

Penerapan Metode Geographically Weighted Logistic Regression Untuk Memodelkan Pencemaran Air Sungai Mahakam Berdasarkan Data Dissolved Oxygen

Application Of Geographically Weighted Logistic Regression Method For Modeling Mahakam River Water Pollution Based On Dissolved Oxygen Data

Wiwit Widyaningsih¹, Suyitno^{2a)}, dan Qonita Qurrota A'yun³

^{1, 2}Program Studi Statistika, Jurusan Matematika, FMIPA Universitas Mulawarman, Indonesia

³Program Studi Matematika, Jurusan Matematika, FMIPA Universitas Mulawarman, Indonesia

^{a)} Corresponding email: suyitno.stat.unmul@gmail.com

ABSTRACT

The Geographically Weighted Logistic Regression (GWLRL) model is a local model of logistic regression applied to spatial heterogeneity data. Parameter estimation of the GWLRL model is conducted at each observation location using spatial weighting. The aim of this research is to obtain the GWLRL model on the Dissolved Oxygen (DO) data of Mahakam River in 2022, and to identify the factors affecting the probability of Mahakam River water is polluted. The research data is secondary data obtained from Environmental Department of East Kalimantan Province. Spatial weight is calculated using the adaptive bisquare weighting function, and the optimal bandwidth is determined using the Generalized Cross Validation (GCV) criterion. Parameter estimation method is Maximum Likelihood Estimation (MLE), and Maximum Likelihood (ML) estimator was obtained using the iterative Newton-Raphson method. Based on the result of the GWLRL model parameter testing, it was concluded that locally influential factors on the probability of Mahakam River water pollution are nitrate concentration and iron concentration, and globally influential factor is nitrate concentration.

Keywords: adaptive bisquare, DO, GCV, GWLRL, MLE

1. Pendahuluan

Analisis regresi merupakan salah satu analisis dalam statistika yang digunakan untuk menaksir pola hubungan sebab-akibat antara peubah prediktor dan peubah respon. Analisis regresi yang digunakan untuk menganalisis hubungan satu atau beberapa peubah prediktor dengan sebuah peubah respon yang bersifat kategorik biner adalah regresi logistik biner. Regresi logistik biner merupakan model regresi linier tergeneralisasi dengan peubah respon berdistribusi Bernoulli. Kategori 1 pada peubah respon menyatakan hasil percobaan sukses dan 0 untuk hasil percobaan gagal (Alwi dkk, 2018). Data respon biner di lapangan sering kali ditemukan berupa data spasial (Inayah dkk, 2021).

Data spasial terdiri dari data spasial tipe titik atau heterogenitas spasial dan tipe area atau autokorelasi spasial (Hidayanti dkk, 2015). Data heterogenitas spasial dikotomus tidak bisa dimodelkan dengan regresi biasa tetapi dengan *Geographically Weighted Logistic Regression* atau GWLRL. Penaksiran parameter dilakukan pada setiap lokasi pengamatan dengan menggunakan pembobot spasial (Safitri dkk, 2020). Pembobot spasial yang digunakan adalah fungsi pembobot *adaptive bisquare* dan kriteria penentuan *bandwidth* optimum menggunakan *akaike information criterion* (AIC).

Penelitian-penelitian sebelumnya tentang pemodelan dengan GWLRL yaitu penelitian yang dilakukan oleh Inayah dkk (2021) tentang pemodelan *Biochemical Oxygen Demand* (BOD) di Sungai Mahakam menggunakan GWLRL dengan fungsi pembobot *adaptive bisquare* dan diperoleh pemodelan dengan GWLRL lebih baik dari regresi logistik berdasarkan nilai koefisien determinasi. Penelitian yang dilakukan oleh Lestari dkk (2021) tentang pemodelan *Dissolved Oxygen* (DO) di Sungai Mahakam menggunakan GWLRL dengan fungsi pembobot *adaptive tricube* dan diperoleh pemodelan dengan GWLRL lebih baik dari regresi logistik dengan peubah yang diduga berpengaruh terhadap DO adalah konsentrasi amonia, konsentrasi fosfat, Daya Hantar Listrik (DHL), *Total Dissolved Solid* (TDS), konsentrasi nitrit, dan konsentrasi sulfat. Penelitian yang dilakukan oleh Chairina dkk (2020) tentang pemodelan DO di Sungai Mahakam menggunakan regresi weibull univariat dimana peubah yang diduga berpengaruh terhadap DO adalah amonia, suhu, deterjen, nitrat, fenol, dan pH. Model GWLRL pada penelitian ini akan diaplikasikan pada data indikator pencemaran air DO di 33 titik lokasi pengamatan Sungai Mahakam pada tahun 2022.

Sungai Mahakam merupakan sungai terbesar di Kalimantan Timur yang bermuara di Selat Makassar. Masyarakat sekitar memanfaatkan Sungai Mahakam sebagai jalur transportasi air untuk distribusi bahan baku industri (batu bara) dan transportasi penumpang antar wilayah, irigasi pertanian dan perikanan, serta untuk kebutuhan rumah tangga (Rahman dkk, 2017). Aktivitas-aktivitas tersebut berpotensi menghasilkan limbah pada aliran air sungai. Limbah tersebut dapat menyebabkan pencemaran air Sungai Mahakam, oleh karena itu diperlukan tindakan pencegahan. Salah satu bentuk tindakan pencegahan pencemaran air Sungai Mahakam

adalah dengan memberikan informasi kepada masyarakat mengenai faktor-faktor yang dapat berpengaruh terhadap meningkatnya peluang air Sungai Mahakam terindikasi tercemar melalui pemodelan statistika yaitu model GWLR yang diaplikasikan pada data indikator pencemaran air.

Salah satu petunjuk yang dapat digunakan untuk mengetahui kualitas perairan adalah DO. DO adalah kadar oksigen terlarut dalam air. DO diperlukan oleh organisme perairan dalam proses pernapasan dan metabolisme tubuh. Penurunan nilai DO dapat menjadi indikator pencemaran air (Chairina dkk, 2021). Menurut Peraturan Pemerintah Nomor 22 Tahun 2021 tentang Penyelenggaraan Perlindungan dan Pengelolaan Lingkungan Hidup menyatakan kriteria mutu air kelas 1 pada parameter DO sebesar 6 mg/l. Berdasarkan peraturan tersebut, nilai DO dapat dikategorikan menjadi 2 kategori yaitu kategori air tidak tercemar apabila kandungan DO lebih dari sama dengan 6 mg/l dan kategori tercemar apabila kandungan DO kurang dari 6 mg/l.

