

## **Modelling the Probability of River Water Pollution Using Geographically Weighted Logistic Regression Model (Case Study: River Water DO Data in East Kalimantan)**

### **Pemodelan Peluang Pencemaran Air Sungai Menggunakan Model *Geographically Weighted Logistic Regression* (Studi Kasus: Data DO Air Sungai di Kalimantan Timur)**

**Adelia Miranda<sup>1</sup>, Suyitno<sup>\*2</sup>, Meirinda Fauziyah<sup>3</sup>**

<sup>1,2,3</sup>*Program Studi S1 Statistika, Jurusan Matematika, FMIPA, Universitas  
Mulawarman*

*Email: <sup>1</sup>adeliamiranda029@gmail.com, <sup>2</sup>suyitno.stat.unmul@gmail.com, <sup>3</sup>meirindaf@gmail.com*

*\*Corresponding Author*

*Received: 17 September 2024, revised: 20 November 2024, accepted: 29 November 2024*

#### **Abstract**

Geographically Weighted Logistic Regression (GWLR) is a local binary logistic regression model, and it's applied to the spatial heterogeneity data. The parameter estimation of GWLR model in this study uses Maximum Likelihood Estimation (MLE) method, and it's conducted at each observation location with spatial weighting. The spatial weight in this study was calculated using the adaptive tricube function. The spatial weighting function depends on distance between observation location and bandwidth, where the determination of optimal bandwidth uses the Akaike Information Criterion (AIC). The aim of this research is to identify the factors influencing the probability of river water pollution in East Kalimantan Province through GWLR modelling to Dissolved Oxygen (DO) data 2022, and to interpret it based on the best model. The research data is secondary data provided by Life Environment Department of East Kalimantan Province. Research concludes that the GWLR was fit model based on the results of similarity testing of the GWLR model and global model, as well as simultaneous parameter testing, with the model fitting measure was a McFadden R-Squared value of 61,1%, and an AIC value of 29,629. Based on partial parameter testing, local factors influencing chance of river water pollution in East Kalimantan can be identified, namely nitrate concentration and water color degree. Based on the GWLR modelling to DO data 2022, it can be interpreted that increasing nitrate concentration and water colour degree respectively will increase the probability of river water pollution.

**Keywords:** AIC, DO, GWLR, MLE, *river water pollution*.



**Abstrak**

Model *Geographically Weighted Logistic Regression* (GWLR) adalah model regresi logistik biner lokal, dan diterapkan pada data heterogenitas spasial (spasial tipe titik). Penaksiran parameter model GWLR dalam penelitian ini menggunakan metode *Maximum Likelihood Estimation* (MLE), dan dilakukan di setiap titik lokasi pengamatan dengan pembobot spasial. Pembobot spasial pada penelitian ini dihitung menggunakan fungsi pembobot *adaptive tricube*. Fungsi Pembobot spasial tergantung pada jarak antar lokasi pengamatan dan *bandwidth*, dimana penentuan *bandwidth* optimum menggunakan kriteria *Akaike Information Criterion* (AIC). Tujuan penelitian ini adalah mengidentifikasi faktor-faktor yang berpengaruh terhadap peluang pencemaran air sungai di Provinsi Kalimantan Timur melalui pemodelan GWLR pada data *Dissolved Oxygen* (DO) Tahun 2022, dan menginterpretasikannya berdasarkan model GWLR terbaik. Data penelitian ini adalah data sekunder yang disediakan oleh Dinas Lingkungan Hidup (DLH) Provinsi Kalimantan Timur. Penelitian menyimpulkan bahwa model GWLR adalah model yang *fit* (layak), yang didasarkan pada hasil pengujian similaritas model GWLR dan model global, serta pengujian parameter secara simultan, dengan ukuran *fitting* (kelayakan) model adalah nilai *McFadden R-Squared* sebesar 61,1%, dan nilai AIC sebesar 29,629. Berdasarkan pengujian parameter secara parsial dapat diidentifikasi faktor-faktor lokal yang memengaruhi peluang air sungai di Kalimantan Timur tercemar, yaitu konsentrasi nitrat dan derajat warna air. Berdasarkan pemodelan GWLR pada data DO 2022 dapat diinterpretasikan bahwa setiap kenaikan konsentrasi nitrat dan derajat warna air masing-masing akan meningkatkan peluang pencemaran air sungai.

**Kata kunci:** AIC, DO, GWLR, MLE, peluang pencemaran air sungai.

**1. Pendahuluan**

Air bersih merupakan kebutuhan dasar kehidupan manusia. Air bersih diperlukan manusia untuk memenuhi konsumsi air minum dan memenuhi kebutuhan sehari-hari seperti mandi, mencuci, dan keperluan lainnya. Perusahaan air bersih memanfaatkan berbagai sumber air untuk diolah menjadi air bersih layak konsumsi, dimana sungai masih mendominasi sebagai sumber utama air baku di Kalimantan Timur. Hal ini didukung oleh topografi Provinsi Kalimantan Timur yang memiliki aliran sungai yang membentang hampir di seluruh wilayah. Sekitar 74,35 persen dari total volume air baku yang digunakan oleh perusahaan air bersih di Kalimantan Timur pada tahun 2022 berasal dari sungai, diikuti oleh air waduk sebesar 15,30 persen dan air tanah sebesar 9,46 persen [3].

Sungai di Provinsi Kalimantan Timur selain berfungsi sebagai sumber air baku utama juga menjadi pusat berbagai aktivitas masyarakat, seperti kegiatan transportasi, bongkar muat hasil tambang, dan disepanjang Daerah Aliran Sungai (DAS) terdapat aktivitas rumah tangga, restoran, industri, dan pertambangan. Kegiatan-kegiatan tersebut berpotensi menghasilkan limbah dalam aliran sungai yang dapat mengakibatkan air sungai terancam tercemar. Air sungai yang tercemar akan mengancam kesehatan masyarakat Kalimantan Timur, oleh karenanya perlu tindakan pencegahan. Salah satu upaya pencegahan pencemaran air secara statistika adalah memberikan informasi faktor-faktor yang memengaruhi pencemaran aliran air sungai di Kalimantan Timur kepada masyarakat dan pemerintah daerah melalui pemodelan *Geographically Weighted Logistic Regression* (GWLR) pada data *Dissolved Oxygen* (DO). Melalui pemodelan GWLR dapat dianalisis faktor-faktor yang berpengaruh signifikan terhadap peluang pencemaran air sungai.

Faktor-faktor yang berpengaruh terhadap peluang pencemaran air sungai perlu diinformasikan kepada masyarakat agar mereka dapat berpartisipasi dalam pencegahan pencemaran air. Informasi tersebut sangat berguna bagi pemerintah daerah sebagai pertimbangan dalam membuat kebijakan pencegahan pencemaran air sungai.

DO adalah kadar oksigen terlarut di air, dan merupakan salah satu indikator untuk mendeteksi pencemaran air [19]. Data DO air sungai dipengaruhi oleh karakteristik geografis lingkungan DAS dan data DO adalah data spasial, oleh karena itu dapat dimodelkan menggunakan GWLR. Metode penaksiran parameter pada penelitian adalah MLE dan penaksir *Maximum Likelihood* (ML) ditemukan menggunakan metode numerik, yakni metode iteratif Newton-Raphson. Pembobot spasial dihitung menggunakan fungsi pembobot *adaptive tricube* dengan penentuan *bandwidth* optimum menggunakan kriteria AIC. Komputasi pada penelitian ini menggunakan *opensource software R* dan *octave*. Pemodelan GWLR pada data DO air Sungai Mahakam 2018 telah diteliti oleh Lestari dkk [12], dimana penentuan *bandwidth* optimum menggunakan kriteria *generalized cross validation* (GCV), hal ini yang menjadi salah satu pembeda dengan penelitian ini.

Berdasarkan penelitian sebelumnya, model GWLR lebih unggul daripada model global (model regresi logistik biner). Hal ini ditunjukkan hasil penelitian yang dilakukan oleh Wardhani dkk [23] dan Salsavira dkk [18], yang menyimpulkan bahwa model GWLR lebih baik daripada model global, berdasarkan nilai AIC model GWLR lebih kecil daripada model global, begitu juga GCV model GWLR lebih kecil dari model global [12].

Penelitian tentang model regresi logistik dan model GWLR yang lainnya dengan pembahasan yang berbeda dengan penelitian ini, yaitu perbandingan pemodelan GWLR dan regresi logistik biner (model global) pada data COVID-19 di Sumatera Barat oleh Haq, I., dkk [7], dimana penelitian ini menyimpulkan bahwa, model GWLR identik dengan model globalnya. Rifada, M., dkk [17] membahas faktor penentu *stunting* pada balita di Kabupaten Probolinggo menggunakan model logistik ordinal parametrik dan non parametrik. Perdana, H., dkk [15] membahas model GWLR untuk mengetahui faktor yang memengaruhi kelulusan Mahasiswa FMIPA Universitas Tanjungpura berdasarkan lokasi sekolah, dengan ukuran kebaikan model adalah ketepatan klasifikasi.

## 2. TINJAUAN PUSTAKA

### 2.1 Model Regresi Logistik Global

Model regresi logistik biner (global) adalah suatu metode statistika yang digunakan untuk memodelkan hubungan antara variabel respon nominal biner dengan satu atau lebih variabel prediktor kontinu dan atau kategorik [19]. Model umum regresi logistik adalah

$$\pi_i = \frac{\exp(\boldsymbol{\beta}^T \mathbf{x}_i)}{1 + \exp(\boldsymbol{\beta}^T \mathbf{x}_i)} ; i = 1, 2, \dots, n, \quad (2.1)$$

dengan  $\boldsymbol{\beta}^T = [\beta_0 \ \beta_1 \ \beta_2 \ \dots \ \beta_p]$ ,  $\mathbf{x}_i = [x_{i0} \ x_{i1} \ x_{i2} \ \dots \ x_{ip}]^T$ , dan  $\pi_i$  menyatakan  $P(y_i = 1)$ , yakni peluang suatu percobaan sukses [1], dan dalam penelitian ini diinterpretasikan peluang pencemaran air sungai. Metode penaksiran parameter adalah MLE, dimana penaksir ML ditemukan menggunakan metode numerik, yaitu metode iteratif Newton-Raphson [10].

Pengujian signifikansi parameter regresi logistik global terdiri dari pengujian secara simultan dan parsial. Pengujian signifikansi parameter secara simultan (serentak) bertujuan untuk mengetahui kelayakan model regresi. Hipotesis pengujian parameter secara simultan adalah

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0 \text{ (Model regresi logistik global tidak layak)}$$

$H_1$  : Minimal ada satu  $\beta_k \neq 0, k = 1, 2, \dots, p$  (Model regresi logistik global layak).

