

## **ROBUST GEOGRAPHICALLY WEIGHTED REGRESSION DENGAN METODE MUTLAK SIMPANGAN TERKECIL PADA PEMODELAN KEJADIAN DIARE DI KOTA SEMARANG**

**Ika Chandra Nurhayati<sup>1</sup>, Agus Rusgiyono<sup>2</sup>, Hasbi Yasin<sup>3</sup>**

<sup>1,2,3</sup> Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro

e-mail : [ikachandra07@gmail.com](mailto:ikachandra07@gmail.com)

### **ABSTRACT**

Diarrhea is one of many health issues in developing country like Indonesia, because the sickness and the death number are still high. According to health profile of Semarang City, the people who suffer from diarrhea from 2010-2015 are decreasing. The lowest point happened at the year 2013 with the total case of 38.001, however there are an increasing number from 2014-2015. The distribution data of diarrhea is a spatial data. The differences between environment and sanitation could cause spatial heterogeneity. The spatial heterogeneity could cause the produced variant value no longer constant, but instead it is different on each region. Therefore, regression model that involves the effects of spatial heterogeneity is needed, which are Geographically Weighted Regression (GWR) that is built by Weighted Least Square (WLS) adjuster. Although, GWR parameter adjuster that used WLS is very sensitive with the existence of outliers. The existence of the outlier in the data will create a huge residual. Thus, more robust method is needed, which is Least Absolute Deviation (LAD) methods in order to estimate the parameter on model GWR. This model is called Robust GWR (RGWR). The result shows that the model events of diarrhea on each region in Semarang City are different. Furthermore, the model events of diarrhea with RGWR model generate MAPE 16,3396% which means the performance of RGWR is formed well.

**Keyword:** Diarrhea, *Robust, Geographically Weighted Regression, Least Absolute Deviation*

### **1. PENDAHULUAN**

Penyakit diare merupakan salah satu masalah kesehatan masyarakat di negara berkembang seperti Indonesia karena angka kesakitan dan kematiannya masih tinggi. Penyakit ini masih menjadi endemis dan juga merupakan penyakit potensial Kejadian Luar Biasa (KLB) yang disertai dengan kematian Berdasarkan profil kesehatan Kota Semarang, penderita diare dari tahun 2010 – 2015 cenderung mengalami penurunan. Kejadian terendah terjadi pada tahun 2013 dengan total kasus sebanyak 38.001, namun pada tahun 2014 – 2015 mengalami peningkatan (Kementerian Kesehatan RI, 2015).

Diare merupakan penyakit menular, sehingga kejadian diare di suatu daerah dapat memengaruhi kejadian diare di sekitar daerah tersebut (Hariyanti, 2010). Salah satu metode statistika yang dapat dilakukan untuk mengetahui faktor-faktor yang mempengaruhi kejadian diare adalah analisis regresi. Faktor-faktor yang memengaruhi kejadian diare di wilayah tertentu bergantung pada kondisi wilayah persekitarnya, yang disebut pengaruh spasial (Bhekti, 2014). Pengaruh spasial yang berkaitan dengan perbedaan karakteristik lingkungan dan geografis antar wilayah pengamatan adalah keragaman spasial atau heterogenitas spasial. Oleh sebab itu, diperlukan model regresi yang melibatkan pengaruh heterogenitas spasial ke dalam, yaitu model *Geographically Weighted Regression* (GWR).

Model GWR dibangun dengan penaksir *Weighted Least Square* (WLS). Namun, penaksiran parameter GWR menggunakan *least square* sangat sensitif terhadap keberadaan pencilan. Oleh sebab itu diperlukan metode yang lebih *robust* guna mengakomodasi adanya pencilan pada model GWR. Salah satu metode yang dapat digunakan adalah dengan menggunakan penaksiran *Least Absolute Deviation* (LAD). LAD merupakan metode estimasi yang meminimalkan jumlah nilai mutlak dari kesalahan. Dengan menggunakan LAD, efek yang ditimbulkan oleh pencilan dapat diatasi secara otomatis

tanpa harus mendeteksi data mana yang menjadi pencilan atau tanpa harus membobot ulang residualnya (Wang dan Scott, 1994). Berdasarkan uraian di atas, maka pada penelitian ini akan dilakukan pengkajian tentang model GWR dengan menggunakan metode LAD dalam penaksiran parameternya pada kasus kejadian diare di 16 kecamatan di Kota Semarang pada tahun 2015.