2. Tinjauan Pustaka

2.1 Model Regresi Logistik Biner

Regresi logistik diaplikasikan pada data respon dikotomis berdistribusi Bernoulli dan merupakan famili distribusi eksponensial dengan fungsi kepadatan peluang (FKP) diberikan oleh persamaan

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

dimana $\pi = P(Y = 1)$ disebut peluang hasil percobaan adalah sukses, dan $1 - \pi = P(Y = 0)$ disebut peluang hasil percobaan adalah gagal. Model regresi logistik merupakan *generalized linear model* (GLM) dan model umum regresi logistik biner yaitu

$$\pi(\mathbf{x}_i) = \frac{\exp(\boldsymbol{\beta}^T \mathbf{x}_i)}{1 + \exp(\boldsymbol{\beta}^T \mathbf{x}_i)} \tag{2}$$

dimana $\pi(\mathbf{x}_i) = P(y_i = 1 | \mathbf{x}_i)$, $\boldsymbol{\beta}^T = [\beta_0 \ \beta_1 \ \dots \ \beta_p]$, dan $\mathbf{x}_i = [1 \ x_{i,1} \ x_{i,2} \ \dots \ x_{i,p}]^T$ (Inayah dkk, 2021).

Penaksiran parameter regresi logistik pada bahasan ini dilakukan dengan metode *Maximum Likelihood Estimation* (MLE). Metode MLE adalah metode penaksiran parameter dengan memaksimalkan fungsi *likelihood* dan fungsi *log-likelihood*. $\hat{\boldsymbol{\beta}}$ yang memaksimalkan fungsi *likelihood* juga memaksimalkan fungsi *log-likelihood*, dan secara matematik $\hat{\boldsymbol{\beta}}$ mudah diperoleh melalui maksimum fungsi *log-likelihood*. Berdasarkan FKP yang diberikan oleh persamaan (1) dan dengan memperhatikan persamaan (2), fungsi *likelihood* didefinisikan oleh

$$L(\boldsymbol{\beta}) = \left(\prod_{i=1}^n (1 + \exp(\boldsymbol{\beta}^T \mathbf{x}_i))^{-1} \right) \exp \left(\sum_{i=1}^n y_i \boldsymbol{\beta}^T \mathbf{x}_i \right). \tag{3}$$

Fungsi *log-likelihood* dari fungsi *likelihood* (3) adalah

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^n y_i \boldsymbol{\beta}^T \mathbf{x}_i - \ln(1 + \exp(\boldsymbol{\beta}^T \mathbf{x}_i)). \tag{4}$$

$\hat{\boldsymbol{\beta}}$ yang memaksimalkan fungsi *log-likelihood* diperoleh dengan menurunkan fungsi (4) terhadap parameter $\boldsymbol{\beta}$ dan disama dengankan nol sebagai berikut

$$\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{0}, \tag{5}$$

dimana $\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ adalah vektor gradien berdimensi $(p + 1)$ dan $\mathbf{0}$ adalah vektor nol yang berdimensi $(p + 1)$.

Dengan memperhatikan ekspresi persamaan (4), persamaan (5) adalah nonlinier sehingga solusi eksak untuk mendapatkan penaksir eksak maximum likelihood (ML) model regresi logistik tidak dapat diperoleh secara analitik. Persamaan (5) dapat diselesaikan secara numerik, dalam penelitian ini menggunakan metode iteratif Newton-Raphson. (Inayah dkk, 2021).

Pengujian parameter model regresi logistik biner terdiri dari pengujian parameter secara simultan dan parsial. Pengujian secara simultan dilakukan untuk mengetahui signifikansi peubah prediktor secara keseluruhan. Hipotesis pengujian secara simultan adalah

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \dots = \beta_p = 0$$

$$H_1 : \text{Minimal ada satu } \beta_k \neq 0; k = 1, 2, 3, \dots, p$$

Statistik uji diberikan oleh

$$G = 2(\ell(\hat{\boldsymbol{\Omega}}) - \ell(\hat{\boldsymbol{\omega}})) \tag{6}$$

dengan $\ell(\hat{\boldsymbol{\Omega}})$ dan $\ell(\hat{\boldsymbol{\omega}})$ diperoleh dari

$$\ell(\hat{\Omega}) = \sum_{i=1}^n y_i \hat{\beta}^T \mathbf{x}_i - \ln(1 + \exp(\hat{\beta}^T \mathbf{x}_i)) \tag{7}$$

dan

$$\ell(\hat{\omega}) = (n_1 \ln(n_1) + n_0 \ln(n_0) - n \ln(n)), \tag{8}$$

dimana $n_0 = \sum_{i=1}^n (1 - y_i)$, $n_1 = \sum_{i=1}^n y_i$, dan $n = n_0 + n_1$. Daerah penolakan H_0 pada taraf uji α adalah jika $G > \chi^2_{(\alpha, p)}$ atau $p_{value} < \alpha$ (Inayah dkk, 2021), dengan G diberikan oleh persamaan (6).

Pengujian secara parsial dilakukan untuk mengetahui signifikansi peubah prediktor secara individual. Hipotesis pengujian secara parsial untuk β_k dengan $k = 1, 2, 3, \dots, p$ adalah

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0$$

Statistik uji diberikan oleh

$$W = \frac{\hat{\beta}_k}{\sqrt{\text{Var}(\hat{\beta}_k)}} \tag{9}$$

Daerah penolakan H_0 pada taraf uji α jika $|W| > Z_{\alpha/2}$ atau jika $p_{value} < \alpha$ (Inayah dkk, 2021).

