

## PEMODELAN TINGKAT KUALITAS AIR DI KOTA PONTIANAK DENGAN MENGGUNAKAN *MULTIVARIATE GEOGRAPHICALLY WEIGHTED REGRESSION*

### *Modeling Water Quality Levels In Pontianak City Using Multivariate Geographically Weighted Regression*

Dadan Kusnandar<sup>1</sup>, Naomi Nessyana Debatara<sup>2\*</sup>, Shindy Utari<sup>3</sup>

<sup>1,2,3</sup> Program Studi Statistika, FMIPA Universitas Tanjungpura  
Jln. Prof. Dr. H. Hadari Nawawi, Kota Pontianak, 78124, Indonesia

Corresponding author e-mail: <sup>2\*</sup> [naominessyana@math.untan.ac.id](mailto:naominessyana@math.untan.ac.id)

#### Abstrak

Ketersediaan air bersih dan sanitasi yang layak merupakan salah satu tujuan dalam *Sustainable Development Goals*. Kualitas air cenderung mengalami penurunan terutama di daerah permukiman akibat tercemar limbah dari hasil kegiatan manusia. Penyebab pencemaran air bisa jadi berbeda-beda di setiap lokasi pengamatan, sehingga faktor letak geografis perlu dipertimbangkan pada proses pengambilan keputusan. *Multivariate Geographically Weighted Regression* digunakan untuk mengatasi adanya pengaruh heterogenitas spasial dalam data yang disebabkan oleh perbedaan kondisi lokasi yang satu dengan lokasi lain. Tujuan dari penelitian ini adalah menentukan model dan faktor-faktor apa saja yang berpengaruh terhadap kualitas air di Kota Pontianak. Data yang digunakan pada penelitian ini adalah data kualitas air di Kota Pontianak sebanyak 42 titik sampel lokasi. Variabel responnya terdiri dari  $Y_1$  (COD) dan  $Y_2$  (TDS), sedangkan untuk variabel prediktor terdiri dari  $X_1$  (warna),  $X_2$  (pH),  $X_3$  (kandungan zat besi), dan  $X_4$  (kesadahan). Hasil penelitian menunjukkan bahwa variabel yang mempengaruhi COD adalah warna, sedangkan variabel TDS dipengaruhi oleh warna dan kesadahan.

**Kata Kunci :** *Weighted least square, cross validation, Akaike's information criterion, spasial, pencemaran air.*

#### Abstract

The availability of clean water and proper sanitation is one of the goals in the *Sustainable Development Goals*. Water quality tends to decreased, especially in residential areas due to contamination of waste from human activities. The causes of water pollution may vary in each observation location, so that the geographical location factor needs to be considered in the decision making process. *Multivariate Geographically Weighted Regression* is used to overcome the influence of spatial heterogeneity caused by differences in conditions from one location to another. The purpose of this study was to determine the models and factors that affect water quality in Pontianak City. The data used in this study were 42 water quality data in Pontianak City as sample points. The response variables consisted of  $Y_1$  (COD) and  $Y_2$  (TDS), while the predictor variables consisted of  $X_1$  (color),  $X_2$  (pH),  $X_3$  (iron content), and  $X_4$  (water hardness). The results showed that the variables affecting COD on water quality were color, while the TDS variables were affected by color and hardness.

**Keywords:** *Weighted least square, cross validation, Akaike's information criterion, spatial, water pollution.*

#### Article info:

Submitted: 30<sup>th</sup> March 2021

Accepted: 25<sup>th</sup> July 2021

#### How to cite this article:

D. Kusnandar, N. N. Debatara, and S. Utari, "ANALISIS REGRESI NON LINEAR PADA DATA PASIEN COVID-19 MENGGUNAKAN METODE BOOTSTRAP", *BAREKENG: J. Il. Mat. & Ter.*, vol. 15, no. 03, pp. 493-502, Sep. 2021.



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).  
Copyright © 2021 Dadan Kusnandar, Naomi Nessyana Debatara, Shindy Utari