Statistik uji pada pengujian ini adalah

$$G = 2 \left( \ell(\hat{\Omega}) - \ell(\hat{\omega}) \right), \quad (2.2)$$

dimana

$$\ell(\hat{\Omega}) = \sum_{i=1}^n y_i \hat{\beta}^T \mathbf{x}_i - \sum_{i=1}^n \ln[1 + \exp(\hat{\beta}^T \mathbf{x}_i)], \quad (2.3)$$

dan

$$\ell(\hat{\omega}) = \ln(\hat{\pi}) \sum_{i=1}^n y_i + \left( n - \sum_{i=1}^n y_i \right) \ln(1 - \hat{\pi}), \quad (2.4)$$

dengan  $G \sim \chi_p^2$ .  $H_0$  ditolak pada taraf uji  $\alpha$  jika  $G > \chi_{(1-\alpha, p)}^2$  atau  $H_0$  ditolak jika  $P_{value} < \alpha$  [10].

Uji signifikansi parameter yang kedua adalah pengujian parameter model secara parsial, yang bertujuan untuk mengetahui pengaruh masing-masing variabel prediktor terhadap variabel respon. Hipotesis pengujian secara parsial untuk parameter  $\beta_k$  untuk  $k$  tertentu ( $k = 1, 2, \dots, p$ ) adalah

$H_0$  :  $\beta_k = 0$  (Variabel prediktor ke- $k$  tidak berpengaruh)

$H_1$  :  $\beta_k \neq 0$  (Variabel prediktor ke- $k$  berpengaruh).

Statistik uji adalah

$$W_k = \frac{\hat{\beta}_k}{\sqrt{\text{Var}(\hat{\beta}_k)}}, \quad k = 1, 2, \dots, p, \quad (2.5)$$

dimana  $\text{Var}(\hat{\beta}_k)$  adalah elemen diagonal utama ke- $k$  dari invers matriks Informasi Fisher. Matriks Informasi Fisher diberikan oleh

$$\mathbf{I}(\hat{\beta}) = \mathbf{X}^T \hat{\mathbf{V}} \mathbf{X}. \quad (2.6)$$

dengan  $\mathbf{X}$  adalah matriks data variabel prediktor berukuran  $n \times (p + 1)$ ,  $\hat{\mathbf{V}} = \text{diag}(v_1, v_2, \dots, v_i, \dots, v_n)$ ;  $v_i = \pi_i(1 - \pi_i)$ , dan  $\pi_i$  diberikan oleh persamaan (2.1). Diketahui bahwa statistik uji  $W_k \sim N(0, 1)$ .  $H_0$  pada pengujian parameter secara parsial ditolak pada taraf uji  $\alpha$  jika nilai  $|W| > Z_{1-\alpha/2}$  atau jika  $P_{value} < \alpha$  [12], [21].

## 2.2 Pendeteksian Multikolinieritas

Pendeteksian multikolinieritas bertujuan untuk mengetahui apakah terdapat korelasi antar variabel prediktor dalam model regresi. Multikolinieritas dapat dideteksi dengan melihat nilai *Variance Inflation Factor* (VIF), dimana nilai  $VIF > 10$  memberikan informasi terdapat multikolinieritas pada model regresi [20]. Nilai VIF dihitung dengan rumus

$$VIF_k = \frac{1}{1 - R_k^2}; \quad k = 1, 2, \dots, p, \quad (2.7)$$

dengan  $R_k^2$  adalah koefisien determinasi model regresi dari variabel prediktor  $X_k$  yang diregresikan terhadap variabel prediktor lainnya [6].

## 2.3 Pengujian Heterogenitas Spasial

Pengujian heterogenitas spasial bertujuan untuk mengetahui apakah data variabel respon merupakan data heterogenitas spasial. Salah satu metode pengujian data heterogenitas spasial adalah uji *Glejser*. Uji *Glejser* mengadopsi uji parameter secara simultan pada model regresi linier berganda [6]. Pengujian heterogenitas spasial pada penelitian ini diterapkan pada data variabel

respon sebelum dikategorikan menjadi data nominal biner. Hipotesis pengujian heterogenitas spasial adalah

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

(Data pengamatan variabel respon bukan data heterogenitas spasial)

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

(Data pengamatan variabel respon adalah data heterogenitas spasial)

Statistik uji adalah statistik  $F_2$ , yaitu

$$F_2 = \frac{(\hat{\alpha}\mathbf{X}^T\mathbf{e} - n\bar{e}^2)/p}{(\mathbf{e}^T\mathbf{e} - \hat{\alpha}\mathbf{X}^T\mathbf{e}/n - p - 1)}, \quad (2.8)$$

dengan  $\hat{\alpha}$  merupakan vektor penaksir *Ordinary Least Square* (OLS) model regresi *auxiliary*, dan  $\mathbf{e}$  adalah vektor nilai mutlak dari residual, yakni  $[|\hat{e}_1| \ |\hat{e}_2| \ \dots \ |\hat{e}_n|]^T$ ;  $\hat{e}_i = y_i - \hat{y}_i$ . Diketahui bahwa statistik uji  $F_2 \sim F_{(p, n-p-1)}$ .  $H_0$  ditolak pada taraf uji  $\alpha$  jika nilai  $F_2 > F_{(1-\alpha; (p, n-p-1))}$  atau jika  $p_{value} < \alpha$  [10], [12].

## 2.4 Model Geographically Weighted Logistic Regression

Model *Geographically Weighted Logistic Regression* (GWLR) adalah model lokal dari model regresi logistik global, dimana setiap parameter regresi bergantung pada lokasi geografis dan ditaksir pada setiap lokasi pengamatan [5]. Misal diketahui koordinat setiap lokasi pengamatan, maka model GWLR [4] pada lokasi pengamatan ke- $i$  adalah

$$\pi_i(u_i, v_i) = \frac{\exp(\boldsymbol{\beta}^T(u_i, v_i)\mathbf{x}_i)}{1 + \exp(\boldsymbol{\beta}^T(u_i, v_i)\mathbf{x}_i)}; i = 1, 2, \dots, n. \quad (2.9)$$

Metode penaksiran parameter model GWLR adalah MLE dengan pembobot spasial, dimana penaksir ML ditemukan menggunakan metode iteratif Newton-Raphson [12]. Penaksiran parameter model GWLR menggunakan pembobot spasial. Pembobot spasial dihitung menggunakan fungsi pembobot yang bergantung pada jarak antar lokasi pengamatan dan *bandwidth*. Fungsi pembobot pada penelitian ini adalah fungsi pembobot *adaptive tricube* [12]. Kriteria penentuan *bandwidth* optimum dalam penelitian ini adalah *Akaike Information Criterion* (AIC). *Bandwidth* optimum adalah nilai *bandwidth* yang memberikan nilai AIC model GWLR minimum [5]. Nilai AIC dihitung menggunakan rumus

$$AIC(u_i, v_i) = -2\ell(\hat{\boldsymbol{\beta}}(u_i, v_i)) + 2K, i = 1, 2, \dots, n \quad (2.10)$$

[22], dengan

$$\ell(\hat{\boldsymbol{\beta}}(u_i, v_i)) = \sum_{i=1}^n y_i \hat{\boldsymbol{\beta}}^T(u_i, v_i) \mathbf{x}_i - \sum_{i=1}^n \ln[1 + \exp(\hat{\boldsymbol{\beta}}^T(u_i, v_i) \mathbf{x}_i)] \quad (2.11)$$

dan  $K$  adalah banyaknya parameter dalam model [12].

Pengujian signifikansi parameter GWLR terdiri dari pengujian kesamaan model regresi logistik global dengan model GWLR, pengujian signifikansi parameter secara simultan dan pengujian signifikansi parameter secara parsial. Hipotesis pada pengujian kesamaan model regresi logistik global dengan model GWLR adalah

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

(Model regresi logistik global dan model GWLR adalah identik)

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

(Model regresi logistik global dan model GWLR adalah tidak identik)

Statistik uji diperoleh berdasarkan metode *likelihood ratio test* (LRT), yaitu

$$D = 2 \sum_{i=1}^n [\ell(\hat{\Omega}(u_i, v_i)) - \ell(\hat{\Omega})], \quad (2.12)$$

[2], [16], dimana  $\ell(\hat{\Omega}(u_i, v_i))$  adalah nilai maksimum fungsi *log-likelihood* di bawah model populasi (model GWLR lengkap) yang diberikan oleh persamaan (2.11) dan  $\ell(\hat{\Omega})$  diberikan oleh persamaan (2.3), dengan  $D \sim \chi^2_{(np)}$ .  $H_0$  ditolak pada taraf uji  $\alpha$  jika  $D > \chi^2_{(1-\alpha, p)}$  atau jika  $P_{value} < \alpha$  [21].

Uji signifikansi parameter yang kedua adalah pengujian parameter model GWLR secara simultan, dengan hipotesis adalah sebagai berikut

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

$H_1$  : Minimal ada satu  $\beta_k(u_i, v_i) \neq 0$ ,  $k = 1, 2, \dots, p$ ;  $i = 1, 2, \dots, n$  (Model GWLR layak).

Statistik uji diturunkan menggunakan metode LRT, yaitu

$$G_s = 2 \sum_{i=1}^n [\ell(\hat{\Omega}(u_i, v_i)) - \ell(\hat{\omega}(u_i, v_i))], \quad (2.13)$$

[2], [16], dengan  $G_s \sim \chi^2_{(np)}$ ,  $\ell(\hat{\Omega}(u_i, v_i))$  diberikan oleh persamaan (2.11), dan

$$\ell(\hat{\omega}(u_i, v_i)) = \sum_{i=1}^n w_{ij} \ln \hat{\pi}_0(u_i, v_i) y_i + (1 - y_i) \ln(1 - \hat{\pi}_0(u_i, v_i)), \quad (2.14)$$

$H_0$  ditolak pada taraf uji  $\alpha$  jika nilai  $G_s > \chi^2_{(1-\alpha, p)}$  atau jika  $P_{value} < \alpha$ .