2.2 Pendeteksian Multikolinieritas

Multikolinieritas adalah suatu kondisi dimana terjadi korelasi kuat antara peubah prediktor dalam model atau dapat dikatakan bahwa antar peubah prediktor tidak saling bebas. Ukuran yang dapat digunakan untuk mendeteksi multikolinieritas adalah *Variance Inflation Factor* (VIF). Nilai VIF lebih besar dari 10 mengindikasikan adanya masalah multikolinieritas. Nilai VIF dihitung merujuk pada Sriningsih dkk, 2018.

2.3 Heterogenitas Spasial

Heterogenitas spasial terjadi apabila terdapat karakteristik yang berbeda di setiap lokasi pengamatan sehingga parameter regresi yang dihasilkan berbeda-beda secara spasial (Hidayanti dkk, 2015). Salah satu pengujian heterogenitas spasial adalah menggunakan uji *Glejser* dengan hipotesis sebagai berikut

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2$$

$$H_1 : \text{Minimal ada satu } \sigma_i^2 \neq \sigma^2; i = 1, 2, \dots, n$$

Statistik uji diberikan oleh

$$F = \frac{(\hat{\mathbf{a}}^T \mathbf{X}^T \mathbf{e} - n \bar{e}^2) / p}{(\mathbf{e}^T \mathbf{e} - \hat{\mathbf{a}}^T \mathbf{X}^T \mathbf{e}) / (n - p - 1)}, \tag{10}$$

dimana $\mathbf{e} = [|\hat{e}_1| \quad |\hat{e}_2| \quad \dots \quad |\hat{e}_n|]$, $|\hat{e}_i| = |\hat{\pi}_i - \bar{Y}|$, dan $\mathbf{a} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{e}$. Daerah penolakan H_0 pada taraf uji α jika nilai $F_{hitung} > F_{(\alpha, p, (n-p-1))}$ atau jika $p_{value} < \alpha$ yang berarti terjadi heterogenitas spasial (Inayah dkk, 2021).

2.4 Model Geographically Weighted Logistic Regression

Model *Geographically Weighted Logistic Regression* (GWLR) merupakan model lokal dari regresi logistik dimana penaksiran parameter dilakukan di setiap lokasi pengamatan dengan menggunakan pembobot spasial. Pembobot spasial dihitung menggunakan fungsi pembobot spasial, dan salah satu pembobot spasial adalah fungsi pembobot *adaptive bisquare* (Tizona dkk, 2017). Salah satu kriteria penentuan bandwidth optimum adalah *Generalized Cross-Validation* atau GCV (Inayah dkk, 2021). Model GWLR pada lokasi pengamatan ke- i dengan koordinat (u_i, v_i) adalah (Pratiwi dkk, 2020)

$$\pi_i(u_i, v_i) = \frac{\exp(\beta^T(u_i, v_i) \mathbf{x}_i)}{1 + \exp(\beta^T(u_i, v_i) \mathbf{x}_i)}, i = 1, 2, \dots, n. \tag{11}$$

Penaksiran parameter model GWLR pada penelitian ini menggunakan metode MLE. Berdasarkan persamaan (3) maka fungsi *likelihood* yang diberi bobot spasial untuk penaksiran parameter GWLR pada lokasi ke- i dengan koordinat (u_i, v_i) sebagai berikut:

$$L(\beta(u_i, v_i)) = \left(\prod_{j=1}^n (1 + \exp(\beta^T(u_i, v_i) \mathbf{x}_j)^{-1}) \right) \exp \left(\sum_{j=1}^n y_j \beta^T(u_i, v_i) \mathbf{x}_j \right)^{w_{ij}}. \tag{12}$$

Fungsi *log-likelihood* dari fungsi *likelihood* (12) adalah

$$\ell(\boldsymbol{\beta}(u_i, v_i)) = \sum_{j=1}^n w_{i,j} (y_j \boldsymbol{\beta}^T(u_i, v_i) \mathbf{x}_j - \ln(1 + \exp(\boldsymbol{\beta}^T(u_i, v_i) \mathbf{x}_j))). \quad (13)$$

$\hat{\boldsymbol{\beta}}(u_i, v_i)$ yang memaksimalkan fungsi *log-likelihood* diperoleh dengan menurunkan fungsi (13) terhadap parameter $\boldsymbol{\beta}(u_i, v_i)$ dan disama dengankan nol sebagai berikut

$$\frac{\partial \ell(\boldsymbol{\beta}(u_i, v_i))}{\partial (\boldsymbol{\beta}(u_i, v_i))} = \mathbf{0}, \quad (14)$$

dimana $\frac{\partial \ell(\boldsymbol{\beta}(u_i, v_i))}{\partial (\boldsymbol{\beta}(u_i, v_i))}$ adalah vektor gradien berdimensi $(p+1)$ dan $\mathbf{0}$ adalah vektor nol yang berdimensi

$(p+1)$. Dengan memperhatikan ekspresi persamaan (13), persamaan (14) adalah nonlinier sehingga solusi eksak untuk mendapatkan penaksir eksak maximum likelihood (ML) model GWLR tidak dapat diperoleh secara analitik. Persamaan (14) dapat diselesaikan secara numerik, dalam penelitian ini menggunakan metode iteratif Newton-Raphson. (Inayah dkk, 2021).

Pengujian kesesuaian model bertujuan untuk mengetahui apakah terdapat perbedaan yang signifikan antara model GWLR dengan model regresi logistik. Hipotesis pengujian kesesuaian model adalah

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

$$H_1 : \text{Minimal ada satu } \beta_k(u_i, v_i) \neq \beta_k; k = 1, 2, \dots, p; i = 1, 2, \dots, n$$

Statistik uji diberikan oleh

$$F_2 = \frac{D(\hat{\boldsymbol{\beta}}) / p}{D(\hat{\boldsymbol{\beta}}^*) / np} \quad (15)$$

dengan $D(\hat{\boldsymbol{\beta}}) = 2(\ell(\hat{\boldsymbol{\Omega}}) - \ell(\hat{\omega}))$ diberikan oleh persamaan (6). $D(\hat{\boldsymbol{\beta}}^*)$ didefinisikan oleh

$$D(\hat{\boldsymbol{\beta}}^*) = 2(\ell(\hat{\boldsymbol{\Omega}}_{\text{GWLR}}) - \ell(\hat{\omega})) \quad (16)$$

dengan

$$\ell(\hat{\boldsymbol{\Omega}}_{\text{GWLR}}) = \sum_{i=1}^n (y_i \ln(\hat{\pi}_i(u_i, v_i)) + (1 - y_i) \ln(1 - \hat{\pi}_i(u_i, v_i))), \quad (17)$$

dimana $\pi_i(u_i, v_i)$ diberikan oleh persamaan (11) dan $\ell(\hat{\omega})$ diberikan oleh persamaan (8) (Pratiwi dkk, 2020).