## 1. PENDAHULUAN

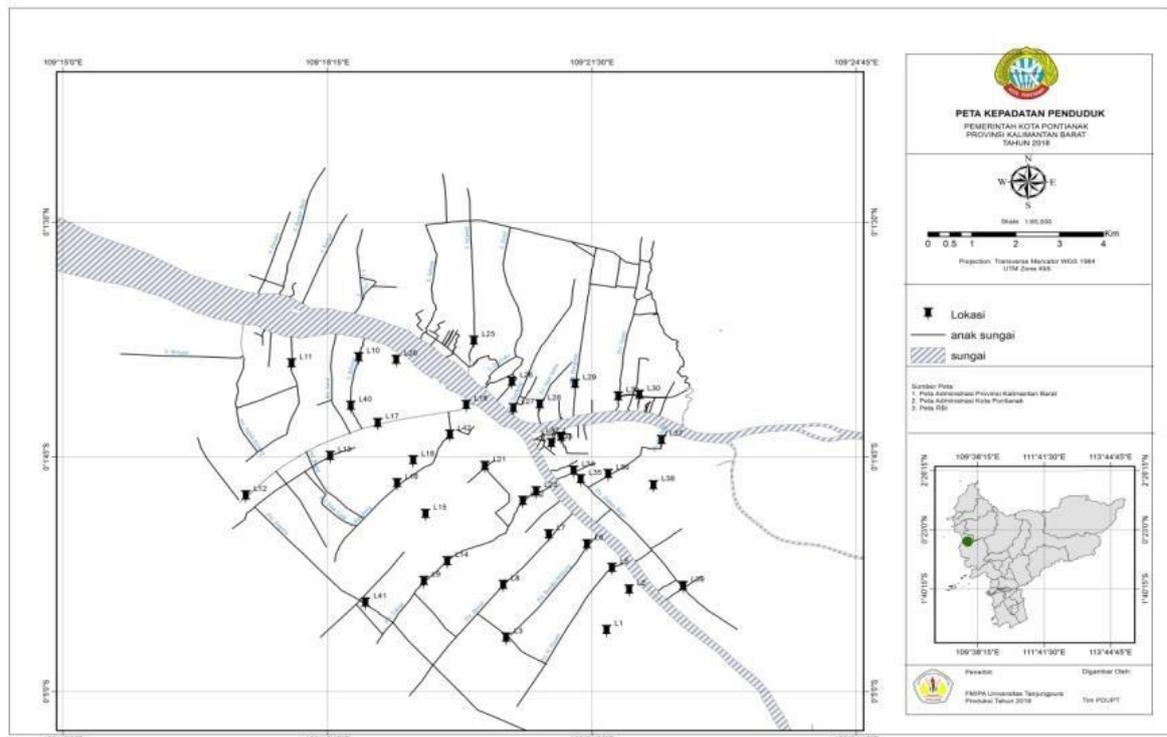
Air merupakan sumber daya alam yang penting bagi manusia. Keberadaannya dibutuhkan dalam hampir semua kegiatan manusia sehari-hari, baik untuk rumah tangga, pertanian atau industri. Akan tetapi, kegiatan tersebut umumnya menghasilkan limbah yang dapat menurunkan kualitas air dan menimbulkan polusi [1]. Beberapa jenis limbah bahkan berbahaya bagi kehidupan manusia. Analisis klasifikasi kualitas air di Kota Pontianak telah dilakukan dengan menggunakan teknik analisis multivariat [1]–[3]. Dalam penelitian ini digunakan perluasan analisis multivariat dengan mempertimbangkan model regresi spasial.

Analisis regresi adalah analisis yang menggunakan dua atau lebih variabel sehingga memungkinkan adanya ketergantungan antara variabel satu dengan variabel yang lain. Dalam analisis regresi terdapat dua jenis variabel yaitu variabel prediktor dan variabel respon. Untuk menghubungkan variabel respon dan prediktor dapat digunakan model regresi berbentuk univariat maupun multivariat [4]. Regresi spasial merupakan analisis yang mengevaluasi hubungan antara satu variabel dengan beberapa variabel lain dan memberikan efek spasial pada beberapa lokasi yang menjadi pusat pengamatan. Hal mendasar dalam pemodelan spasial yaitu adanya matriks pembobot spasial. Matriks ini digunakan untuk menggambarkan hubungan antar wilayah yang berdekatan [5]. Apabila unsur spasial digunakan untuk model regresi klasik akan menyebabkan kesimpulan yang kurang tepat karena asumsi homogenitas tidak terpenuhi. Oleh karena itu, dibutuhkan metode statistika untuk mengatasi penambahan unsur variabel spasial. Matriks pembobotan spasial dengan pendekatan titik sebagai letak geografis suatu wilayah berdasarkan posisi koordinat garis lintang dan garis bujur yang mencerminkan hubungan antara lokasi yang satu dengan yang lainnya [6].