Uji signifikansi parameter yang ketiga adalah pengujian parameter model GWLR secara parsial. Hipotesis pengujian parameter secara parsial  $\beta_k(u_i, v_i)$  pada  $i$  dan  $k$  tertentu,  $k=1, 2, \dots, p$ ;  $i = 1, 2, \dots, n$  adalah

$H_0$  :  $\beta_k(u_i, v_i) = 0$  Versus  $H_1$  :  $\beta_k(u_i, v_i) \neq 0$

Statistik uji adalah statistik Wald, yaitu

$$W_k(u_i, v_i) = \frac{\hat{\beta}_k(u_i, v_i)}{\sqrt{\text{Var}(\hat{\beta}_k(u_i, v_i))}}, \quad (2.15)$$

dimana  $\text{Var}(\hat{\beta}_k(u_i, v_i))$  adalah elemen diagonal utama ke- $k$  dari invers matriks Informasi Fisher pada penaksiran parameter model GWLR pada lokasi  $(u_i, v_i)$ , yang diberikan oleh

$$\mathbf{I}(\hat{\beta}(u_i, v_i)) = \mathbf{X}^T \mathbf{W}(u_i, v_i) \hat{\mathbf{V}}(u_i, v_i) \mathbf{X}. \quad (2.16)$$

dimana  $\hat{\mathbf{V}} = \text{diag}[\pi_1(u_i, v_i)[1 - \pi_1(u_i, v_i)] \quad \pi_2(u_i, v_i)[1 - \pi_2(u_i, v_i)] \quad \dots \quad \pi_n(u_i, v_i)[1 - \pi_n(u_i, v_i)]]$ ,

$\mathbf{W}(u_i, v_i) = \text{diag}[w_{i1}(u_i, v_i) \quad w_{i2}(u_i, v_i) \quad \dots \quad w_{in}(u_i, v_i)]$  dan  $\mathbf{X}$  adalah matriks data pengamatan variabel prediktor berukuran  $n \times (p + 1)$ , dengan  $W_k(u_i, v_i) \sim N(0, 1)$ ,  $n \rightarrow \infty$ .

$H_0$  ditolak pada taraf uji  $\alpha$  jika nilai  $|W_k(u_i, v_i)| > Z_{1-\alpha/2}$  atau jika  $P_{value} < \alpha$  [10], [12].

## 2.5 Ukuran Kebaikan dan Interpretasi Model GWLR

Ukuran kebaikan model regresi logistik global dan GWLR adalah AIC yang diberikan oleh persamaan (2.10) dan ukuran kebaikan tambahan adalah *McFadden R-Squared* ( $R_{MF}^2$ ). Ukuran  $R_{MF}^2$  menjelaskan besar pengaruh variabel prediktor terhadap variabel respon. Semakin besar nilai  $R_{MF}^2$  maka model regresi logistik global atau GWLR semakin baik, dengan  $0 \leq R_{MF}^2 \leq 1$ . Nilai  $R_{MF}^2$  model regresi logistik global dan model GWLR masing-masing diperoleh dengan rumus

$$R_{MF}^2 = 1 - \frac{\ell(\hat{\Omega})}{\ell(\hat{\omega})}, \quad (2.17)$$

[10], [12] dan

$$R_{MF}^2(u_i, v_i) = 1 - \frac{\ell(\widehat{\Omega}(u_i, v_i))}{\ell(\widehat{\omega}(u_i, v_i))}. \quad (2.18)$$

Interpretasi model GWLR menggunakan nilai *odds ratio*. Nilai *odds ratio* (*OR*) model GWLR berdasarkan variabel prediktor kontinu atau nominal global [12] dihitung berdasarkan formula

$$OR(X_k) = \exp(\widehat{\beta}_k(u_i, v_i)). \quad (2.19)$$

## 2.6 Variabel dan Data Penelitian

Variabel penelitian terdiri dari variabel respon (*Y*) dan variabel prediktor (*X*). Variabel penelitian ditayangkan pada Tabel 2.1.

**Tabel 2.1.** Variabel Penelitian

Notasi	Nama Variabel	Tipe Variabel	Satuan
<i>Y</i>	<i>Dissolved Oxygen</i> (DO)	Nominal Biner	-
<i>X</i> <sub>1</sub>	<i>Total Dissolved Solid</i> (TDS)	Kontinu	mg/L ( <i>ppm</i> )
<i>X</i> <sub>2</sub>	Konsentrasi Nitrat	Kontinu	<i>ppm</i> <sup>*)</sup>
<i>X</i> <sub>3</sub>	Derajat Warna Air	Kontinu	Pt-Co Unit
<i>X</i> <sub>4</sub>	pH	Kontinu	-
<i>X</i> <sub>5</sub>	Sulfat	Kontinu	<i>ppm</i>
<i>X</i> <sub>6</sub>	Suhu	Kontinu	°C

<sup>\*)</sup>*ppm* = *part per million*

Data variabel DO (*Y*) adalah nominal biner dan didefinisikan oleh

$$y_i = \begin{cases} 1, & \text{Jika nilai DO} < 6 \\ 0, & \text{Jika nilai DO} \geq 6 \end{cases} \quad (2.20)$$

Nilai  $y_i = 1$  adalah hasil pengamatan DO air sungai  $< 6$  dan diinterpretasikan air sungai tercemar [11], sehingga  $\pi_i = P(y_i = 1)$  diartikan sebagai peluang air sungai tercemar [13], [14]. Data penelitian adalah data sekunder yang diperoleh dari Dinas Lingkungan Hidup (DLH) Provinsi Kalimantan Timur. Ukuran sampel adalah 33, yaitu titik-titik lokasi pemantauan kualitas air sungai di Provinsi Kalimantan Timur yang sudah ditetapkan DLH tahun 2022.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Deskripsi Data Variabel Respon

Deskripsi data variabel respon dinyatakan dalam statistik presentase data nominal biner dari variabel respon, dimana  $y = 0$  jika nilai hasil pengamatan DO  $\geq 6$  (berarti air sungai tidak tercemar) dan  $y = 1$  jika nilai hasil pengamatan DO  $< 6$  (air sungai terindikasi tercemar) [11], [13]. Statistik deskriptif data variabel respon di 33 titik lokasi pengamatan dapat dilihat pada Tabel 3.1, sedangkan deskripsi data variabel prediktor dibahas pada sub 3.9, pada bahasan pengelompokan model GWLR berdasarkan Tabel 3.9 dan 3.10.

**Tabel 3.1.** Statistik Deskriptif Data Variabel Respon

	Tidak Tercemar ( $y_i = 0$ )	Tercemar ( $y_i = 1$ )	<i>n</i>
<b>Banyaknya lokasi pengamatan</b>	15	18	33
<b>Presentase</b>	45%	55%	

Berdasarkan Tabel 3.1, diperoleh informasi bahwa, 15 dari 33 titik pengamatan atau sebesar 45% dari titik pengamatan air sungai di Provinsi Kalimantan Timur tidak terindikasi tercemar, dan 18 titik lokasi pengamatan atau 55% terindikasi tercemar.

### 3.2 Pendeteksian Multikolinieritas dan Pemilihan Model Regresi Logistik Global Terbaik

Pendeteksian multikolinieritas pada penelitian ini menggunakan kriteria VIF. Adanya multikolinier antar variabel prediktor merupakan petunjuk awal ketidakakuratan hasil analisis model regresi. Hasil perhitungan nilai VIF berdasarkan persamaan (2.7) disajikan pada Tabel 3.2.

**Tabel 3.2.** Nilai VIF Variabel Prediktor

Variabel	Nilai VIF	Indikasi Multikolinieritas
TDS ( $X_1$ )	7,874	Tidak terjadi multikolinieritas
Konsentrasi Nitrat ( $X_2$ )	2,255	Tidak terjadi multikolinieritas
Derajat Warna Air ( $X_3$ )	1,569	Tidak terjadi multikolinieritas
pH ( $X_4$ )	5,127	Tidak terjadi multikolinieritas
Sulfat ( $X_5$ )	7,447	Tidak terjadi multikolinieritas
Suhu ( $X_6$ )	4,132	Tidak terjadi multikolinieritas

Berdasarkan nilai VIF pada Tabel 3.2, dapat disimpulkan bahwa tidak terdapat multikolinieritas antar variabel prediktor dalam model regresi. Hal ini ditunjukkan oleh nilai VIF pada setiap variabel prediktor yang kurang dari 10, sehingga pemodelan regresi dapat melibatkan enam variabel prediktor seperti pada Tabel 3.2.

Pemilihan model regresi logistik global terbaik menggunakan metode *backward elimination* dengan mengeluarkan satu persatu variabel prediktor yang memberikan nilai  $p_{value}$  terbesar yang lebih dari 0,1. Kriteria model regresi logistik terbaik adalah model yang *fit* (layak), yang memberikan nilai AIC optimal, dan variabel prediktor yang berpengaruh paling banyak (dengan nilai  $p_{value} < 0,1$ ). Berdasarkan metode *backward elimination*, diperoleh model regresi logistik terbaik adalah model dengan empat variabel prediktor, yaitu TDS ( $X_1$ ), konsentrasi nitrat ( $X_2$ ), derajat warna air ( $X_3$ ), dan pH ( $X_4$ ). Model regresi logistik dengan 4 variabel prediktor tersebut menghasilkan model yang layak dengan nilai AIC optimal yaitu sebesar 32,710, dan variabel prediktor yang berpengaruh signifikan paling banyak, yaitu  $X_2$ ,  $X_3$ , dan  $X_4$ . Pendekatan *backward elimination* dan kriteria AIC dalam pemilihan model terbaik pada penelitian ini menjadi salah satu pembeda dengan penelitian yang telah dilakukan oleh Lestari dkk [12].

### 3.3 Model Regresi Logistik Global

Model umum regresi logistik global terbaik dengan empat variabel prediktor yang mengacu pada persamaan (2.1) adalah sebagai berikut

$$\pi = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4})} \quad (4.1)$$

Model yang diberikan oleh persamaan (4.1) dinamakan model regresi logistik global. Berdasarkan algoritma iteratif Newton-Raphson diperoleh penaksir ML model regresi logistik global yang disajikan pada Tabel 3.3 berikut.



**Tabel 3.3.** Penaksir ML Parameter Model Regresi Logistik global

Parameter	Nilai Taksiran
$\beta_0$	15,243
$\beta_1$	0,007
$\beta_2$	14,117
$\beta_3$	0,031
$\beta_4$	-2,891

Berdasarkan hasil penaksiran parameter pada Tabel 3.3, diperoleh model regresi logistik global adalah

$$\hat{\pi} = \frac{\exp(15,243 + 0,007x_1 + 14,117x_2 + 0,031x_3 - 2,891x_4)}{1 + \exp(15,243 + 0,007x_1 + 14,117x_2 + 0,031x_3 - 2,891x_4)}, \quad (4.2)$$

dimana  $\hat{\pi}$  menyatakan taksiran peluang pencemaran air Sungai,  $X_1$  adalah TDS,  $X_2$  adalah konsentrasi nitrat dalam air sungai,  $X_3$  menyatakan derajat warna air, dan  $X_4$  adalah pH air sungai.

### 3.4 Pengujian Parameter Model Regresi Logistik Global Secara Simultan dan Parsial

Hipotesis pengujian parameter secara simultan adalah

$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$  (Model regresi logistik global tidak layak)

$H_1: \text{Terdapat } \beta_k \neq 0; k = 1, 2, 3, 4$  (Model regresi logistik global layak).