Statistik uji pada persamaan (15) berdistribusi F dengan derajat bebas p dan np . Daerah penolakan H_0 pada taraf uji α jika $F_2 > F_{(\alpha, n, np)}$ atau $p_{\text{value}} < \alpha$ (Inayah dkk, 2021).

Pengujian parameter model GWLR terdiri dari pengujian parameter secara simultan dan parsial. Hipotesis pengujian secara simultan adalah

$$H_0 : \beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \dots = \beta_p(u_i, v_i) = 0; i = 1, 2, \dots, n$$

$$H_1 : \text{Minimal ada satu } \beta_k(u_i, v_i) \neq 0; k = 1, 2, \dots, p; i = 1, 2, \dots, n$$

Statistik uji diberikan oleh

$$G_2 = 2(\ell(\hat{\boldsymbol{\Omega}}_{\text{GWLR}}) - \ell(\hat{\omega}_{\text{GWLR}})) \quad (18)$$

dimana $\ell(\hat{\boldsymbol{\Omega}}_{\text{GWLR}})$ diberikan oleh persamaan (17). $\ell(\hat{\omega}_{\text{GWLR}})$ didefinisikan oleh

$$\ell(\hat{\omega}_{\text{GWLR}}) = \sum_{i=1}^n (y_i \ln(\hat{\pi}_i(u_i, v_i) | \hat{\beta}_0(u_i, v_i)) + (1 - y_i) \ln(1 - \hat{\pi}_i(u_i, v_i) | \hat{\beta}_0(u_i, v_i))) \quad (19)$$

Daerah penolakan H_0 pada taraf uji α jika $G_2 > \chi^2_{(\alpha, v)}$ atau $p_{\text{value}} < \alpha$ (Pratiwi dkk, 2020).

Setelah pengujian secara simultan, selanjutnya dilakukan pengujian parameter secara parsial. Hipotesis pengujian secara parsial dengan $k = 0, 1, 2, \dots, p$ adalah

$$H_0 : \beta_k(u_i, v_i) = 0; k = 0, 1, 2, \dots, p; i = 1, 2, \dots, n$$

$$H_1 : \beta_k(u_i, v_i) \neq 0$$

Statistik uji diberikan oleh

$$W_2 = \frac{\hat{\beta}_k(u_i, v_i)}{\sqrt{\text{Var}(\hat{\beta}_k(u_i, v_i))}} \quad (20)$$

Daerah penolakan H_0 ditolak pada taraf uji α jika $|W_2| > Z_{\alpha/2}$ atau jika $p_{\text{value}} < \alpha$ (Inayah dkk, 2021).

2.5 Ukuran Kebaikan Model

Ukuran kebaikan model regresi dapat dianalisis dengan menggunakan metode *Akaike Information Criterion* (AIC) AIC adalah metode untuk mendapatkan model terbaik dengan membandingkan semua

kemungkinan model yang bisa terbentuk dari peubah prediktor. Model terbaik diperoleh berdasarkan nilai AIC yang terkecil. Nilai AIC dapat diperoleh dari persamaan berikut

$$AIC = -2\ell(\hat{\beta}(u_i, v_i)) + 2k \tag{21}$$

dimana $\ell(\hat{\beta}(u_i, v_i))$ adalah nilai maksimum fungsi *log-likelihood* untuk semua lokasi dan k adalah banyaknya parameter dalam model (Sembiring, 2019).

2.6 Interpretasi Model GWLR

Interpretasi parameter regresi logistik dari peubah kategorik berskala nominal adalah dengan *odds ratio*. *Odds ratio* adalah perbandingan nilai *odds* untuk kategori $x_k = 1$ terhadap *odds* untuk kategori $x_k = 0$. Hubungan *odds ratio* dengan koefisien regresi model GWLR sebagai berikut (Aliu dkk, 2022)

$$OR(u_i, v_i) = \frac{\frac{\pi((u_i, v_i) | x_k = 1)}{1 - \pi((u_i, v_i) | x_k = 1)}}{\frac{\pi((u_i, v_i) | x_k = 0)}{1 - \pi((u_i, v_i) | x_k = 0)}} = \exp(\hat{\beta}_k(u_i, v_i)) \tag{22}$$

Berdasarkan persamaan (22) di atas diketahui bahwa setiap kenaikan satu satuan x_k maka akan menurunkan atau meningkatkan nilai π di lokasi pengamatan ke- i sebesar $e^{\hat{\beta}_k}$ (Inayah dkk, 2021).

2.7 Dissolved Oxygen (DO)

Dissolved Oxygen (DO) adalah jumlah (mg/l) oksigen terlarut di dalam air. Oksigen terlarut dalam air berasal dari hasil fotosintesis fitoplankton atau tumbuhan air lainnya dan difusi dari atmosfer. Oksigen terlarut diperlukan oleh semua organisme akuatik untuk proses respirasi dan metabolisme. Selain itu DO juga diperlukan untuk proses dekomposisi bahan organik dan oksidasi bahan anorganik dalam air (Anggraini dkk, 2014). Kandungan DO berdasarkan standar baku mutu air kelas 1 dalam Peraturan Pemerintah Republik Indonesia Nomor 22 Tahun 2021 ditunjukkan pada Tabel 1.