## 2. TINJAUAN PUSTAKA

### 2.1. Diare dan Faktor yang Mempengaruhi

Diare atau penyakit diare (*Diarrheal Disease*) berasal dari bahasa Yunani, yaitu *Diarroi* yang artinya mengalir terus (Ariani, 2016). Diare adalah penyakit yang ditandai dengan meningkatnya frekuensi buang air besar lebih dari 3 kali sehari disertai perubahan konsistensi tinja menjadi lebih cair dengan atau tanpa lendir atau darah. Banyak faktor yang mempengaruhi terjadinya diare. Cara penularan diare pada umumnya melalui cara *fekal-oral*, yaitu melalui makanan atau minuman yang tercemar oleh enteropatogen atau kontak langsung dengan tangan penderita atau barang-barang yang telah tercemar tinja penderita atau tidak langsung melalui lalat.

### 2.2. Regresi Linier

Menurut Gujarati (2006), regresi merupakan suatu metode untuk mengukur besarnya pengaruh variabel bebas terhadap variabel terikat dan memprediksi variabel terikat dengan menggunakan variabel bebas. Secara matematis, model regresi linier dapat dinyatakan sebagai berikut :

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

dengan

$y_i$  = Variabel respon pada observasi ke -  $i$

$\beta_0, \beta_k$  = Intesep dan koefisien regresi dari variabel penjelas ke- $k$

$x_{ik}$  = Variabel penjelas ke- $k$  pada observasi ke- $i$

$\varepsilon_i$  = Komponen sisaan pada observasi ke- $i$  yang diasumsikan berdistribusi Normal ( $\varepsilon \sim N(0, \sigma^2)$ )

Apabila dinyatakan dalam bentuk persamaan matriks, akan diperoleh rumus

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

dimana:

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}; \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}; \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}; \mathbf{X} = \begin{bmatrix} 1 & X_{11} & \dots & X_{1p} \\ 1 & X_{21} & \dots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \dots & X_{np} \end{bmatrix}$$

Salah satu metode yang dapat digunakan untuk memperoleh estimasi tersebut adalah metode *Ordinary Least Square* (OLS). Prinsip dari metode OLS adalah dengan meminimalkan jumlah kuadrat residual yang dihasilkan oleh model dengan persamaan berikut:

$$L = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \{y_i - \sum_{k=0}^p \hat{\beta}_k X_{ik}\}^2 \quad (2)$$

Kemudian didapat estimasi parameter  $\boldsymbol{\beta}$  dengan rumus  $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$ .

### 2.3. Uji Heterogenitas Spasial

Menurut Anselin (1988) untuk mendeteksi adanya efek heterogenitas spasial dapat digunakan uji Breusch-Pagan. Uji hipotesis yang digunakan pada uji Breusch-Pagan adalah:

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

$$H_1 : \text{minimal terdapat satu } i \text{ sedemikian sehingga } \sigma_i^2 \neq \sigma^2 \quad (i = 1, 2, \dots, n)$$

Taraf signifikansi =  $\alpha$

Statistik Uji

$$BP = \frac{1}{2} \mathbf{f}^T \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{f} \quad (3)$$

$\mathbf{f} = (f_1, f_2, \dots, f_n)^T$  dengan  $f_i = \left( \frac{e_i^2}{\sigma^2} - 1 \right)$ . Matriks  $\mathbf{Z}$  merupakan matriks berukuran  $n \times (p + 1)$  yang berisi variabel penjelas yang telah distandarkan. Nilai BP merupakan sebuah nilai skalar yang mengikuti distribusi *Chi-Square* berderajat bebas  $p$ .

Kriteria Keputusan

Tolak  $H_0$  jika  $BP > \chi_{(\alpha; p)}^2$  dengan  $p$  adalah banyaknya variabel bebas atau jika  $p\text{-value} < \alpha$ . Apabila  $H_0$  ditolak maka terjadi adanya perbedaan karakteristik satu wilayah dengan wilayah lainnya, sehingga perlu dilakukan permodelan menggunakan *Geographically Weighted Regression* (GWR).