Model *Multivariat Geographically Weighted Regression* merupakan model regresi multivariat spasial yang digunakan untuk mengatasi adanya pengaruh heterogenitas spasial yang disebabkan oleh perbedaan kondisi lokasi yang satu dengan lokasi lain [7]. Penaksir parameter pada model Multivariat GWR bersifat lokal untuk setiap pengamatan yang artinya mempunyai koefisien regresi yang berbeda-beda, dengan variabel respon lebih dari satu dan dengan menambahkan pembobotan lokasi [6]. Pembobotan pada Multivariat GWR merupakan kedekatan titik lokasi ke- $i$  dengan titik lokasi pengamatan lainnya. Fungsi tersebut dinyatakan dalam diagonal matriks dimana elemen-elemen diagonalnya merupakan sebuah fungsi dari setiap titik lokasi pengamatan. Pembobotan matriks juga dapat ditentukan dengan menggunakan fungsi *Kernel*. Fungsi pembobotan *Kernel* memberikan pembobotan yang sesuai dengan nilai *bandwidth* optimum tergantung pada kondisi data. Terdapat beberapa fungsi pembobotan yang dapat digunakan model Multivariat GWR yaitu fungsi *Kernel Fixed Gaussian*, fungsi *Kernel Bisquare Gaussian*, dan fungsi *Kernel Tricube Gaussian* [8]. Dalam penelitian ini, model Multivariat GWR diaplikasikan untuk menentukan sebaran dan faktor-faktor apa saja yang berpengaruh terhadap *Chemical Oxygen Demand* (COD) dan *Total Dissolved Solid* (TDS) pada kualitas air di Kota Pontianak.

## 2. METODE PENELITIAN

Dalam penelitian ini, pemodelan Multivariat GWR dilakukan pada kualitas air di Kota Pontianak. Penelitian ini menggunakan data kualitas air dari 42 titik lokasi yang berbeda yang disajikan pada Gambar 1. Pengambilan sampel dilakukan dengan metode *stratified random sampling* [9]. Variabel yang digunakan dalam penelitian ini terdiri dari dua variabel respon dan empat variabel prediktor. Variabel respon yang digunakan yaitu COD ( $Y_1$ ) dan TDS ( $Y_2$ ), sedangkan untuk variabel prediktor yang digunakan yaitu warna ( $X_1$ ), pH ( $X_2$ ), kandungan zat besi ( $X_3$ ), dan kesadahan ( $X_4$ ).



Gambar 1. Peta Titik Lokasi Sampel

Sumber: [10]

## 2.1 Regresi Multivariat

Tahapan pertama dalam penelitian ini adalah melakukan uji asumsi klasik pada model dengan menggunakan metode regresi multivariat. Model tersebut dapat dituliskan sebagai berikut [11]:

$$\mathbf{Y}_{(n \times q)} = \mathbf{X}_{(n \times (p+1))} \boldsymbol{\beta}_{((p+1) \times q)} + \boldsymbol{\varepsilon}_{(n \times q)} \quad (1)$$

dimana  $\mathbf{Y}$  merupakan matriks data pada variabel respon,  $\mathbf{X}$  merupakan matriks data pada variabel prediktor,  $\boldsymbol{\beta}$  adalah matriks parameter regresi,  $\boldsymbol{\varepsilon}$  merupakan matriks residual,  $n$  merupakan jumlah pengamatan,  $q$  merupakan banyaknya variabel respon, dan  $p$  merupakan banyaknya variabel prediktor. Asumsi yang digunakan dalam analisis regresi linear multivariat yaitu:

### a. Multikolinearitas

Multikolinearitas terjadi karena adanya korelasi yang cukup tinggi antar variabel prediktor. Hal ini dapat terdeteksi salah satunya dengan menentukan nilai *Variance Inflation Factor (VIF)*. Nilai VIF dapat dihitung dengan menggunakan rumus [12]:

$$VIF = \frac{1}{1 - R_j^2} \quad (2)$$

dimana  $R_j^2$  merupakan koefisien determinasi yang dihasilkan dari regresi variabel prediktor  $X_j$  dengan variabel prediktor lainnya. Jika nilai  $VIF < 10$  maka tidak terdapat multikolinearitas.

### b. Residual Berdistribusi Normal

Uji normalitas merupakan salah satu syarat dalam model regresi linear yang mengharuskan terpenuhinya asumsi bahwa eror berdistribusi normal. Uji yang digunakan dalam menguji normalitas adalah uji *Kolmogorov-Smirnov*. Pemeriksaan residual berdistribusi normal dapat dihitung dengan menggunakan rumus [13]:

$$KS_{hitung} = \max |F(z_i) - S(z_i)| \quad (3)$$

dengan,

$$z_i = \frac{x_i - \bar{x}}{s}$$

$$F(z_i) = P(z \leq z_i)$$

$$S(z_i) = \frac{z_1, z_2, \dots, z_n \leq z_i}{n}$$

dimana  $\bar{x}$  merupakan rata-rata sampel dan  $s$  merupakan simpangan baku sampel. Dikatakan berdistribusi normal jika  $p\text{-value} < \alpha$ .

### c. Heteroskedastisitas

Pengujian heteroskedastisitas dilakukan sebagai syarat salah satu uji asumsi eror untuk membentuk model OLS. Salah satu cara digunakan dalam pengujian heteroskedastisitas adalah menggunakan uji *Bartlett's Sphericity* yang dapat mendeteksi heterogenitas spasial dengan rumus [14]:

$$\chi^2_{hitung} = -\left(N - 1 - \frac{2q + 5}{6}\right) \ln |\mathbf{R}| \quad (4)$$

dimana  $\mathbf{R}$  merupakan matriks korelasi dari masing-masing variabel respon. Variabel respon berkorelasi jika  $\chi^2_{hitung} > \chi^2_{\alpha; \frac{1}{2}q(q-1)}$ .