Tabel 1. Kriteria Mutu Air Kelas 1 Berdasarkan DO

DO	Indikasi
≥ 6 mg/l	Tidak Tercemar
< 6 mg/l	Tercemar

3. Metodologi Penelitian

3.1 Sumber Data dan Peubah Penelitian

Data penelitian merupakan data sekunder yang diperoleh dari Dinas Lingkungan Hidup (DLH) Provinsi Kalimantan Timur dalam Pradipa (2023). Peubah penelitian terdiri dari peubah respon (Y), peubah prediktor (X), dan peubah lokasi geografis (u, v). Peubah respon yaitu indikator pencemaran air DO. Peubah respon adalah kategori (nominal) dengan nilai 1 apabila kandungan DO < 6 mg/l (air sungai tercemar) dan bernilai 0 apabila kandungan DO ≥ 6 mg/l (air sungai tidak tercemar). Peubah prediktor terdiri dari konsentrasi deterjen (X_1), konsentrasi nitrat (X_2), *total dissolved solid* atau TDS (X_3), dan konsentrasi besi (X_4). Peubah lokasi geografis adalah koordinat lokasi pengamatan dengan u menyatakan letak garis lintang dan v menyatakan letak garis bujur.

3.2 Populasi, Teknik Sampling, dan Sampel Penelitian

Populasi penelitian ini adalah seluruh Daerah Aliran Sungai (DAS) Mahakam. Teknik sampling menggunakan *purposive sampling*, yaitu teknik penentuan sampel dengan pertimbangan peneliti. Pertimbangan peneliti dalam menentukan sampel adalah kebaruan data di DLH Provinsi Kalimantan Timur. Sampel yang digunakan adalah 33 titik lokasi pengamatan yang bermuara ke Sungai Mahakam tahun 2022.

4. Hasil dan Pembahasan

4.1 Deskripsi Data Penelitian

Deskripsi data penelitian dinyatakan dalam statistik deskriptif yang meliputi nilai minimum, nilai maksimum, rata-rata, dan simpangan baku. Statistik deskriptif data peubah prediktor dapat dilihat pada Tabel 2.

Tabel 2. Deskripsi Data Peubah Prediktor

Peubah	Minimum	Maksimum	Rata-Rata	Simpangan Baku
X_1	0,0020	0,1300	0,0162	0,0205

Peubah	Minimum	Maksimum	Rata-Rata	Simpangan Baku
X_2	0,0200	1,6000	0,3169	0,4278
X_3	12,0000	184,0000	59,7600	47,0112
X_4	0,0400	4,0000	0,6094	0,7585

Berdasarkan Tabel 2 diketahui bahwa rata-rata konsentrasi deterjen dari 33 titik lokasi pengamatan sebesar 0,0162 mg/l, rata-rata konsentrasi nitrat sebesar 0,3169 mg/l, rata-rata TDS sebesar 59,76 mg/l, dan rata-rata konsentrasi besi sebesar 0,6094 mg/l.

Persentase banyak lokasi pengamatan terindikasi tercemar (kategori 1) dan terindikasi tidak tercemar (kategori 0) berdasarkan nilai DO disajikan pada Tabel 3.

Tabel 3. Deskripsi Data Peubah Respon

Banyak Lokasi Pengamatan (n)	Kategori 1		Kategori 0	
	Banyak Lokasi	Persentase	Banyak Lokasi	Persentase
33	18	55%	15	45%

Berdasarkan Tabel 3, diketahui bahwa 55% dari seluruh lokasi pengamatan terindikasi tercemar dan 45% lokasi pengamatan terindikasi tidak tercemar.

4.2 Pendeteksian Multikolinieritas

Pendeteksian multikolinieritas dapat dilakukan dengan menggunakan nilai VIF. Hasil perhitungan disajikan pada Tabel 4.

Tabel 4. Nilai VIF

Peubah	VIF	Peubah	VIF
X_1	1,0073	X_3	1,6171
X_2	1,0337	X_4	1,5920

Berdasarkan Tabel 4, diketahui bahwa nilai VIF dari setiap peubah prediktor lebih kecil dari 10. Hal tersebut menunjukkan bahwa tidak terdapat multikolinieritas antar peubah prediktor, sehingga pemodelan regresi logistik biner pada penelitian ini menggunakan keempat peubah prediktor yaitu konsentrasi deterjen, konsentrasi nitrat, TDS, dan konsentrasi besi.

4.3 Pemodelan Regresi Logistik Biner pada Data DO Air Sungai Mahakam

Pemodelan dengan regresi logistik biner meliputi penaksiran parameter, pengujian parameter secara simultan, dan pengujian parameter secara parsial. Penaksiran parameter model regresi biner menggunakan metode MLE yang diselesaikan dengan metode iterative Newton-Raphson. Hasil perhitungan penaksiran parameter dan pengujian parameter menggunakan *software Octave* disajikan pada Tabel 5.

Tabel 5. Hasil Penaksiran dan Pengujian Parameter Model Regresi Logistik Biner

Parameter	Nilai Taksiran	Standar Error	$ W_{hitung} $	P_{value}
β_0	-1,5096	1,1778	1,2817	0,1999
β_1	-20,1474	36,1813	0,5568	0,5776
β_2	7,6216	4,5339	1,6810	0,0928*
β_3	-0,0040	0,0118	0,3399	0,7339
β_4	1,4409	1,1958	1,1958	0,2282

Berdasarkan Tabel 5, diperoleh model regresi logistik biner sebagai berikut

$$\hat{\pi}_i = \frac{\exp(-1,5096 - 20,1474x_{i,1} + 7,6216x_{i,2} - 0,0040x_{i,3} + 1,4409x_{i,4})}{1 + \exp(-1,5096 - 20,1474x_{i,1} + 7,6216x_{i,2} - 0,0040x_{i,3} + 1,4409x_{i,4})} \tag{23}$$

Pengujian parameter model regresi logistik biner dilakukan secara simultan dan parsial. Hasil pengujian parameter secara simultan diperoleh bahwa secara simultan peubah konsentrasi deterjen, konsentrasi nitrat, TDS, dan konsentrasi besi berpengaruh terhadap peluang air Sungai Mahakam terindikasi tercemar. Hal ini didukung oleh hasil uji simultan dengan nilai statistik uji G pada persamaan (6) sebesar 12,5253 lebih besar dari nilai kritis $\chi^2_{(0,1;4)} = 7,7794$, dan $p_{value} = 0,0138$ lebih kecil dari taraf uji $\alpha = 0,1$.