Statistik uji pada pengujian secara simultan adalah statistik  $G$  yang diberikan oleh persamaan (2.2).

Hasil perhitungan nilai statistik  $G$  dan  $p_{value}$  dapat dilihat pada Tabel 3.4.

**Tabel 3.4.** Hasil Pengujian Parameter Model Regresi Logistik Global Secara Simultan

$G$	$\chi^2_{(0,9;4)}$	$p_{value}$	Keputusan Uji
22,762	7,779	0,0001	$H_0$ ditolak

Berdasarkan Tabel 3.4, diketahui  $G = 22,762 > \chi^2_{(0,9;4)} = 7,779$  dan  $p_{value} = 0,0001 < \alpha = 0,1$  sehingga diputuskan  $H_0$  ditolak. Kesimpulan uji simultan adalah model regresi logistik global layak atau variabel TDS, konsentrasi nitrat, derajat warna air, dan pH secara simultan berpengaruh terhadap peluang pencemaran air sungai.

Hipotesis pengujian parameter secara parsial parameter  $\beta_k$  untuk  $k$  tertentu;  $k = 0, 1, 2, 3, 4$  adalah

$H_0: \beta_k = 0$  (Variabel prediktor  $X_k$  tidak berpengaruh terhadap peluang pencemaran air sungai).

$H_1: \beta_k \neq 0$  (Variabel prediktor  $X_k$  berpengaruh terhadap peluang pencemaran air sungai).

Statistik uji diberikan oleh persamaan (2.5), dan hasil perhitungan statistik  $W_k$  dan  $p_{value}$  dapat dilihat pada Tabel 3.5.

**Tabel 3.5.** Hasil Pengujian Parameter Model Regresi Logistik Global Secara Parsial

Variabel	$\beta$	Nilai Taksiran	SE	$W_k$	$p_{value}$	Keputusan Uji
-	$\beta_0$	15,243	11,442	1,332	0,183	$H_0$ gagal ditolak
TDS ( $X_1$ )	$\beta_1$	0,007	0,011	0,649	0,516	$H_0$ gagal ditolak
Konsentrasi Nitrat ( $X_2$ )	$\beta_2$	14,117	8,295	1,702	0,089	$H_0$ ditolak
Derajat Warna Air ( $X_3$ )	$\beta_3$	0,031	0,014	2,291	0,022	$H_0$ ditolak
pH ( $X_4$ )	$\beta_4$	-2,891	1,729	-1,672	0,094	$H_0$ ditolak

Berdasarkan Tabel 3.5, diperoleh nilai statistik uji  $W_2$ ,  $W_3$  dan  $W_4$  berturut-turut untuk variabel konsentrasi nitrat, derajat warna air, dan pH air masing-masing lebih besar dari nilai kritis  $Z_{(0,95)} = 1,645$ , dan  $p_{value}$  masing-masing variabel tersebut kurang dari 0,1, sehingga keputusan uji adalah  $H_0$  ditolak. Kesimpulan uji adalah variabel konsentrasi nitrat, derajat warna air, dan pH air masing-masing berpengaruh secara individual terhadap peluang pencemaran air sungai. Berdasarkan Tabel 3.5 bahwa variabel TDS tidak berpengaruh. Hal ini didukung oleh nilai statistik uji  $W_1 = 0,649 < Z_{(0,95)} = 1,645$  dan  $p_{value} = 0,516 > \alpha = 0,1$ , sehingga diputuskan  $H_0$  gagal ditolak. Kesimpulan uji adalah variabel TDS tidak berpengaruh secara individual terhadap peluang pencemaran air sungai. Berdasarkan hasil pengujian secara simultan dan parsial, model regresi logistik adalah model yang layak dengan variabel prediktor berpengaruh  $X_2$ ,  $X_3$ , dan  $X_4$ , serta memberikan nilai AIC sebesar 32,712, dan nilai *McFadden R-Squared* sebesar 50,1%. Bahasan berikutnya adalah pengujian heterogenitas spasial pada data DO, sebagai tahap awal dalam pemodelan GWLR untuk mendapatkan model yang lebih baik, dari pada model global.

### 3.5 Pengujian Heterogenitas Spasial

Hipotesis pengujian heterogenitas spasial adalah

$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_{33}^2 = \sigma^2$  (Data DO bukan data heterogenitas spasial)

$H_1$ : Minimal ada satu  $\sigma_k^2 \neq \sigma^2$ ;  $k = 1,2,3, \dots, 33$  (Data DO adalah data heterogenitas spasial).

Statistik uji diberikan oleh persamaan (2.8), dan hasil perhitungan nilai statistik  $F_2$  dan  $p_{value}$  dapat dilihat pada Tabel 3.6.

**Tabel 3.6.** Hasil Pengujian Heterogenitas Spasial

$F_2$	$F_{(0,9;4;28)}$	$p_{value}$	Keputusan Uji
15,16	2,157	$1,041 \times 10^{-6}$	$H_0$ ditolak

Berdasarkan Tabel 3.6, diketahui nilai  $F_2 = 3,410 > F_{(0,9;4;28)} = 2,157$  dan nilai  $p_{value} = 1,041 \times 10^{-6} < \alpha = 0,1$ , sehingga diputuskan  $H_0$  ditolak dan disimpulkan data DO adalah data heterogenitas spasial. Berdasarkan hasil pengujian ini data DO adalah heterogenitas spasial, sehingga data DO dapat dimodelkan menggunakan model GWLR.

### 3.6 Model Geographically Weighted Logistic Regression (GWLR)

Model umum GWLR pada lokasi ke- $i$  dengan empat variabel prediktor adalah

$$\pi_i(u_i, v_i) = \frac{\exp(\beta_0(u_i, v_i) + \beta_1(u_i, v_i)x_{i1} + \beta_2(u_i, v_i)x_{i2} + \beta_3(u_i, v_i)x_{i3} + \beta_4(u_i, v_i)x_{i4})}{1 + \exp(\beta_0(u_i, v_i) + \beta_1(u_i, v_i)x_{i1} + \beta_2(u_i, v_i)x_{i2} + \beta_3(u_i, v_i)x_{i3} + \beta_4(u_i, v_i)x_{i4})} \quad (4.2)$$

dengan  $i = 1,2, \dots, 33$ . Penaksiran parameter model GWLR menggunakan metode MLE dengan mempertimbangkan efek spasial, melalui pemberian pembobot spasial pada fungsi *likelihood*. Berdasarkan metode iteratif Newton-Raphson diperoleh penaksir ML parameter model GWLR, nilai statistik  $W_k(u_i, v_i)$ ,  $p_{value}$  dan nilai *odds ratio* variabel berpengaruh untuk setiap lokasi pengamatan disajikan pada Tabel 3.9 (pada halaman berikutnya).

### 3.7 Pengujian Kesamaan Model Regresi Logistik Global dan Model GWLR

Hipotesis pengujian kesamaan model regresi logistik global dan model GWLR adalah

$H_0: \beta_k(u_1, v_1) = \beta_k(u_2, v_2) \dots = \beta_k(u_{33}, v_{33}) = \beta_k$ ;  $k = 1,2,3,4$

(Model regresi logistik global dan model GWLR adalah identik).

$H_1$ : Minimal ada satu  $\beta_k(u_i, v_i) \neq 0$ ;  $k = 1,2,3,4$ ;  $i = 1,2,3, \dots, 33$

(Model regresi logistik global dan model GWLR adalah tidak identik).

Statistik uji diberikan oleh persamaan (2.12), dan hasil perhitungan nilai statistik uji dan  $p_{value}$  dapat dilihat pada Tabel 3.7

Tabel 3.7. Hasil Pengujian Hipotesis Kesamaan Model

D	$\chi^2_{(0,9;132)}$	Pvalue	Keputusan Uji
258,040	153,204	$3,480 \times 10^{-10}$	H <sub>0</sub> ditolak

Berdasarkan Tabel 3.7, diketahui nilai  $D = 258,040 > \chi^2_{(0,9;132)} = 153,204$  dan nilai  $p_{value} = 3,480 \times 10^{-10} < \alpha = 0,1$  sehingga diputuskan H<sub>0</sub> ditolak pada taraf signifikansi 10%. Kesimpulan uji kesamaan model regresi logistik global dan model GWLR adalah model regresi logistik global tidak identik dengan model GWLR. Keputusan uji menolak hipotesis nol mengkonfirmasi bahwa model yang sesuai berdasarkan data sampel adalah model lokal atau model GWLR. Pada penelitian ini statistik uji kesamaan model, dan uji simultan dikembangkan berdasarkan metode LRT, sedangkan pada penelitan Lestari [12] statistik uji diturunkan berdasarkan devian.

### 3.8 Pengujian Parameter Model GWLR Secara Simultan

Hipotesis pengujian parameter Model GWLR secara simultan adalah

$$H_0: \beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \beta_3(u_i, v_i) = \beta_4(u_i, v_i) = 0; i = 1,2,3 \dots, 33$$

(Model GWLR tidak layak)

$$H_1: \text{Minimal ada satu } \beta_k(u_i, v_i) \neq 0; k = 1,2,3,4; i = 1,2,3, \dots, 33$$

(Model GWLR layak).

Statistik uji diberikan oleh persamaan (2.13), dan hasil perhitungan nilai statistik  $G_s$  dan  $p_{value}$  dapat dilihat pada Tabel 3.8. Berdasarkan Tabel 3.8, nilai  $G_s = 773,478 > \chi^2_{(0,9;132)} = 153,204$ , dan nilai  $p_{value} = 0 < \alpha = 0,1$  sehingga diputuskan H<sub>0</sub> ditolak pada taraf uji 10%. Kesimpulan uji simultan adalah model GWLR layak atau variabel TDS, konsentrasi nitrat, derajat warna air, dan pH secara simultan berpengaruh terhadap peluang pencemaran air sungai di setiap lokasi pengamatan.

Tabel 3.8. Hasil Pengujian Hipotesis Parameter Model GWLR Secara Simultan

$G_s$	$\chi^2_{(0,9;132)}$	Pvalue	Keputusan Uji
773,478	153,204	0	H <sub>0</sub> ditolak

Berdasarkan nilai statistik uji pada Tabel 3.8, keputusan uji adalah menerima H<sub>1</sub>, hal ini mengkonfirmasi bahwa model GWLR telah memenuhi pemodelan regresi lokal, dimana minimal terdapat satu nilai parameter model lokal (selain intersep) signifikan berbeda dengan nol, dan berarti terdapat variabel prediktor berpengaruh.