## 2.2 Multivariat GWR

Multivariat GWR merupakan pengembangan dari model linear multivariat spasial dengan penaksir parameter bersifat lokal untuk setiap lokasi pengamatan. Apabila terdapat variabel respon  $Y_1, Y_2, \dots, Y_q$  dan variabel prediktor  $X_1, X_2, \dots, X_p$  pada lokasi ke- $i$ , maka model Multivariat GWR sebagai berikut [15]:

$$Y_{hi} = \beta_{0h}(u_i, v_i) + \beta_{1h}(u_i, v_i)X_{1i} + \dots + \beta_{kh}(u_i, v_i)X_{ki} + \varepsilon_{hi} \quad (5)$$

dimana  $Y_{hi}$  merupakan variabel respon ke- $h$  pada lokasi ke- $i$  dengan  $h = 1, 2, \dots, q$ ,  $X_{ki}$  merupakan variabel prediktor ke- $k$  pada lokasi ke- $i$  dengan  $k = 1, 2, \dots, p$ ,  $(u_i, v_i)$  merupakan koordinat titik lokasi pengamatan ke- $i$  (*longitude, latitude*),  $\beta_{0h}(u_i, v_i)$  merupakan konstanta/*intercept* dari Multivariat GWR pada variabel respon ke- $h$  dan lokasi ke- $i$ ,  $\beta_{kh}$  merupakan koefisien regresi pada variabel respon ke- $h$ , variabel prediktor ke- $k$ , dan lokasi ke- $i$ , dan  $\varepsilon_{hi}$  merupakan eror pada variabel respon ke- $h$  dan lokasi ke- $i$ .

Penaksiran koefisien regresi pada metode Multivariat GWR dilakukan dengan meminimumkan kuadrat eror pada pembobotan yang diberikan berbeda setiap lokasi pengamatan, sehingga koefisien regresi pada lokasi pengamatan  $(u_i, v_i)$  ditaksir dengan penambahan bobot  $w_{ij}$  dengan fungsi *Kernel* sebagai berikut [8]:

### 1. Fungsi Gaussian

$$w_{ij}(u_i, v_i) = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right) \quad (6)$$

### 2. Fungsi Bisquare

$$w_{ij}(u_i, v_i) = \begin{cases} \left(1 - \left(\frac{d_{ij}}{b}\right)^2\right)^2 & , \text{untuk } d_{ij} \leq b \\ 0 & , \text{untuk } d_{ij} > b \end{cases} \quad (7)$$

### 3. Fungsi Tricube

$$w_{ij}(u_i, v_i) = \begin{cases} \left(1 - \left(\frac{d_{ij}}{b}\right)^3\right)^3 & , \text{untuk } d_{ij} \leq b \\ 0 & , \text{untuk } d_{ij} > b \end{cases} \quad (8)$$

dengan,

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

dimana  $b$  merupakan *bandwidth*,  $d_{ij}$  merupakan jarak *Euclidean* antar lokasi  $i$  dan lokasi  $j$ .

Estimasi parameter yang digunakan dalam model Multivariat GWR adalah metode *weighted least square* (WLS) atau metode kuadrat terkecil tertimbang dengan memberikan pembobot yang berbeda untuk setiap lokasi yang diamati. Estimasi tersebut adalah sebagai berikut [15]:

$$\hat{\beta}(u_i, v_i) = (\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{Y} \quad (9)$$

Selanjutnya dilakukan pengujian signifikansi parameter Multivariat GWR dengan uji serentak dan uji parsial. Perumusan hipotesis yang digunakan pada uji serentak sebagai berikut [15]:

$$H_0 : \beta_{1h}(u_i, v_i) = \beta_{2h}(u_i, v_i) = \dots = \beta_{ph}(u_i, v_i) = 0$$

$$H_1 : \text{minimal ada satu } \beta_{kh}(u_i, v_i) \neq 0$$

Statistik uji:

$$F_{hitung} = \frac{\left| \mathbf{Y}^T (\mathbf{I} - \mathbf{S})^T (\mathbf{I} - \mathbf{S}) \mathbf{Y} \right| / \left( \frac{\delta_1^2}{\delta_2} \right)}{\left| \mathbf{Y}^T (\mathbf{I} - \mathbf{S}_\omega)^T (\mathbf{I} - \mathbf{S}_\omega) \mathbf{Y} \right| / \left( \frac{\delta_{1\omega}^2}{\delta_{2\omega}} \right)} \quad (10)$$