### 2.4. Pendeteksian Pencilan Spasial

Menurut Shekar (2003) dalam Maimon dan Rokach (2010), pencilan spasial (*spatial outlier*) didefinisikan sebagai ketidakstabilan lokal atau objek spasial yang menggambarkan atribut non-spasial yang relatif ekstrim atau berbeda secara signifikan dari objek spasial lainnya dalam lingkup ketetanggaan. Salah satu metode yang dapat digunakan untuk mendeteksi adanya pencilan spasial adalah Statistik Uji Spasial  $Z$  dengan uji hipotesis sebagai berikut:

$$H_0 : \mu_i = \mu_{S(x)} \text{ (Nilai amatan lokasi ke-} i \text{ bukan merupakan pencilan spasial)}$$

$$H_1 : \mu_i \neq \mu_{S(x)} \text{ (Nilai amatan lokasi ke-} i \text{ merupakan pencilan spasial)}$$

Taraf signifikansi =  $\alpha$

Statistik Uji

$$Z_{hitung} = \left| \frac{S(x) - \mu_{S(x)}}{\sigma_{S(x)}} \right| \quad (4)$$

dengan,  $S(x)$  adalah selisih antara nilai amatan ke- $i$  dengan rata-rata nilai amatan lokasi yang dekat dengan lokasi  $i$ .

Kriteria Keputusan

Nilai amatan dari lokasi ke- $i$  dideteksi sebagai pencilan spasial apabila nilai  $Z_{hitung} > Z_{\frac{\alpha}{2}}$ .

Nilai  $Z_{\frac{\alpha}{2}}$  merupakan nilai  $Z$  tabel distribusi Normal Baku untuk taraf signifikansi  $\alpha$ .

### 2.5. Geographically Weighted Regression

*Geographically Weighted Regression* (GWR) merupakan pengembangan dari model regresi dengan parameter dihitung pada setiap lokasi pengamatan. Model GWR menurut Fotheringham *et al.* (2002) dapat ditulis sebagai berikut:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i) x_{ik} + \varepsilon_i, \quad i=1, 2, \dots, n \quad (5)$$

Penaksiran parameter menggunakan WLS dilakukan sama seperti pada regresi linier, yaitu dengan meminimalkan jumlah kuadrat residual dengan persamaan sebagai berikut:

$$\sum_{j=1}^n w_j(u_i, v_i) \varepsilon_j^2 = \sum_{j=1}^n w_j(u_i, v_i) [y_j - \beta_0(u_i, v_i) - \sum_{k=1}^p \beta_k(u_i, v_i) x_{ik}]^2 \quad (6)$$

Apabila dituliskan dalam bentuk matriks akan menjadi

$$\begin{aligned} \boldsymbol{\varepsilon}^T \mathbf{W}(u_i, v_i) \boldsymbol{\varepsilon} &= [\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}(u_i, v_i)]^T \mathbf{W}(u_i, v_i) [\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}(u_i, v_i)] \\ &= \mathbf{Y}^T \mathbf{W}(u_i, v_i) \mathbf{Y} - 2\boldsymbol{\beta}^T(u_i, v_i) \mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{Y} + \boldsymbol{\beta}^T(u_i, v_i) \mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X} \boldsymbol{\beta}(u_i, v_i) \end{aligned}$$

Sehingga estimasi parameter model GWR dapat diperoleh dari persamaan di bawah:

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

Matriks pembobot pada model GWR merupakan aspek yang sangat penting. Peran pembobot mewakili letak data observasi satu dengan yang lainnya. Matriks pembobot merupakan matriks diagonal dari pembobot masing-masing lokasi di sekitar amatan yang digambarkan sebagai berikut:

$$\mathbf{W}(u_i, v_i) = \begin{bmatrix} w_{i1} & 0 & \dots & 0 \\ 0 & w_{i2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_{in} \end{bmatrix}$$

dimana  $w_{ij} = w_j(u_i, v_i)$  dengan  $j = 1, 2, \dots, n$  merupakan pembobot lokasi ke- $j$  untuk mengestimasi parameter pada lokasi ke- $i$ .