### 3.9 Pengujian Parameter Model GWLR Secara Parsial

Pengujian parameter regresi secara parsial untuk mengetahui variabel bebas yang berpengaruh terhadap peluang pencemaran air sungai di Kalimantan Timur. Hipotesis pengujian parameter secara parsial  $\beta_k(u_i, v_i)$  pada  $i$  dan  $k$  tertentu,  $k = 1,2, \dots, p; i = 1,2, \dots, 33$  adalah

$$H_0: \beta_k(u_i, v_i) = 0 \text{ Versus } H_1: \beta_k(u_i, v_i) \neq 0$$

Hasil perhitungan penaksiran parameter, nilai statistik uji  $W_k(u_i, v_i)$  dan  $p_{value}$  pengujian parameter model GWLR secara parsial dapat dilihat pada Tabel 3.9. Variabel prediktor berpengaruh dapat dilihat berdasarkan nilai  $p_{value}$  bertanda (\*) pada Tabel 3.9.

Tabel 3.9. Penaksir ML dari Parameter, Nilai  $W_k(u_i, v_i)$ ,  $p_{value}$  dan Odds Ratio

Lokasi (i)	Variabel	Parameter	Taksiran	SE	$W_k(u_i, v_i)$	pvalue	Odds Ratio
1	-	$\beta_0$	15,356	12,796	1,200	0,230	-
	$X_1$	$\beta_1$	0,011	0,012	0,860	0,390	-
	$X_2$	$\beta_2$	16,584	9,846	1,684	0,092*	$1,593 \times 10^7$
	$X_3$	$\beta_3$	0,022	0,014	1,563	0,118	-
	$X_4$	$\beta_4$	-2,968	1,986	-1,495	0,135	-
2	-	$\beta_0$	15,366	12,445	1,235	0,217	-
	$X_1$	$\beta_1$	0,010	0,012	0,816	0,414	-
	$X_2$	$\beta_2$	15,755	9,311	1,692	0,091*	$6,958 \times 10^6$
	$X_3$	$\beta_3$	0,023	0,014	1,667	0,096*	1,024
	$X_4$	$\beta_4$	-2,945	1,915	-1,538	0,124	-
3	-	$\beta_0$	19,602	18,326	1,070	0,285	-
	$X_1$	$\beta_1$	0,006	0,014	0,406	0,685	-
	$X_2$	$\beta_2$	18,253	12,609	1,448	0,148	-
	$X_3$	$\beta_3$	0,031	0,018	1,675	0,094*	1,031
	$X_4$	$\beta_4$	-3,514	2,722	-1,291	0,197	-
4	-	$\beta_0$	19,648	18,368	1,070	0,285	-
	$X_1$	$\beta_1$	0,006	0,014	0,402	0,688	-
	$X_2$	$\beta_2$	18,214	12,596	1,446	0,148	-
	$X_3$	$\beta_3$	0,031	0,018	1,674	0,094*	1,031
	$X_4$	$\beta_4$	-3,519	2,726	-1,291	0,197	-
5	-	$\beta_0$	19,702	18,352	1,074	0,283	-
	$X_1$	$\beta_1$	0,006	0,014	0,396	0,692	-
	$X_2$	$\beta_2$	18,091	12,475	1,450	0,147	-
	$X_3$	$\beta_3$	0,031	0,018	1,679	0,093*	1,031
	$X_4$	$\beta_4$	-3,521	2,719	-1,295	0,195	-
6	-	$\beta_0$	19,682	18,368	1,072	0,284	-
	$X_1$	$\beta_1$	0,006	0,014	0,398	0,690	-
	$X_2$	$\beta_2$	18,147	12,534	1,448	0,148	-
	$X_3$	$\beta_3$	0,031	0,018	1,676	0,094*	1,031
	$X_4$	$\beta_4$	-3,521	2,723	-1,293	0,196	-

Keterangan: \*) variabel prediktor berpengaruh pada pada taraf signifikansi 10%

**Tabel 3.9.** Penaksir ML dari Parameter, Nilai  $W_k(u_i, v_i)$ , pvalue dan Odds Ratio (lanjutan)

Lokasi (i)	Variabel	Parameter	Taksiran	SE	$W_k(u_i, v_i)$	Pvalue	Odds Ratio
7	-	$\beta_0$	15,277	14,017	1,090	0,276	-
	$X_1$	$\beta_1$	0,011	0,012	0,914	0,361	-
	$X_2$	$\beta_2$	18,805	10,760	1,748	0,081*	$1,468 \times 10^8$
	$X_3$	$\beta_3$	0,025	0,015	1,671	0,095*	1,026
	$X_4$	$\beta_4$	-3,001	2,179	-1,377	0,168	-
8	-	$\beta_0$	17,961	16,611	1,081	0,280	-
	$X_1$	$\beta_1$	0,016	0,015	1,082	0,279	-
	$X_2$	$\beta_2$	23,928	13,650	1,753	0,080*	$2,466 \times 10^{10}$
	$X_3$	$\beta_3$	0,022	0,017	1,351	0,177	-
	$X_4$	$\beta_4$	-3,531	2,632	-1,341	0,180	-
9	-	$\beta_0$	15,242	13,224	1,153	0,249	-
	$X_1$	$\beta_1$	0,007	0,012	0,577	0,564	-
	$X_2$	$\beta_2$	13,202	8,591	1,537	0,124	-
	$X_3$	$\beta_3$	0,024	0,014	1,653	0,098*	1,024
	$X_4$	$\beta_4$	-2,837	1,966	-1,443	0,149	-
10	-	$\beta_0$	15,466	13,938	1,110	0,267	-
	$X_1$	$\beta_1$	0,013	0,013	0,999	0,318	-
	$X_2$	$\beta_2$	19,367	11,099	1,745	0,081*	$2,576 \times 10^8$
	$X_3$	$\beta_3$	0,021	0,015	1,442	0,149	-
	$X_4$	$\beta_4$	-3,056	2,186	-1,398	0,162	-
11	-	$\beta_0$	16,813	14,522	1,158	0,247	-
	$X_1$	$\beta_1$	0,015	0,014	1,082	0,279	-
	$X_2$	$\beta_2$	21,598	12,463	1,733	0,083*	$2,397 \times 10^9$
	$X_3$	$\beta_3$	0,022	0,015	1,430	0,153	-
	$X_4$	$\beta_4$	-3,336	2,317	-1,440	0,150	-
12	-	$\beta_0$	16,708	15,696	1,064	0,287	-
	$X_1$	$\beta_1$	0,005	0,014	0,370	0,711	-
	$X_2$	$\beta_2$	12,294	9,287	1,324	0,186	-
	$X_3$	$\beta_3$	0,026	0,016	1,663	0,096*	1,027
	$X_4$	$\beta_4$	-2,993	2,284	-1,310	0,190	-
13	-	$\beta_0$	16,148	14,809	1,090	0,276	-
	$X_1$	$\beta_1$	0,006	0,013	0,423	0,672	-
	$X_2$	$\beta_2$	12,371	8,934	1,385	0,166	-
	$X_3$	$\beta_3$	0,025	0,015	1,661	0,097*	1,026
	$X_4$	$\beta_4$	-2,923	2,166	-1,350	0,177	-
14	-	$\beta_0$	16,438	14,406	1,141	0,254	-
	$X_1$	$\beta_1$	0,015	0,014	1,066	0,286	-
	$X_2$	$\beta_2$	21,110	12,138	1,739	0,082*	$1,472 \times 10^9$
	$X_3$	$\beta_3$	0,022	0,015	1,428	0,153	-
	$X_4$	$\beta_4$	-3,262	2,289	-1,425	0,154	-
15	-	$\beta_0$	15,338	12,827	1,196	0,232	-
	$X_1$	$\beta_1$	0,011	0,012	0,880	0,379	-
	$X_2$	$\beta_2$	16,860	9,757	1,728	0,084*	$2,100 \times 10^7$
	$X_3$	$\beta_3$	0,024	0,014	1,660	0,097	-
	$X_4$	$\beta_4$	-2,972	1,986	-1,497	0,134	-

**JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI**  
**Adelia Miranda, Suyitno, Meirinda Fauziyah**

**Tabel 3.9.** Penaksir ML dari Parameter, Nilai  $W_k(u_j, v_i)$ ,  $p_{value}$  dan *Odds Ratio* (lanjutan)

Lokasi (i)	Variabel	Parameter	Taksiran	SE	$W_k(u_i, v_i)$	pvalue	Odds Ratio
16	-	$\beta_0$	15,026	13,928	1,079	0,281	-
	$X_1$	$\beta_1$	0,013	0,013	0,968	0,333	-
	$X_2$	$\beta_2$	18,803	10,763	1,747	0,081*	$1,465 \times 10^8$
	$X_3$	$\beta_3$	0,021	0,015	1,424	0,155	-
	$X_4$	$\beta_4$	-2,964	2,171	-1,365	0,172	-
17	-	$\beta_0$	17,540	14,999	1,169	0,242	-
	$X_1$	$\beta_1$	0,017	0,015	1,119	0,263	-
	$X_2$	$\beta_2$	23,009	13,360	1,722	0,085*	$9,834 \times 10^9$
	$X_3$	$\beta_3$	0,022	0,016	1,403	0,160	-
	$X_4$	$\beta_4$	-3,490	2,415	-1,445	0,148	-
18	-	$\beta_0$	19,603	18,334	1,069	0,285	-
	$X_1$	$\beta_1$	0,006	0,014	0,406	0,685	-
	$X_2$	$\beta_2$	18,259	12,619	1,447	0,148	-
	$X_3$	$\beta_3$	0,031	0,018	1,674	0,094*	1,031
	$X_4$	$\beta_4$	-3,514	2,723	-1,290	0,197	-
19	-	$\beta_0$	19,569	18,291	1,070	0,285	-
	$X_1$	$\beta_1$	0,006	0,014	0,409	0,682	-
	$X_2$	$\beta_2$	18,275	12,610	1,449	0,147	-
	$X_3$	$\beta_3$	0,031	0,018	1,675	0,094*	1,031
	$X_4$	$\beta_4$	-3,510	2,718	-1,291	0,197	-
20	-	$\beta_0$	14,281	13,764	1,038	0,299	-
	$X_1$	$\beta_1$	0,010	0,012	0,827	0,408	-
	$X_2$	$\beta_2$	16,615	9,647	1,722	0,085*	$1,643 \times 10^7$
	$X_3$	$\beta_3$	0,020	0,014	1,405	0,160	-
	$X_4$	$\beta_4$	-2,765	2,100	-1,317	0,188	-
21	-	$\beta_0$	14,182	13,867	1,023	0,306	-
	$X_1$	$\beta_1$	0,010	0,012	0,807	0,420	-
	$X_2$	$\beta_2$	16,407	9,571	1,714	0,087*	$1,334 \times 10^7$
	$X_3$	$\beta_3$	0,020	0,014	1,384	0,166	-
	$X_4$	$\beta_4$	-2,740	2,110	-1,298	0,194	-
22	-	$\beta_0$	14,588	12,526	1,165	0,244	-
	$X_1$	$\beta_1$	0,009	0,012	0,695	0,487	-
	$X_2$	$\beta_2$	14,217	9,095	1,563	0,118	-
	$X_3$	$\beta_3$	0,024	0,014	1,648	0,099*	1,024
	$X_4$	$\beta_4$	-2,800	1,909	-1,467	0,142	-
23	-	$\beta_0$	14,616	12,492	1,170	0,242	-
	$X_1$	$\beta_1$	0,009	0,012	0,695	0,487	-
	$X_2$	$\beta_2$	14,199	9,055	1,568	0,117	-
	$X_3$	$\beta_3$	0,024	0,014	1,666	0,096*	1,024
	$X_4$	$\beta_4$	-2,804	1,903	-1,474	0,141	-
24	-	$\beta_0$	17,273	15,420	1,120	0,263	-
	$X_1$	$\beta_1$	0,017	0,015	1,129	0,259	-
	$X_2$	$\beta_2$	23,581	13,602	1,734	0,083*	$1,743 \times 10^{10}$
	$X_3$	$\beta_3$	0,021	0,016	1,348	0,178	-
	$X_4$	$\beta_4$	-3,456	2,478	-1,395	0,163	-

**JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI**  
**Adelia Miranda, Suyitno, Meirinda Fauziah**

**Tabel 3.9.** Penaksir ML dari Parameter, Nilai  $W_k(u_i, v_i)$ ,  $p_{value}$  dan *Odds ratio* (lanjutan)



Lokasi (i)	Variabel	Parameter	Taksiran	SE	$W_k(u_i, v_i)$	Pvalue	Odds Ratio
25	-	$\beta_0$	17,294	15,480	1,117	0,264	-
	$X_1$	$\beta_1$	0,017	0,015	1,130	0,258	-
	$X_2$	$\beta_2$	23,678	13,657	1,734	0,083*	$1,919 \times 10^{10}$
	$X_3$	$\beta_3$	0,021	0,016	1,343	0,179	-
	$X_4$	$\beta_4$	-3,461	2,487	-1,392	0,164	-
26	-	$\beta_0$	15,381	14,478	1,062	0,288	-
	$X_1$	$\beta_1$	0,011	0,012	0,885	0,376	-
	$X_2$	$\beta_2$	19,020	10,946	1,738	0,082*	$1,821 \times 10^8$
	$X_3$	$\beta_3$	0,026	0,015	1,649	0,099*	1,026
	$X_4$	$\beta_4$	-3,007	2,242	-1,341	0,180	-
27	-	$\beta_0$	19,452	17,550	1,108	0,268	-
	$X_1$	$\beta_1$	0,016	0,015	1,065	0,287	-
	$X_2$	$\beta_2$	24,991	14,386	1,737	0,082*	$7,136 \times 10^{10}$
	$X_3$	$\beta_3$	0,024	0,017	1,360	0,174	-
	$X_4$	$\beta_4$	-3,770	2,782	-1,355	0,175	-
28	-	$\beta_0$	16,701	15,684	1,065	0,287	-
	$X_1$	$\beta_1$	0,005	0,014	0,371	0,711	-
	$X_2$	$\beta_2$	12,295	9,278	1,325	0,185	-
	$X_3$	$\beta_3$	0,026	0,016	1,662	0,097*	1,027
	$X_4$	$\beta_4$	-2,992	2,282	-1,311	0,190	-
29	-	$\beta_0$	14,730	12,664	1,163	0,245	-
	$X_1$	$\beta_1$	0,009	0,012	0,761	0,447	-
	$X_2$	$\beta_2$	15,183	8,889	1,708	0,088*	$3,924 \times 10^6$
	$X_3$	$\beta_3$	0,023	0,014	1,652	0,099*	1,023
	$X_4$	$\beta_4$	-2,813	1,922	-1,463	0,143	-
30	-	$\beta_0$	14,752	12,704	1,161	0,246	-
	$X_1$	$\beta_1$	0,009	0,012	0,739	0,460	-
	$X_2$	$\beta_2$	14,939	8,797	1,698	0,089*	$3,076 \times 10^6$
	$X_3$	$\beta_3$	0,023	0,014	1,649	0,099*	1,023
	$X_4$	$\beta_4$	-2,807	1,922	-1,460	0,144	-
31	-	$\beta_0$	15,630	12,607	1,240	0,215	-
	$X_1$	$\beta_1$	0,010	0,012	0,838	0,402	-
	$X_2$	$\beta_2$	16,235	9,639	1,684	0,092*	$1,124 \times 10^7$
	$X_3$	$\beta_3$	0,024	0,014	1,658	0,097*	1,024
	$X_4$	$\beta_4$	-3,004	1,950	-1,540	0,124	-
32	-	$\beta_0$	15,581	12,546	1,242	0,214	-
	$X_1$	$\beta_1$	0,010	0,012	0,829	0,407	-
	$X_2$	$\beta_2$	16,064	9,537	1,684	0,092*	$9,472 \times 10^6$
	$X_3$	$\beta_3$	0,024	0,014	1,677	0,094*	1,024
	$X_4$	$\beta_4$	-2,991	1,938	-1,544	0,123	-
33	-	$\beta_0$	16,457	15,341	1,073	0,283	-
	$X_1$	$\beta_1$	0,005	0,014	0,387	0,699	-
	$X_2$	$\beta_2$	12,244	9,142	1,339	0,180	-
	$X_3$	$\beta_3$	0,026	0,016	1,664	0,096*	1,027
	$X_4$	$\beta_4$	-2,960	2,236	-1,324	0,186	-

Sebagai contoh, berdasarkan Tabel 3.9 bahwa variabel prediktor berpengaruh terhadap peluang pencemaran air sungai di lokasi ke-14 (Kantor Gubernur Samarinda) adalah konsentrasi nitrat. Hal ini dikonfirmasi oleh nilai  $|W_2(u_{14}, v_{14})| = 1,73$  lebih besar dari nilai kritis  $Z_{0,95} = 1,645$  dan  $p_{value} = 0,082 < \alpha = 0,1$ , sehingga disimpulkan konsentrasi nitrat berpengaruh terhadap peluang pencemaran air sungai di lokasi pengamatan Kantor Gubernur. Variabel prediktor berpengaruh untuk model GWLR di lokasi yang lain dapat diperoleh dengan cara yang sama. Berdasarkan hasil pengujian parameter secara parsial pada Tabel 3.9, model GWLR dapat dikelompokkan menjadi tiga kelompok berdasarkan variabel prediktor yang berpengaruh signifikan terhadap peluang pencemaran air sungai. Hasil pengelompokkan model GWLR dapat dilihat pada Tabel 3.10.

**Tabel 3.10.** Pengelompokkan Model GWLR Berdasarkan Variabel Prediktor Berpengaruh

Kelompok	Variabel Berpengaruh	Lokasi ( <i>i</i> )
1	$X_2$	1). Titik Nursery Suring PPU, 8). Jembatan Marangkayu Santan Ulu Kukar; 10). Pulau Kumala Kukar; 11). Palaran Samarinda; 14). Kantor Gubernur Samarinda; 15). Kalamur Samarinda; 16). Bloro Kukar; 17). Anggana Kukar; 20). Desa Siran Kukar; 21). Desa Kedang Kepala Kukar; 24). Jembatan Desa Pampang Samarinda; 25). Pampang Dalam Samarinda; dan 27). Jembatan Jl. Poros Bontang-Samarinda Km.8 Kutai Timur.
2	$X_3$	3). Keraton Gunung Tabur Berau; 4). Jembatan Gunung Tabur Berau; 5). Hulu PT. BBE Berau; 6). Hilir Berau Coal; 9). Tering Kubar; 12). Nyan Mahulu; 13). Long Bagun Mahulu; 18). Keraton Sambaliung Berau; 19). Kantor Bupati Berau; 22). Titik Hulu Sungai Kasungai Paser; 23). Titik Hilir Sungai Kasungai; 28). Muara Sungai Boh Mahulu; dan 33). Batoq Kelo Mahulu
3	$X_2$ dan $X_3$	2). Titik Jembatan Sei Sepaku; 7). Jembatan Santan Tengah Kukar; 26). Jembatan Jl. Soekarno-Hatta Bontang; 29). Muara Sungai Belayan Kukar; 30). Desa Sebelimbingan Kukar; 31). Titik Hulu Sungai 28; dan 32). Titik Hilir (Pos Pon 1) Balikpapan

Berdasarkan Tabel 3.10, faktor yang berpengaruh terhadap peluang air sungai tercemar di titik-titik pengamatan pada kelompok 1 adalah konsentrasi nitrat. Deskripsi (kimiawi) umum air sungai pada kelompok 1 adalah sebagai berikut: rata-rata konsentrasi nitrat sebesar 0,48 ppm (di atas rata-rata data sampel sebesar 0,32 ppm), rata-rata DO sebesar 4,71 ppm (di bawah rata-rata DO data sampel sebesar 5,36 ppm), rata-rata TDS sebesar 72,20 ppm (di atas rata-rata data sampel sebesar 59,76 ppm), rata-rata derajat warna air sebesar 97,23 Pt-Co Unit (di atas rata-rata data sampel sebesar 86,79 Pt-Co Unit), dan pH air sebesar 6,99 (di bawah rata-rata data sampel sebesar 7,08). Berdasarkan TDS, konsentrasi nitrat dan ukuran derajat warna, diperoleh informasi bahwa air sungai di kelompok 1 relatif kurang jernih [13], dengan kandungan nitrat relatif tinggi (di atas rata-rata), serta terindikasi kurang baik kualitasnya berdasarkan DO yang jauh di bawah angka baku mutu 6 ppm [13].

Faktor yang berpengaruh terhadap peluang air sungai tercemar di titik-titik pengamatan pada kelompok 2 adalah derajat warna air. Rata-rata derajat warna air pada kelompok ini sebesar 76,46 Pt-Co Unit (di bawah rata-rata data sampel), rata-rata DO sebesar 5,53 ppm (di atas rata-rata data sampel), rata-rata TDS sebesar 47,97 ppm, (di bawah rata-rata data sampel), rata-rata konsentrasi nitrat sebesar 0,27 ppm (di bawah rata-rata data sampel), dan pH air sebesar 7,21 (di atas rata-rata data sampel). Berdasarkan TDS dan ukuran derajat warna [13], air sungai pada

kelompok 1 relatif lebih jernih dibandingkan dengan air sungai pada kelompok 1, dan kualitas air sedikit kurang baik berdasarkan DO yang sedikit di bawah 6 ppm [13], tetapi kualitasnya lebih baik dari pada air sungai pada kelompok 1.