Dengan:

$$\delta_1 = tr((\mathbf{I} - \mathbf{S})^T (\mathbf{I} - \mathbf{S}))$$

$$\delta_2 = tr((\mathbf{I} - \mathbf{S})^T (\mathbf{I} - \mathbf{S}))^2$$

$$\delta_{1\omega} = tr((\mathbf{I} - \mathbf{S}_\omega)^T (\mathbf{I} - \mathbf{S}_\omega))$$

$$\delta_{2\omega} = tr((\mathbf{I} - \mathbf{S}_\omega)^T (\mathbf{I} - \mathbf{S}_\omega))^2$$

$$\mathbf{S} = \begin{bmatrix} \mathbf{X}_1^T (\mathbf{X}_1^T \mathbf{W}(u_1, v_1) \mathbf{X}_1)^{-1} \mathbf{X}_1^T \mathbf{W}(u_1, v_1) \\ \mathbf{X}_2^T (\mathbf{X}_2^T \mathbf{W}(u_2, v_2) \mathbf{X}_2)^{-1} \mathbf{X}_2^T \mathbf{W}(u_2, v_2) \\ \vdots \\ \mathbf{X}_n^T (\mathbf{X}_n^T \mathbf{W}(u_n, v_n) \mathbf{X}_n)^{-1} \mathbf{X}_n^T \mathbf{W}(u_n, v_n) \end{bmatrix}$$

$$\mathbf{S}_\omega = \begin{bmatrix} \mathbf{1}^T (\mathbf{1}^T \mathbf{W}(u_i, v_i) \mathbf{1})^{-1} \mathbf{1}^T \mathbf{W}(u_i, v_i) \\ \mathbf{1}^T (\mathbf{1}^T \mathbf{W}(u_2, v_2) \mathbf{1})^{-1} \mathbf{1}^T \mathbf{W}(u_2, v_2) \\ \vdots \\ \mathbf{1}^T (\mathbf{1}^T \mathbf{W}(u_n, v_n) \mathbf{1})^{-1} \mathbf{1}^T \mathbf{W}(u_n, v_n) \end{bmatrix}$$

dimana  $\mathbf{1}$  merupakan vektor satu berukuran  $n \times 1$  yang merupakan baris ke-1 dari matriks  $\mathbf{X}$ . Kriteria pengujian tolak  $H_0$  jika  $F_{hitung} > F \left( \alpha, \left( \frac{\delta_1^2}{\delta_2} \right), \left( \frac{\delta_{1\omega}^2}{\delta_{2\omega}} \right) \right)$ .

Sedangkan perumusan hipotesis yang digunakan pada uji parsial sebagai berikut [15]:

$$H_0 : \beta_{kh}(u_i, v_i) = 0$$

$$H_1 : \beta_{kh}(u_i, v_i) \neq 0$$

Statistik uji:

$$t_{hitung} = \frac{\hat{\beta}_{kh}(u_i, v_i)}{SE(\hat{\beta}_{kh}(u_i, v_i))} \quad (11)$$

Kriteria pengujian tolak  $H_0$  jika  $|t_{hitung}| \geq t_{\left(\frac{\alpha}{2}, \left(n - \frac{\delta_1^2}{\delta_2}\right)\right)}$ .

### 2.3 Pemilihan Model Terbaik

Pemilihan model terbaik dapat dilakukan dengan menghitung nilai AIC (*Akaike's Information Criterion*), dengan rumus [8]:

$$AIC = 2n \ln |\hat{\Sigma}| + n \ln(2\pi) + n \left\{ \frac{n + \text{tr}(\mathbf{S})}{n - 2 - \text{tr}(\mathbf{S})} \right\} \quad (12)$$

dengan  $\hat{\Sigma}$  merupakan penaksir matriks varian kovarian dari vektor error model Multivariat GWR. Model terbaik adalah model yang memiliki nilai AIC terkecil.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Pemodelan Regresi Multivariat

Analisis model regresi multivariat ini digunakan untuk menentukan variabel prediktor mana saja yang berpengaruh secara signifikan terhadap variabel respon tanpa melibatkan faktor lokasi pengamatan. Pada pengujian dengan menggunakan metode Multivariat GWR, langkah awal yang dilakukan yaitu menguji dengan menggunakan model regresi multivariat.

