dengan Pembobot Fixed Bisquare Kernel Pada Data Spasial (Studi Kasus Balita Gizi Buruk di Provinsi Jawa Timur Tahun 2008).* Malang: Universitas Brawijaya. 2012.

[3] Fitriyaningsih, I., & Sutikno. *Geographically Weighted Lasso dan PCA untuk mengatasi multikolinieritas data spasial (Studi kasus: Perumahan Pondok Indah Jakartas Selatan).* Surabaya: Institut Teknologi Sepuluh Nopember. 2015.

[4] Pradita. *Geographically Weighted Logistic Regression dan Aplikasinya (Studi Kasus: IPM di Provinsi Jawa Timur).* Surabaya: Institut Teknolni Sepuluh Nopember. 2011.

[5] Yan, X., & Su, X. G. *inier Regression Analysis: Theory and Computing.* Singapore: World Scientific. 2009.

[6] Mar'ah, & Zakiyah. *Pemodelan Regresi Terboboti Geografis Semiparametrik Dengan Model Linier Koregionalisasi.* Bogor: Sekolah Pascasarjana, Institut Pertanian Bogor. 2017.

[7] Anselin, L. *Spatial Econometrics.* Dalls: School of Social Science. 2009.

[8] Fotheringham, A., Brundson, C., dan Charlthon, M. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships.* New York: John Wiley and Sons. 2002.

[9] Pravitasary, Hajarisman, dan Sunendiari. *Pemodelan Faktor-faktor yang Berpengaruh Terhadap Angka Buta Huruf di Provinsi Jawa Barat dengan Geographically Weighted Logistic Regression.* Bandung: Universitas Islam Bandung. 2015.