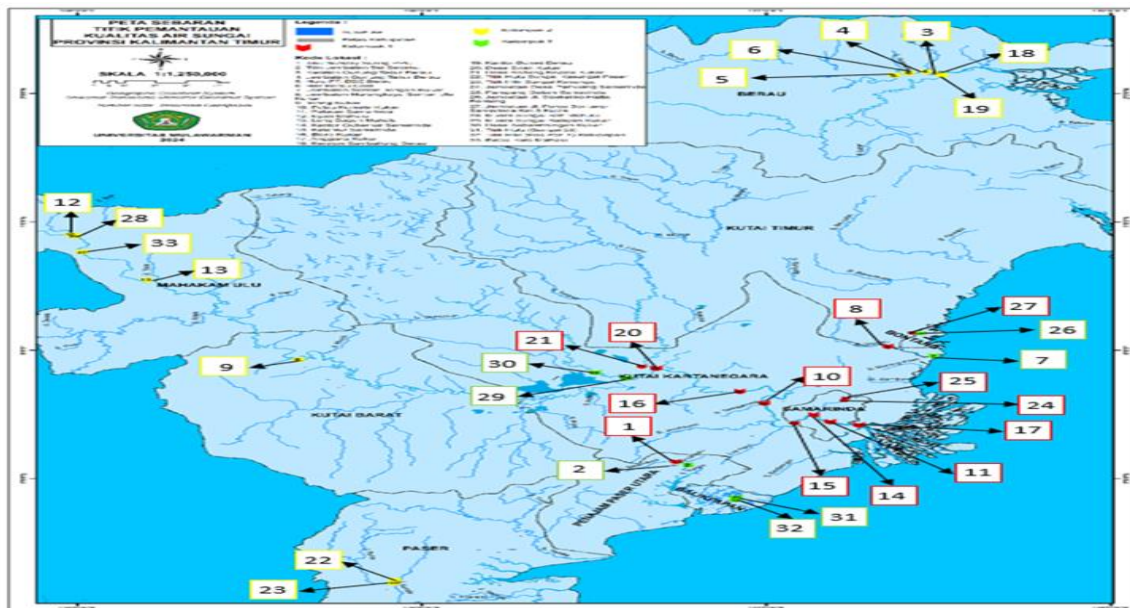
Faktor yang yang berpengaruh terhadap peluang air sungai tercemar di titik-titik pengamatan pada kelompok 3 adalah konsentrasi nitrat dan derajat warna air. Rata-rata konsentrasi nitrat air sungai pada kelompok 3 sebesar 0,11 ppm ( di bawah rata-rata data sampel), rata-rata derajat warna sebesar 86,57 Pt-Co Unit (sedikit di bawah rata-rata data sampel). Rata-rata DO sebesar 6,27 ppm (di atas rata-rata data sampel), TDS sebesar 58,54 ppm (sedikit di bawah rata-rata data sampel), dan pH air sebesar 6.99 (sedikit di bawah rata-rata data sampel). Berdasarkan TDS dan ukuran derajat warna air, bahwa air sungai pada kelompok 3 relatif jernih dibanding dengan air sungai pada kelompok 1, serta didasarkan pada DO dan konsentrasi nitrat, kualitas air sudah baik dengan DO lebih besar dari 6 ppm [13].

Pengelompokan model GWLR berdasarkan variabel prediktor (faktor) yang berpengaruh pada Tabel 3.10 merupakan pemetaan kualitas air yang dapat diinformasikan kepada pemerintah, sebagai pertimbangan dalam menyusun kebijakan pencegahan pencemaran air sungai di Provinsi Kalimantan Timur. Pengelompokan model GWLR berdasarkan variabel prediktor (faktor) berpengaruh pada Tabel 3.10 dapat divisualisasikan menggunakan peta yang ditunjukkan pada Gambar 3.1. Model GWLR di titik-titik lokasi pengamatan yang berwarna merah pada Gambar 3.1 adalah model GWLR kelompok 1 dengan faktor yang berpengaruh adalah konsentrasi nitrat. Model GWLR Kelompok 1 adalah model GWLR di titik-titik pengamatan Nursery Suring PPU, Jembatan Marangkayu Santan Ulu Kukar, Pulau Kumala Kukar, Palaran Samarinda, Kantor Gubernur Samarinda, Kalamur Samarinda, Bloro Kukar, Anggana Kukar, Desa Siran Kukar, Desa Kedang Kepala Kukar, Jembatan Desa Pampang Samarinda, Pampang Dalam Samarinda, dan Jembatan Jl. Poros Bontang-Samarinda Km.8 Kutai Timur.

Model GWLR di titik-titik lokasi pengamatan yang berwarna kuning pada Gambar 3.1 adalah model GWLR kelompok 2 dengan faktor yang berpengaruh adalah derajat warna air. Model Kelompok 2 adalah model GWLR di titik-titik pengamatan Keraton Gunung Tabur Berau, Jembatan Gunung Tabur Berau, Hulu PT. BBE Berau, Hilir Berau Coal, Tering Kubar, Nyan Mahulu, Long Bagun Mahulu, Keraton Sambaliung Berau, Kantor Bupati Berau, Titik Hulu Sungai Kasungai Paser, Titik Hilir Sungai Kasungai, Muara Sungai Boh Mahulu, dan Batoq Kelo Mahulu.

Model GWLR di titik-titik lokasi pengamatan yang berwarna hijau pada Gambar 3.1 adalah model GWLR kelompok 3 dengan faktor yang berpengaruh adalah konsentrasi nitrat dan derajat warna air. Model Kelompok 3 adalah model GWLR di titik-titik pengamatan Titik Jembatan Sei Sepaku, Jembatan Santan Tengah Kukar, Jembatan Jl. Soekarno-Hatta Bontang, Muara Sungai Belayan Kukar, Desa Sebelimbangan Kukar, Titik Hulu Sungai 28, dan Titik Hilir (Pos Pon 1) Balikpapan.

Berdasarkan peta pada Gambar 3.1, terlihat bahwa variabel prediktor yang berpengaruh terhadap model GWLR di titik-titik pengamatan yang berdekatan pada umunya sama (ditandai warna titik-titik pengamatan yang berdekatan sama). Atau dapat ditegaskan bahwa peluang pencemaran air sungai di titik-titik lokasi pengamatan yang berdekatan dipengaruhi oleh faktor yang sama. Berdasarkan Tabel 3.10 dan Gambar 3.1, bahwa faktor yang berpengaruh terhadap peluang pencemaran air sungai di Kalimantan Timur bersifat lokal, hal ini dapat menjadi pertimbangan Pemerintah Provinsi Kalimantan Timur atau Pemerintah Kabupaten/Kota dalam menyusun program pencegahan pencemaran air sungai. Program yang diimplementasikan disarankan tidak seragam untuk setiap wilayah, dan memperhatikan karakteristik setiap sungai dan lingkungannya.



Gambar 3.1. Pemetaan Model GWLR Berdasarkan Variabel Prediktor Berpengaruh

### 3.10 Pemilihan Model Terbaik dan Interpretasi Model

Berdasarkan uji parameter secara simultan dan parsial bahwa model regresi logistik global merupakan model yang layak. Pemilihan model ini bertujuan untuk memilih model yang lebih baik diantara model GWLR dan model regresi logistik global. Ukuran kebaikan model pada penelitian ini adalah nilai AIC, dan sebagai ukuran kebaikan tambahan adalah  $R_{MF}^2$ . Nilai AIC dan  $R_{MF}^2$  kedua model dapat dilihat pada Tabel 3.11.

Model	AIC	$R_{MF}^2$
Regresi Logistik Global	32,712	0,501
GWLR	29,629	0,611

Berdasarkan nilai AIC dan  $R_{MF}^2$  pada Tabel 3.11, disimpulkan bahwa model GWLR lebih baik dari model regresi logistik global, karena nilai AIC model GWLR lebih kecil, yaitu 29,629, serta nilai koefisien determinasi yang lebih besar dari model global, yaitu 61,1%. Oleh karena itu, peluang pencemaran air sungai di Provinsi Kalimantan Timur tahun 2022 lebih baik dimodelkan menggunakan model GWLR.

Interpretasi model GWLR pada bahasan ini dibatasi pada satu model GWLR di setiap kelompok berdasarkan pada Tabel 3.10. Salah satu model GWLR yang diinterpretasikan untuk kelompok pertama adalah model GWLR di lokasi pengamatan Kantor Gubernur Samarinda dengan persamaan berdasarkan penaksir ML pada Tabel 3.9 adalah

$$\hat{\pi}(u_{14}, v_{14}) = \frac{\exp(16,438 + 0,015x_{14,1} + 21,110x_{14,2} + 0,022x_{14,3} - 3,262x_{14,4})}{1 + \exp(16,438 + 0,015x_{14,1} + 21,110x_{14,2} + 0,022x_{14,3} - 3,262x_{14,4})}$$

Berdasarkan hasil pengujian parsial yang diberikan oleh Tabel 3.9, diketahui bahwa faktor yang berpengaruh terhadap peluang pencemaran air sungai di lokasi pengamatan Kantor Gubernur Samarinda adalah konsentrasi nitrat ( $X_2$ ) dengan nilai *odds ratio*  $1,472 \times 10^9$ . Interpretasinya adalah setiap kenaikan satu *ppm* konsentrasi nitrat dan dianggap nilai variabel lainnya tetap akan meningkatkan peluang air sungai di Kantor Gubernur tercemar menjadi  $1,472 \times 10^9$  kali.

Salah satu model GWLR yang diinterpretasikan untuk kelompok kedua adalah model GWLR di lokasi pengamatan Kantor Bupati Berau, dengan persamaan model GWLR adalah

$$\hat{\pi}(u_{19}, v_{19}) = \frac{\exp(19,569 + 0,006x_{19,1} + 18,275x_{19,2} + 0,031x_{19,3} - 3,510x_{14,4})}{1 + \exp(19,569 + 0,006x_{19,1} + 18,275x_{19,2} + 0,031x_{19,3} - 3,510x_{14,4})}$$

Berdasarkan hasil pengujian parameter yang diberikan pada Tabel 3.9, faktor yang berpengaruh terhadap peluang pencemaran air sungai di lokasi pengamatan Kantor Bupati Berau adalah derajat warna air ( $X_3$ ) dengan nilai *odds ratio* 1,031. Interpretasinya adalah setiap kenaikan satu Pt-Co Unit warna air dan dianggap nilai variabel lainnya tetap akan meningkatkan peluang air sungai di Kantor Bupati Berau tercemar menjadi 1,031 kali atau meningkat 3,1%.