**Tabel 1. Uji Serentak Model Regresi Multivariat**

| Y1 (COD) |           |    |           |              |                |
|----------|-----------|----|-----------|--------------|----------------|
| SK       | JK        | db | KT        | $F_{hitung}$ | <i>P-value</i> |
| Model    | 30346,160 | 4  | 7586,540  | 31,762       | 0,000          |
| Error    | 8837,712  | 37 | 238,857   |              |                |
| Total    | 39183,872 | 41 |           |              |                |
| Y2 (TDS) |           |    |           |              |                |
| SK       | JK        | db | KT        | $F_{hitung}$ | <i>P-value</i> |
| Model    | 47313,167 | 4  | 11828,292 | 38,267       | 0,000          |
| Error    | 11436,650 | 37 | 309,099   |              |                |
| Total    | 58749,816 | 41 |           |              |                |

Tabel 1 menunjukkan bahwa nilai statistik uji F ( $F_{hitung}$ ) untuk COD ( $Y_1$ ) sebesar 31,762 dengan *p-value* sebesar 0,000 dan nilai statistik uji F untuk TDS ( $Y_2$ ) sebesar 38,267 dan *p-value* sebesar 0,000. Hasil statistik uji  $F_{hitung} > F_{tabel}(2,63)$  dengan menggunakan tingkat signifikansi  $\alpha$  (0,05) dapat disimpulkan variabel prediktor secara serentak berpengaruh signifikan terhadap COD dan TDS. Setelah didapatkan hasil uji serentak dari analisis regresi multivariat, selanjutnya dilakukan pengujian parsial. Tabel 2 menyajikan nilai parameter hasil taksiran untuk model Regresi Multivariat.

**Tabel 2. Penaksir Parameter Model Regresi Multivariat**

| Variabel respon | Parameter       | Penaksir ( $\beta$ ) | $t_{hitung}$ | <i>P-value</i> |
|-----------------|-----------------|----------------------|--------------|----------------|
| COD ( $Y_1$ )   | Intercept       | 49,893               | 1,980        | 0,055          |
|                 | Warna ( $X_1$ ) | 0,082                | 7,974        | 0,000          |
|                 | pH ( $X_2$ )    | 2,654                | 0,740        | 0,464          |
|                 | Besi ( $X_3$ )  | -4,687               | -2,491       | 0,017          |

| Variabel respon | Parameter           | Penaksir ( $\beta$ ) | $t_{hitung}$ | $P$ -value |
|-----------------|---------------------|----------------------|--------------|------------|
| TDS ( $Y_2$ )   | Kesadahan ( $X_4$ ) | -0,214               | -0,916       | 0,366      |
|                 | Intercept           | 17,091               | 0,596        | 0,555      |
|                 | Warna ( $X_1$ )     | -0,063               | -5,383       | 0,000      |
|                 | pH ( $X_2$ )        | 7,166                | 1,756        | 0,087      |
|                 | Besi ( $X_3$ )      | -1,022               | -0,478       | 0,636      |
|                 | Kesadahan ( $X_4$ ) | 0,841                | 3,169        | 0,003      |

Berdasarkan hasil dari Tabel 2 pada variabel COD ( $Y_1$ ) dengan tingkat signifikansi ( $\alpha$ ) sebesar 5% didapatkan  $p$ -value parameter  $\beta_1$  dan  $\beta_3$  lebih kecil dari  $\alpha$  (0,05). Hal ini menunjukkan bahwa variabel besi dan warna berpengaruh signifikan terhadap peningkatan COD pada kualitas permukaan air, dari hasil penaksir parameter tersebut dapat dibuat model regresi multivariat dari COD adalah sebagai berikut:

$$\hat{Y}_1 = 49,893 + 0,082 X_1 - 4,687 X_3 \quad (13)$$

Persamaan (13) dapat diinterpretasikan bahwa setiap kenaikan warna sebesar satu satuan, maka COD akan meningkat sebesar 0,082 mg/l dengan asumsi besi yang terjadi tetap dan setiap kenaikan besi sebesar satu satuan maka COD akan menurun sebesar 4,687 mg/l dengan asumsi bahwa warna tetap.

Pada variabel TDS ( $Y_2$ ) dengan tingkat signifikansi  $\alpha$  (0,05) didapatkan parameter  $\beta_1$  dan  $\beta_4$  yang signifikan, sehingga dapat dinyatakan bahwa variabel  $X_1$  (warna) dan  $X_4$  (kesadahan) yang berpengaruh signifikan terhadap TDS pada kualitas permukaan air. Dari hasil penaksir parameter tersebut dapat dibuat model regresi multivariat untuk TDS adalah:

$$\hat{Y}_2 = 17,091 - 0,063 X_1 + 0,841 X_4 \quad (14)$$

Dengan interpretasi bahwa setiap kenaikan warna sebesar satu satuan maka TDS akan menurun sebesar 0,063 mg/l dengan asumsi bahwa kesadahan tetap dan setiap kenaikan kesadahan sebesar satu satuan maka TDS akan meningkat sebesar 0,841 mg/l dengan asumsi bahwa warna tetap. Selanjutnya dilakukan uji asumsi residual pada model regresi multivariat yaitu asumsi multikolinearitas, asumsi distribusi normal, dan asumsi heteroskedastisitas.