Berdasarkan hasil pengujian parameter secara parsial pada Tabel 5, diperoleh kesimpulan bahwa

peubah konsentrasi nitrat (X_2) secara individual berpengaruh terhadap peluang air Sungai Mahakam terindikasi tercemar. Hal ini didukung oleh hasil uji parsial dengan nilai statistik uji $|W_{hitung}|$ pada persamaan (9) sebesar 1,6810 lebih besar dari nilai kritis $Z_{(0,05)}$ berdasarkan tabel distribusi normal baku sebesar 1,64, dan $p_{value} = 0,0928$ lebih kecil dari taraf uji $\alpha = 0,1$. Peubah konsentrasi deterjen (X_1), TDS (X_3), dan konsentrasi besi (X_4) secara individual tidak berpengaruh terhadap peluang air Sungai Mahakam terindikasi tercemar. Hal ini karena nilai statistik uji $|W_{hitung}|$ peubah X_1, X_3 , dan X_4 kurang dari nilai kritis $Z_{(0,05)} = 1,64$, dan p_{value} lebih besar dari taraf uji $\alpha = 0,1$.

4.4 Pengujian Heterogenitas Spasial

Pengujian heterogenitas spasial data respon menggunakan uji *Glejser*. Berdasarkan hasil pengujian heterogenitas spasial, diperoleh bahwa data pengamatan merupakan data spasial. Hal ini didukung oleh hasil uji *Glejser* dengan nilai statistik uji F pada persamaan (10) sebesar 9,677 lebih besar dari nilai kritis $F_{(0,1;4;28)}$ berdasarkan tabel distribusi F sebesar 2,1571, dan p_{value} sebesar $4,8 \times 10^{-5}$ lebih kecil dari taraf uji $\alpha = 0,1$. Pemodelan dengan regresi logistik biner diduga tidak sesuai untuk memodelkan DO di DAS Mahakam sehingga model yang sesuai adalah model GWLR.

4.5 Pemodelan GWLR pada Data DO Air Sungai Mahakam

Pemodelan dengan GWLR pada data DO meliputi perhitungan jarak *euclidean*, penentuan *bandwidth* optimum, perhitungan pembobot spasial, penaksiran parameter, pengujian kesesuaian model, pengujian parameter secara simultan, dan pengujian parameter secara parsial. Setelah memperoleh jarak *euclidean*, langkah berikutnya adalah menentukan nilai *bandwidth* optimum di seluruh lokasi pengamatan berdasarkan nilai GCV minimum. Pembobot spasial dapat dihitung berdasarkan nilai *bandwidth* optimum yang telah diperoleh.

Setelah memperoleh nilai pembobot spasial untuk seluruh lokasi, langkah berikutnya dilakukan penaksiran parameter model GWLR. Hasil perhitungan penaksiran parameter dan pengujian parameter menggunakan *software Octave* disajikan pada Tabel 6.

Tabel 6. Hasil Penaksiran Parameter Model GWLR

Lokasi	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	Lokasi	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
1	-3,0636	-13,030	8,1279	-0,0164	4,0755*	18	-1,4544	-20,316	7,9938*	-0,0027	1,2131
2	-3,0787	-13,089	8,2152	-0,0161	4,0440*	19	-1,4541	-20,316	7,9970*	-0,0027	1,2116
3	-1,4542	-20,316	7,9965*	-0,0027	1,2118	20	-1,6126	-19,612	7,7508*	-0,0048	1,5812
4	-1,4544	-20,316	7,9961*	-0,0027	1,2123	21	-1,6133	-19,610	7,7484*	-0,0048	1,5835
5	-1,4546	-20,314	8,0094*	-0,0026	1,2079	22	-3,1482	-12,467	6,6635	-0,0229	5,2547*
6	-1,4553	-20,313	7,9966*	-0,0027	1,2135	23	-3,1500	-12,470	6,6601	-0,0229	5,2599*
7	-1,5096	-20,147	7,6216*	-0,0040	1,4409	24	-1,5500	-19,926	7,6825*	-0,0043	1,4928
8	-1,5762	-19,715	7,9455*	-0,0041	1,4456	25	-1,5877	-19,722	7,7444*	-0,0046	1,5374
9	-1,5096	-20,147	7,6216*	-0,0040	1,4409	26	-1,5614	-19,768	8,0075*	-0,0038	1,3980
10	-1,5974	-19,691	7,7088*	-0,0047	1,5681	27	-1,5637	-19,756	8,0132*	-0,0038	1,4000
11	-1,5096	-20,147	7,6216*	-0,0040	1,4409	28	-1,5404	-20,010	7,6688*	-0,0041	1,4790
12	-1,5395	-20,014	7,6679*	-0,0041	1,4777	29	-1,6097	-19,640	7,7098*	-0,0048	1,5908
13	-1,5531	-19,951	7,6578*	-0,0043	1,5063	30	-1,6098	-19,643	7,7073*	-0,0048	1,5922
14	-1,5880	-19,737	7,6797*	-0,0047	1,5548	31	-1,5740	-19,841	7,5974*	-0,0047	1,5663
15	-1,5891	-19,739	7,6793*	-0,0047	1,5636	32	-3,1410	-13,131	8,3120	-0,0159	4,0662*
16	-1,6033	-19,658	7,7281*	-0,0047	1,5718	33	-1,5439	-19,995	7,6611*	-0,0042	1,4881
17	-1,5796	-19,778	7,6891*	-0,0046	1,5430						

Hasil penaksiran pada Tabel 6 yang bertanda (*) menunjukkan berpengaruh signifikan pada taraf uji 10% berdasarkan uji parsial.

Berdasarkan hasil pengujian kesamaan model, disimpulkan bahwa model GWLR dan model regresi logistik tidak identik. Hal ini didukung oleh hasil uji kesamaan dengan nilai statistik uji F_2 pada persamaan (15) sebesar 23,7510 lebih besar dari nilai kritis $F_{(0,1;4;132)}$ berdasarkan tabel distribusi F sebesar 1,9879, dan $p_{value} = 8,2 \times 10^{-15}$ lebih kecil dari taraf uji $\alpha = 0,1$.