Salah satu model GWLR yang diinterpretasikan untuk kelompok ketiga adalah model GWLR di lokasi pengamatan Muara Sungai Belayan Kukar, dengan persamaan model GWLR adalah

$$\hat{\pi}(u_{29}, v_{29}) = \frac{\exp(14,730 + 0,009x_{29,1} + 15,183x_{29,2} + 0,023x_{29,3} - 2,813x_{29,4})}{1 + \exp(14,730 + 0,009x_{29,1} + 15,183x_{29,2} + 0,023x_{29,3} - 2,813x_{29,4})}$$

Berdasarkan hasil pengujian parsial yang diberikan oleh Tabel 3.9, faktor yang berpengaruh terhadap peluang pencemaran air sungai di lokasi pengamatan Muara Sungai Belayan Kukar adalah konsentrasi nitrat ( $X_2$ ) dan derajat warna air ( $X_3$ ). Nilai *odds ratio* variabel konsentrasi nitrat adalah  $3,924 \times 10^6$ , berarti setiap kenaikan satu *ppm* konsentrasi nitrat dan dianggap nilai variabel lainnya tetap akan meningkatkan peluang air sungai di Muara Sungai Belayan Kukar tercemar menjadi  $3,924 \times 10^6$  kali. Nilai *odds ratio* variabel derajat warna air adalah 1,023, artinya setiap kenaikan satu Pt-Co Unit warna air dan dianggap nilai variabel lainnya tetap akan meningkatkan peluang air sungai di Muara Sungai Belayan Kukar tercemar menjadi 1,023 kali, atau meningkat 2,3%.

Berdasarkan faktor-faktor yang berpengaruh terhadap peluang pencemaran air sungai di setiap lokasi pengamatan di rekomendasikan kepada masyarakat setempat agar menjaga kebersihan air sungai sehingga konsentrasi nitrat dan derajat warna air tidak meningkat. Rekomendasi juga ditujukan kepada pengelola pertambangan, dan pemerintah dalam melakukan operasional dan atau membuat kebijakan agar memperhatikan faktor konsentrasi nitrat dan derajat warna air sungai.

#### 4. KESIMPULAN DAN SARAN

1. Faktor-faktor yang berpengaruh terhadap peluang air sungai di Provinsi Kalimantan Timur tercemar bersifat lokal, faktor-faktor tersebut adalah konsentrasi nitrat dan derajat warna air.
2. Setiap kenaikan konsentrasi nitrat dan derajat warna air masing-masing akan meningkatkan peluang pencemaran air sungai di Provinsi Kalimantan Timur.
3. Berdasarkan faktor lokal yang berpengaruh terhadap peluang pencemaran air sungai di setiap lokasi pengamatan, direkomendasikan kepada masyarakat setempat agar menjaga kebersihan,

tidak membuang sampah di aliran sungai sehingga konsentrasi nitrat dan derajat warna air tidak meningkat. Rekomendasi ditujukan kepada pengelola pertambangan dalam melakukan operasionalnya, dan pemerintah dalam membuat kebijakan pencegahan pencemaran air sungai harus berorientasi pada program-program yang dapat mengurangi konsentrasi nitrat dan derajat warna air sungai, serta program yang melibatkan masyarakat setempat.

## DAFTAR PUSTAKA

- [1] Agresti, A., 2013. *Categorical Data Analysis Third Edition*. New Jersey: John Wiley & Sons, Inc.
- [2] Berliana, S. M., Purhadi, Sutikno & Rahayu, S. P., 2020. Parameter estimation and hypothesis testing of geographically weighted multivariate generalized poisson regression. *mathematics*. 8 (1523), pp. 1-14.  
<https://doi.org/10.3390/math8091523>
- [3] BPS, 2023. *Statistik Air Bersih Provinsi Kalimantan Timur 2022*. Kalimantan Timur: Badan Pusat Statistik.
- [4] Caraka, R. E. & Yasin, H., 2017. *Geographically Weighted Regression: Sebuah Pendekatan Regresi Geografis*. Yogyakarta: MOBIUS.
- [5] Fotheringham, A. S., Brundson, C. & Charlton, M. E., 2002. *Geographically Weighted Logistic Regression: The Analysis of Spatially Varying Relationship*. England: John Wiley & Sons.
- [6] Gujarati, D. A., 2003. *Basic Econometric Fourth Edition*. New York: McGraw-Hill Inc.
- [7] Haq, I., Aidi, M. N., Kurnia, A. & Efriwati, 2023. A Comparison of Logistics Regression and Geographically Weighted Logistics Regression (GWLR) on COVID-19 Data in West Sumatra. *Barekeng: Journal of Mathematics and Its Applications*. 17 (3), pp. 1749-1760.  
<https://doi.org/10.30598/barekengvol17iss3pp1749-1760>
- [8] Harlan, J., 2018. *Analisis Regresi Logistik*. Depok: Gunadarma
- [9] Hosmer, D. W. & Lemeshow, S., 2000. *Applied Logistic Regression, Second Edition*. New York: John Willey & Sons.
- [10] Inayah, U. R., Suyitno & Siringoringo, M., 2021. Upaya Pencegahan Pencemaran Air Sungai Mahakam melalui Pemodelan Geographically Weighted Logistic Regression pada Data BOD. *Jurnal Eksponensial*, 12 (1), pp. 17-26.  
<https://doi.org/10.30872/eksponensial.v12i1.755>
- [11] Kementerian Lingkungan Hidup dan Kehutanan, 2021. *Peraturan Menteri Lingkungan Hidup dan Kehutanan Republik Indonesia Nomor 5 Tahun 2021 Tentang Tata Cara Penerbitan Persetujuan Teknis dan Surat Kelayakan Operasional Bidang Pengendalian Pencemaran Lingkungan*. Jakarta: Menteri Lingkungan Hidup dan Kehutanan Republik Indonesia.
- [12] Lestari, V. D., Suyitno, S. & Siringoringo, M., 2021. Analisis Faktor-Faktor Yang Berpengaruh Terhadap Pencemaran Air Sungai Mahakam Menggunakan Pemodelan *Geographically Weighted Logistic Regression* Pada Data *Dissolved Oxygen*. *Jurnal Eksponensial*, 12 (1), pp. 37-46.  
<https://doi.org/10.30872/eksponensial.v12i1.757>
- [13] Pemerintah Republik Indonesia, 2022. *Lampiran VI, Peraturan Pemerintah Republik Indonesia Nomor 22 Tahun 2021, Tentang Penyelenggaraan Perlindungan dan Pengelolaan Lingkungan Hidup*. Jakarta: Pemerintah Republik Indonesia.  
[https://jdih.setkab.go.id/PUUdoc/176367/Lampiran\\_VI\\_Salinan\\_PP\\_Nomor\\_22\\_Tahun\\_2021.pdf](https://jdih.setkab.go.id/PUUdoc/176367/Lampiran_VI_Salinan_PP_Nomor_22_Tahun_2021.pdf)

- [14] Pemerintah Republik Indonesia, 2022. *Peraturan Pemerintah Republik Indonesia Nomor 22 Tahun 2021, Tentang Penyelenggaraan Perlindungan dan Pengelolaan Lingkungan Hidup*. Jakarta: Pemerintah Republik Indonesia.  
[https://jdih.setkab.go.id/PUUdoc/176367/PP\\_Nomor\\_22\\_Tahun\\_2021.pdf](https://jdih.setkab.go.id/PUUdoc/176367/PP_Nomor_22_Tahun_2021.pdf)
- [15] Perdana, H., Satyahadewi, N. & Arsy, F. M., 2023. Determining Student Graduation Based on School Location Using Geographically Weighted Logistics Regression. *Barekeng: Journal of Mathematics and Its Applications*. 17 (4), pp. 2273-2280.  
<https://doi.org/10.30598/barekengvol17iss4pp2273-2280>
- [16] Purwanti, S. I., Sutikno & Puhadi, 2021. Parameter and hypothesis testing of geographically and temporally weighted bivariate generalized Poisson Regression. *IOP Conference Series: Earth and Environmental Science*. 880 (2021) 012043, pp. 1-11.  
<https://iopscience.iop.org/article/10.1088/1755-1315/880/1/012043>
- [17] Rifada, M., Chamidah, N., Ningrum, R. A. & Muniroh, L., (2023) Stunting Determinants Among Toddlers in Probolinggo of Indonesia Using Parametric and Nonparametric Ordinal Logistic Regression Models. *Commun. Math. Biol. Neurosci.* 2023 (2023), 8, pp. 1-14.  
<http://www.scik.org/index.php/cmbn/article/view/6690>
- [18] Salsavira, S., Afifah, J., Mahendra, F. T. & Dzakiyah, L., 2021. Analisis Spasial Prevalensi Perkawinan Dini dan IPM di Indonesia. *Jurnal Matematika, Statistika Dan Komputasi*, 18 (1), pp. 31-41.  
<https://doi.org/10.20956/j.v18i1.13975>
- [19] Scabra, A. R., Afriadin, A. & Marzuki, M., 2022. Efektivitas Peningkatan Oksigen Terlarut Menggunakan Perangkat Microbubble Terhadap Produktivitas Ikan Nila. *Jurnal Perikanan Unram*, 12 (1), pp. 13–21. <https://jperairan.unram.ac.id/index.php/JP/article/view/269>
- [20] Sifriyani, S., Rasjid, M., Rosadi, D., Anwar, S., Wahyuni, R. D. & Jalaluddin, S., 2022. Spatial-Temporal Epidemiology of Covid-19 Using a Geographically and Temporally Weighted Regression Model. *Symmetry*, 14 (4), 742, pp. 1-8.  
<https://www.mdpi.com/2073-8994/14/4/742>
- [21] Suyitno, Puhadi, Sutikno, & Irhamah, 2017. Multivariate Weibull regression model. *Far East Journal of Mathematical Sciences*, 101 (9), pp. 1977-1992.  
<https://doi.org/10.17654/MS101091977>
- [22] Suyitno, Puhadi, Sutikno, and Irhamah, 2016. Parameter estimation of geographically weighted trivariate Weibull regression model,” *Applied Mathematical Sciences*, 10 (18), 861–878. <https://doi.org/10.12988/ams.2016.6129>.
- [23] Wardhani, Q. S., Handajani, S. S. & Susanto, I., 2022. Pemodelan Indeks Pembangunan Kesehatan Masyarakat dengan Metode Geographically Weighted Logistic Regression. *Jurnal Aplikasi Statistika & Komputasi Statistik*, 14 (1), 1-12.  
<https://doi.org/10.34123/jurnalasks.v14i2.333>