#### a. Uji Multikolinearitas

Uji asumsi ini bertujuan untuk menentukan apakah variabel-variabel prediktornya telah memenuhi kondisi tidak saling berkorelasi. Kriteria yang digunakan untuk memeriksa multikolinearitas antar variabel prediktor adalah dengan menggunakan nilai *Variance Inflation Factors* (VIF) pada variabel-variabel prediktor.

**Tabel 3. Nilai VIF dari Variabel Prediktor**

| Variabel            | VIF   |
|---------------------|-------|
| Warna ( $X_1$ )     | 1,981 |
| pH ( $X_2$ )        | 2,441 |
| Besi ( $X_3$ )      | 1,102 |
| Kesadahan ( $X_4$ ) | 1,668 |

Hasil uji multikolinearitas didapatkan hasil VIF bahwa tidak ada multikolinearitas antar variabel prediktor karena nilai VIF dari masing-masing variabel prediktor menunjukkan nilai kurang dari 10 (nilai VIF minimum mendekati satu).

### b. Uji Asumsi Normal

Berdasarkan hasil pengujian menunjukkan bahwa nilai  $p$ -value pada variabel  $Y_1$  sebesar 0,597 dan nilai  $p$ -value pada variabel  $Y_2$  sebesar 0,765 lebih besar dari  $\alpha$  (0,05) sehingga dapat disimpulkan bahwa residual model regresi yang terbentuk mengikuti distribusi normal.

### c. Uji Asumsi Heteroskedastisitas

Tabel 4. Uji Bartlett's

| Statistik Uji                     | Nilai |
|-----------------------------------|-------|
| Chi Kuadrat ( <i>Chi-Square</i> ) | 2,206 |
| db                                | 2     |
| Sig.                              | 0,332 |

Hasil uji Bartlett's menunjukkan bahwa nilai  $\chi^2_{hitung} = 2,206$  lebih kecil dari nilai  $\chi^2_{2;0,05} = 5,991$  dan  $p$ -value (0,332)  $>$   $\alpha$  (0,05) yang berarti bahwa varian kovarian dari vektor error sama atau dengan kata lain asumsi heteroskedastisitas terpenuhi.

## 3.2 Multivariat GWR

Langkah pertama dalam analisis data ini adalah mencari koordinat lokasi pengamatan dalam hal ini adalah letak geografis 42 lokasi. Setelah koordinat lokasi pengamatan didapatkan, selanjutnya digunakan untuk mencari Jarak *Euclidean* antara lokasi. Kemudian nilai *bandwidth* optimum berdasarkan koordinat lokasi pengamatan dicari dengan menggunakan Fungsi Kernel dengan pembobot *Tricube*. Dari hasil uji didapatkan nilai *bandwidth* optimum sebesar 20430,03. Hasil dari nilai *bandwidth* optimum tersebut digunakan untuk mendapatkan matriks pembobot  $\mathbf{W}(u_i, v_i)$ . Hasil pembobot ini juga digunakan untuk mencari penaksir parameter beta dari COD dan TDS model Multivariat GWR secara serentak untuk setiap lokasi pengamatan. Berdasarkan hasil perhitungan dengan program R menggunakan Pembobot Kernel *Tricube* didapatkan penaksir parameter di setiap lokasi pengamatan  $(u_i, v_i)$ ,  $i = 1, 2, \dots, 42$ .

Selanjutnya dilakukan uji signifikansi model yang terdiri dari uji serentak dan uji parsial. Pengujian pengaruh variabel global secara serentak dari model Multivariat GWR menunjukkan bahwa nilai statistik uji F secara serentak dari model adalah 2,191 yang lebih kecil dari  $F_{tabel}$  yaitu 1,912. Hal ini dapat disimpulkan bahwa variabel prediktor global berpengaruh secara serentak.

Untuk menentukan variabel prediktor global mana yang berpengaruh signifikan terhadap variabel respon pada tiap lokasi, maka dilakukan pengujian parameter global secara parsial. Berdasarkan perhitungan didapat bahwa pemodelan Multivariat GWR menggunakan Fungsi *Kernel Tricube* menghasilkan satu model sebaran COD dan 2 model sebaran TDS pada kualitas permukaan air di Kota Pontianak.

## 3.3 Pemilihan Model Terbaik

Kriteria pemilihan model terbaik yang digunakan adalah dengan membandingkan nilai AIC dari kedua model tersebut. Model terbaik adalah model dengan nilai AIC yang terkecil. Hasil yang diperoleh dapat dilihat pada Tabel 6.

Tabel 6. Nilai AIC Model

| Statistik | Variabel    | Model               |                 |
|-----------|-------------|---------------------|-----------------|
|           |             | Regresi Multivariat | Multivariat GWR |
| AIC       | COD dan TDS | 1071,436            | 1066,631        |

Berdasarkan Tabel 6 diperoleh bahwa dengan menggunakan fungsi pembobot *Kernel Tricube* didapat model Multivariat GWR mempunyai nilai AIC yang lebih kecil dibandingkan dengan model Regresi Multivariat. Maka dari itu, model Multivariat GWR lebih baik digunakan untuk sebaran COD dan TDS di Kota Pontianak.