Salah satu metode yang sering digunakan dalam pembobotan model GWR adalah dengan fungsi kernel. Pada penelitian ini digunakan fungsi *Adaptive Gaussian Kernel* dengan rumus:

$$w_j(u_i, v_i) = \exp\left(-\frac{1}{2} \left(\frac{d_{ij}}{h_{i(q)}}\right)^2\right)$$

dengan,

$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2}$  merupakan jarak *euclidean* antara lokasi ke- $i$  dengan lokasi ke- $j$

$h_{i(q)}$  = *bandwidth adaptive* yang menetapkan  $q$  sebagai jarak terdekat dari lokasi ke- $i$

Pemilihan *bandwidth* memiliki pengaruh yang besar terhadap hasil yang diperoleh dari GWR. Maka dari itu perlu dilakukan pemilihan nilai *bandwidth* yang optimum agar menghasilkan model yang baik. Salah satu metode yang dapat digunakan untuk memilih nilai *bandwidth* yang optimum adalah menggunakan *Cross-Validation* (CV) yang dirumuskan sebagai berikut:

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

Dimana  $\hat{y}_{\neq i}(h)$  adalah nilai estimasi dari  $y_i$  dimana pengamatan di lokasi ke- $i$  dihilangkan dari proses estimasi.

## 2.6. Robust Geographically Weighted Regression (RGWR)

Keberadaan pencilan spasial seringkali susah untuk dideteksi pada model GWR. Oleh sebab itu, diperlukan sebuah metode penaksiran yang lebih *robust*. *Robust* dapat diartikan sebagai ketidaksensitifan atau ketegaran terhadap perubahan-perubahan kecil dari asumsi. Salah satu metode yang dapat diterapkan pada model GWR adalah dengan LAD. Model yang digunakan pada RGWR sama dengan model yang digunakan pada model GWR. Begitu pula pada matriks pembobot yang digunakan, yang membedakan adalah pada kriteria yang digunakan pada pemilihan *bandwidth* yang optimum.

Kriteria pemilihan *bandwidth optimum* pada RGWR dapat dilakukan dengan prosedur kriteria Absolute Cross Validation (ACV). Menurut Wang dan Scott (1994), skor ACV tidak terpengaruh oleh keberadaan pencilan dan akibatnya skor ACV lebih *robust*

daripada skor CV. Kriteria ACV menggunakan nilai mutlak dari selisih variabel respon  $y_i$  dan nilai estimasi  $\hat{y}_{\neq i}(h)$  yang dirumuskan sebagai berikut:

$$ACV(h) = \sum_{i=1}^n |y_i - \hat{y}_{\neq i}(h)| \quad (8)$$

Nilai optimal dari *bandwidth*  $h$  dapat dipilih dengan cara yang sama dengan kriteria pada CV, yaitu dengan memilih nilai *bandwidth*  $h$  yang menghasilkan ACV(h) terkecil.

Estimasi parameter dari model RGWR dapat dilakukan dengan mengadopsi persamaan berikut:

$$L = \sum_{i=1}^n |\varepsilon_i| = \sum_{i=1}^n |y_i - \sum_{k=0}^p \hat{\beta}_k X_{ik}| \quad (9)$$

Solusi untuk menghasilkan parameter regresi  $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)$  tidak dapat dilakukan dengan proses diferensiasi seperti pada metode OLS. Menurut Wagner (1959) penyelesaian solusi dari regresi dengan LAD sama dengan penyelesaian menggunakan program linier. Estimasi parameter model RGWR dalam bentuk pemrograman linier berdasarkan sebagai berikut:

meminimalkan  $\sum_{i=1}^n (\varepsilon_i^+ + \varepsilon_i^-) w_j(u_i, v_i)$ ,

dengan kendala

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

dan  $\varepsilon_i^+, \varepsilon_i^- \geq 0$

$\varepsilon_i^+ = \varepsilon_i I(\varepsilon_i > 0)$ ,  $\varepsilon_i^- = -\varepsilon_i I(\varepsilon_i < 0)$

Permasalahan program linier di atas dapat diselesaikan menggunakan algoritma Simpleks.