Hasil pengujian parameter secara simultan diperoleh kesimpulan bahwa model GWLR layak. Hal ini didukung oleh hasil uji simultan dengan nilai statistik uji G_2 pada persamaan (18) sebesar 15,7427 lebih besar dari nilai kritis $\chi^2_{(0,1;6)} = 10,6446$, dan $p_{value} = 0,0152$ lebih kecil dari taraf uji $\alpha = 0,1$.

Hasil pengujian parameter secara parsial diperoleh kesimpulan bahwa peubah konsentrasi besi (X_4) secara individual berpengaruh terhadap peluang air Sungai Mahakam terindikasi tercemar di lokasi

pengamatan Nursery Suring PPU. Hal ini didukung oleh hasil uji parsial dengan nilai statistik uji $|Z_{hitung}|$ peubah X_4 sebesar 1,6613 lebih besar dari nilai kritis $Z_{(0,05)}$ berdasarkan tabel distribusi normal baku sebesar 1,64, dan $p_{value} = 0,0966$ lebih kecil dari taraf uji $\alpha = 0,1$. Peubah konsentrasi deterjen (X_1), konsentrasi nitrat (X_2), dan TDS (X_3) secara individual tidak berpengaruh terhadap peluang air Sungai Mahakam terindikasi tercemar di lokasi pengamatan Nursery Suring PPU. Hal ini karena nilai statistik uji $|Z_{hitung}|$ peubah X_1, X_2 , dan X_3 kurang dari nilai kritis $Z_{(0,05)} = 1,64$, dan p_{value} lebih besar dari taraf uji $\alpha = 0,1$.

Berdasarkan uji kesesuaian model, uji simultan, dan uji parsial, maka sebagai contoh model GWLR di lokasi pengamatan ke-1 (Nursery Suring PPU) adalah

$$\hat{\pi}(u_1, v_1) = \frac{\exp(-3,0636 - 13,030x_{1,1} + 8,1279x_{1,2} - 0,0164x_{1,3} + 4,0755x_{1,4})}{1 + \exp(-3,0636 - 13,030x_{1,1} + 8,1279x_{1,2} - 0,0164x_{1,3} + 4,0755x_{1,4})} \tag{24}$$

dimana faktor yang berpengaruh terhadap peluang air sungai Mahakan terindikasi tercemar di lokasi Nursery Suring PPU adalah konsentrasi besi.

Berdasarkan peubah prediktor yang berpengaruh pada Tabel 6, maka hasil pengelompokan model GWLR seluruh lokasi pengamatan disajikan pada Tabel 7.

Tabel 7. Pengelompokan Model GWLR Berdasarkan Peubah Prediktor yang Berpengaruh

Kelompok	Model Terbaik	Peubah yang Berpengaruh	Lokasi Pengamatan
1	Lokal	X_4	Nursery Suring PPU, Jembatan Sei Sepaku, Titik Hulu Sungai Kasungai Paser, Titik Hilir Sungai Kasungai, dan Titik Hilir (Pos Pon 1) Balikpapan
2	Global	X_2	Jembatan Santan Tengah Kukar, Tering Kubar, dan Palaran Samarinda
3	Lokal	X_2	Keraton Gunung Tabur Berau, Jembatan Gunung Tabur Berau, Hulu PT. BBE Berau, Hilir Berau Coal, Jembatan Marangkayu Santan Ulu Kukar, Pulau Kumala Kukar, Nyan Mahulu, Long Bagun Mahulu, Kantor Gubernur Samarinda, Kalamur Samarinda, Bloro Kukar, Anggana Kukar, Keraton Sambaliung Berau, Kantor Bupati Berau, Desa Siran Kukar, Desa Kedang Kepala Kukar, Jembatan Desa Pampang Samarinda, Pampang Dalam Samarinda, Jembatan Jl. Soekarno-Hatta Bontang, Jembatan Jl. Poros Bontang-Samarinda, Muara Sungai Boh Mahulu, Muara Sungai Belayan Kukar, Desa Sebelimbingan Kukar, Titik Hulu (Sungai 28), dan Batoq Kelo Mahulu

4.6 Pemilihan Model Terbaik

Pemilihan model terbaik antara model regresi logistik biner dengan model GWLR menggunakan nilai GCV dan AIC. Hasil perhitungan nilai GCV dan AIC disajikan pada Tabel 8

Tabel 8. Nilai GCV dan AIC Model regresi logistik biner dan Model GWLR

Model Regresi	GCV	AIC
Regresi Logistik Biner	0,2411	40,9480
GWLR	0,2052	36,0702

Berdasarkan Tabel 8, diperoleh nilai GCV dan AIC model GWLR lebih kecil dari model regresi logistik biner sehingga dapat disimpulkan bahwa model GWLR lebih baik dari model regresi logistik biner dalam memodelkan peluang pencemaran air Sungai Mahakam melalui indikator DO.

4.7 Interpretasi Model GWLR

Interpretasi model GWLR didasarkan pada nilai *odds ratio* dari peubah prediktor yang berpengaruh di setiap lokasi pengamatan. Model GWLR pada lokasi pengamatan Nursery Suring PPU diberikan oleh persamaan (24). Berdasarkan uji parsial model GWLR di lokasi pengamatan Nursery Suring PPU, diperoleh peubah prediktor yang berpengaruh berdasarkan Tabel 6 adalah konsentrasi besi. Nilai *odds ratio* berdasarkan peubah konsentrasi besi adalah 58,8809 yang berarti setiap peningkatan 1 mg/l konsentrasi besi dan nilai peubah lain dianggap tetap maka peluang air sungai di lokasi Nursery Suring PPU terindikasi tercemar

meningkat menjadi 58,8809 kali.

5. Kesimpulan

Berdasarkan hasil penelitian diperoleh kesimpulan sebagai berikut.

1. Pemodelan peluang air Sungai Mahakam terindikasi tercemar terbaik adalah model *Geographically Weighted Logistic Regression* (GWLR).
2. Faktor-faktor yang berpengaruh terhadap peluang air Sungai Mahakam terindikasi tercemar bersifat lokal, yaitu konsentrasi nitrat dan konsentrasi besi.