#### 4. KESIMPULAN

Dari hasil analisis diperoleh model terbaik untuk tingkat kualitas air COD dan TDS di Kota Pontianak adalah model Multivariat GWR dengan nilai AIC 1066,631. Dengan menggunakan model tersebut diperoleh faktor-faktor yang mempengaruhi kasus tingkat kualitas air di Kota Pontianak, untuk variabel COD adalah warna yang terdapat pada semua 42 titik lokasi pengamatan, sedangkan untuk variabel TDS adalah warna yang terdapat pada semua 42 titik lokasi pengamatan dan kesadahan yang terdapat pada 38 titik lokasi pengamatan.

#### DAFTAR PUSTAKA

- [1] D. Kusnandar, N. N. Debataraaja, S. W. Rizki, and E. Saputri, "Water Quality Mapping in Pontianak City Using Multiple Discriminant Analysis," in *The 4th IndoMS International Conference on Mathematics and Its Applications (IICMA, 2019)*, Pontianak, 2020, vol. 2268, pp. 020006-1-020006-6. doi: 10.1063/5.0016809.
- [2] D. Kusnandar, N. N. Debataraaja, and P. R. Dewi, "Classification of Water Quality in Pontianak City Using Multivariate Statistical Techniques," *Appl. Math. Sci.*, vol. 13, no. 22, pp. 1069–1075, 2019.
- [3] N. N. Debataraaja, D. Kusnandar, and R. W. Nusantara, "Exploratory Analysis of Water Quality in The Settlement Area of Pontianak City," in *The 4th IndoMS International Conference on Mathematics and Its Applications (IICMA, 2019)*, Pontianak, 2020, vol. 2268. doi: 10.1063/5.0016807.
- [4] A. C. Rencher, *Methods of Multivariate Analysis*, 2nd ed. New York: John Wiley & Sons, 2003.
- [5] N. Oktaviani, "Analisis Regresi Spasial Pada Data Jumlah Penduduk Miskin Provinsi Lampung Tahun 2017," Universitas Lampung, Bandar Lampung, 2018.
- [6] P. Andini, N. Herrhyanto, and M. Suherman, "Aplikasi Multivariate Geographically Weighted Regression Menggunakan Software Matlab," *J. EurekaMatematika*, vol. 5, no. 1, pp. 97–105, 2017.
- [7] Y. N. Dianati, N. W. S. Wardhani, and R. Fitriani, "Pemodelan Kerawanan Pangan dan Kemiskinan dengan Geographically Weighted Multivariate Linear Model di Kabupaten Sampang," *Nat. B*, vol. 2, no. 3, pp. 235–241, 2014.
- [8] A. S. Fotheringham, C. Brunson, and M. Charlton, *Geographically Weighted Regression*. West Sussex, England: John Wiley & Sons, 2002.
- [9] N. N. Debataraaja, D. Kusnandar, and R. W. Nusantara, "Identifikasi Lokasi Sebaran Pencemaran Air di Kawasan Permukiman Kota Pontianak," *J. Mat. Stat. Dan Komputasi*, vol. 15, no. 1, pp. 37–41, 2018.
- [10] M. Fikri, N. N. Debataraaja, and D. Kusnandar, "Penentuan Sebaran Spasial Pencemaran Air di Kota Pontianak Menggunakan Analisis Diskriminan Dua Kelompok," *Media Stat.*, vol. 12, no. 2, pp. 226–235, 2019.
- [11] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 6th ed. Upper Saddle River: Pearson Education, Inc., 2007.
- [12] Rahmadeni, and Anggreni, D., "Analisis Jumlah Tenaga Kerja Terhadap Jumlah Pasien RSUD Arifin Achmad Pekanbaru Menggunakan Metode Regresi Gulud," *J. Sains Teknol. Dan Ind.*, vol. 12, no. 1, pp. 48–57, 2014.
- [13] F. J. Massey, "The Kolmogorov-Smirnov Test for Goodness of Fit," *J. Am. Stat. Assoc.*, vol. 46, no. 253, pp. 68–78, 1951.
- [14] K. Daely, U. Sinulingga, and A. Manurung, "Analisis Statistik Faktor-faktor yang Mempengaruhi Indeks Prestasi Mahasiswa," *Saintia Mat.*, vol. 1, no. 5, pp. 483–494, Sep. 2013.
- [15] D. Hanum and Purhadi, "Faktor-faktor yang Mempengaruhi Morbiditas Penduduk Jawa Timur dengan Multivariate Geographically Weighted Regression (MGWR)," *J. Sains Dan Seni Pomits*, vol. 2, no. 2, pp. 184–194, 2013.