6. Daftar Pustaka

- Aliu, M.A., Zubedi, F., Yahya, L., & Oroh, F.A. (2022). Perbandingan Fungsi Pembobot Kernel pada Geographically Weighted Logistic Regression dalam Memodelkan Kasus Kemiskinan di Indonesia. *Jurnal Matematika, Statistika & Komputasi*. 18(3), 362-384. <https://journal.unhas.ac.id/index.php/jmsk/article/view/19567/8139>.
- Alwi, W., Ermawati, & Husain, S. (2018). Analisis Regresi Logistik Biner untuk Memprediksi Kepuasan Pengunjung pada Rumah Sakit Umum Daerah Majene. *Jurnal MSA*. 6(1), 20-26. <https://journal.uin-alauddin.ac.id/index.php/msa/article/view/4783/pdf>.
- Anggraini, N., Simarmata, A.H., & Sihotang, C. (2014). Dissolved Oxygen Concentration from the Water Around the Floating Cage Fish Culture Area and From the Area with No Cage. In the DAM Site of the Koto Panjang Reservoir. *Jurnal Online Mahasiswa (JOM) Bidang Perikanan dan Ilmu Kelautan*. 2(1), 1-7. <https://jom.unri.ac.id/index.php/JOMFAPERIKA/article/view/4704/4586>.
- Chairina, P., Suyitno, & Siringoringo, M. (2020). Model-Model Regresi Weibull Univariat pada Indikator Pencemaran Air Dissolved Oxygen di Daerah Aliran Sungai Lingkungan Hutan Hujan Tropis Kalimantan Timur. *Jurnal Ekspensial*. 11(1), 19-28. <https://jurnal.fmipa.unmul.ac.id/index.php/exponensial/article/view/641/278>.
- Hidayanti, U., Latra, I.N., & Purhadi. (2015). Pemodelan dan Pemetaan Jumlah Kasus DBD di Kota Surabaya dengan Geographically Weighted Negative Binomial Regression (GWNBR) dan Flexibly Shaped Spatial Scan Statistic. *Jurnal Sains dan Seni ITS*. 4(2), 272-278. https://ejournal.its.ac.id/index.php/sains_seni/article/view/11183/2426.
- Inayah, U.R., Suyitno, & Siringoringo, M. (2021). Upaya Pencegahan Pencemaran Air Sungai Mahakam Melalui Pemodelan Geographically Weighted Logistic Regression pada Data BOD. *Jurnal Ekspensial*. 12(1), 17-26. <https://jurnal.fmipa.unmul.ac.id/index.php/exponensial/article/view/755/311>.
- Lestari, V.D., Suyitno, & Siringoringo, M. (2021). Analisis Faktor-Faktor yang Berpengaruh terhadap Pencemaran Air Sungai Mahakam menggunakan Pemodelan Geographically Weighted Logistic Regression pada Data Dissolved Oxygen. *Jurnal Ekspensial*. 12(1), 37-46. <https://jurnal.fmipa.unmul.ac.id/index.php/exponensial/article/view/757/313>.
- Pemerintah Republik Indonesia. (2021). *Peraturan Pemerintah Republik Indonesia Nomor 22 Tahun 2021 Tentang Penyelenggaraan Perlindungan dan Pengelolaan Lingkungan Hidup*. Jakarta: Pemerintah Republik Indonesia.
- Pradipa, Z. (2023). *Analisis Potensi Pencemaran Air Sungai Mahakam menggunakan Pemodelan Regresi Weibull (Studi Kasus: Data Indikator Pencemaran Air DO Tahun 2022)* (Skripsi). Universitas Mulawarman.
- Pratiwi, N., Suyitno, & Siringoringo, M. (2020). Penerapan Model Geographically Weighted Logistic Regression pada Data Status Kesejahteraan Masyarakat di Kalimantan Timur Tahun 2017. *Jurnal Ekspensial*. 11(1), 83-92. <https://jurnal.fmipa.unmul.ac.id/index.php/exponensial/article/view/648/285>.
- Rahman, R.A., Amalia, A.R., Riandhis, J.A., Hidayah, H., & Mardiah. (2017). Peningkatan Kualitas Air Baku Sungai Mahakam dengan Teknologi MOCI (Moringa Oleifera and Cellulose Installation). *Prosiding Seminar Nasional Teknologi IV*. e-ISSN: 2598-7429, 7-14. https://www.researchgate.net/publication/331645449_PENINGKATAN_KUALITAS_AIR_BAKU_SUNGAI_MAHAKAM_DENGAN_TEKNOLOGI_MOCI_MORINGA_OLEIFERA_AND_CELLULOSE_INSTALLATION.
- Safitri, R.N., Suyitno, & Hayati, M.N. (2020). Penerapan Model Mixed Geographically Weighted Regression dengan Fungsi Pembobot Adaptive Tricube pada IPM 30 Kabupaten/Kota di Provinsi Kalimantan Timur. Kalimantan Tengah. dan Kalimantan Selatan Tahun 2016. *Jurnal Ekspensial*. 11(2), 107-116. <https://jurnal.fmipa.unmul.ac.id/index.php/exponensial/article/view/651/288>.
- Sembiring, A.P. (2019). Faktor-Faktor yang Memengaruhi Besar Klaim Asuransi Jiwa dengan Menggunakan Model Regresi Tobit. *Bimaster*. 8(4), 729-736. <https://jurnal.untan.ac.id/index.php/jbmstr/article/view/35996/75676583139>.

- Sriningsih, M., Hatidja, D., & Prang, J.D. (2018). Penanganan Multikolinearitas Dengan Menggunakan Analisis Regresi Komponen Utama Pada Kasus Impor Beras di Provinsi Sulut. *Jurnal Ilmiah Sains*. 18(1), 18-24. <https://ejournal.unsrat.ac.id/v3/index.php/JIS/article/view/19396>.
- Tizona, A.R., Goejantoro, R., & Wasono. (2017). Pemodelan Geographically Weighted Regression (GWR) dengan Fungsi Pembobot Adaptive Bisquare untuk Angka Kesakitan Demam Berdarah di Kalimantan Timur Tahun 2015. *Jurnal Eksponensial*. 8(1), 87-94. <https://jurnal.fmipa.unmul.ac.id/index.php/exponensial/article/view/81/50>.