## 2.7. Ukuran Ketepatan Model

*Mean Absolute Percentage Error* atau (MAPE) merupakan salah satu ukuran yang dapat dijadikan acuan dalam mengevaluasi keakuratan suatu model. MAPE mengukur nilai dugaan model yang dinyatakan dalam bentuk rata-rata persentase absolut residual. Formula MAPE dapat dituliskan sebagai berikut (Myttenaere, et al., 2015) :

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \left( \frac{y_i - \hat{y}_i}{y_i} \right) \times 100\% \right|$$

dengan

$y_i$  : nilai  $y$  aktual

$\hat{y}_i$  : nilai  $y$  prediksi dari model

Semakin kecil nilai MAPE maka model yang didapatkan semakin baik. Berikut merupakan tabel kriteria MAPE (Chang, et al., 2007) :

**Tabel 1.** Nilai Signifikansi dan Kriteria MAPE

MAPE	Signifikansi
<10%	Kinerja model sangat baik
10% – 20%	Kinerja model baik
20% - 50%	Kinerja model masih dalam batas wajar
>50%	Kinerja model kurang baik

## 3. METODE PENELITIAN

### 3.1. Sumber Data dan Variabel Penelitian

Data yang digunakan dalam penelitian tugas akhir ini merupakan data sekunder tentang faktor-faktor yang memengaruhi kejadian diare di 16 kecamatan di Kota Semarang pada tahun 2015. Data ini bersumber Buku Profil Kesehatan Kota Semarang tahun 2015.

### 3.2. Variabel Penelitian

Variabel yang digunakan terdiri dari variabel respon dan variabel dan 6 variabel bebas yang diduga memengaruhi kejadian diare. Variabel respon (Y) adalah persentase kejadian diare di tiap kecamatan. Sedangkan yang merupakan variabel bebas antara lain, persentase rumah sehat ( $X_1$ ), persentase rumah tangga dengan sumber mata air sumur tidak terlindung ( $X_2$ ), persentase kepadatan penduduk ( $X_3$ ), persentase penduduk yang memiliki akses air minum layak ( $X_4$ ), persentase rumah tangga tanpa fasilitas BAB ( $X_5$ ), dan persentase rumah tangga dengan jamban sendiri ( $X_6$ ).

### 3.3. Tahapan Analisis Data

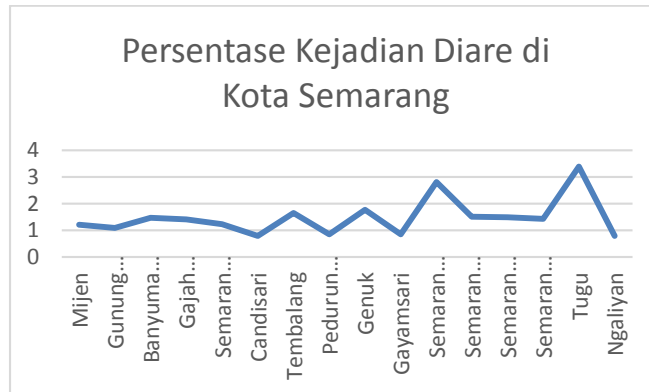
Analisis data yang digunakan pada penelitian ini ada beberapa langkah yang digunakan. Langkah-langkah yang akan digunakan antara lain:

1. Menentukan variabel bebas yang diduga mempengaruhi kejadian diare di Kota Semarang
2. Melakukan analisis deskriptif terhadap data kejadian diare per kecamatan di Kota Semarang pada tahun 2015
3. Melakukan analisis regresi linier berganda terhadap variabel-variabel bebas awal. Variabel bebas tersebut adalah persentase rumah sehat ( $X_1$ ), persentase rumah tangga dengan sumber mata air sumur tidak terlindung ( $X_2$ ), persentase kepadatan penduduk ( $X_3$ ), persentase penduduk yang memiliki akses air minum layak ( $X_4$ ), persentase rumah tangga tanpa fasilitas BAB ( $X_5$ ), dan persentase rumah tangga dengan jamban sendiri ( $X_6$ ).
4. Melakukan uji heterogenitas spasial pada regresi linier berganda dengan menggunakan uji *Breusch Pagan*.
5. Melakukan pendekteksian pencilan dengan menggunakan Statistik Uji Spasial Z. Apabila uji heterogenitas spasial terpenuhi dan terdapat pencilan, maka model regresi linier berganda tidak dapat digunakan untuk pemodelan.
6. Melakukan analisis *Robust Geographically Weighted Regression*, apabila terjadi heterogenitas spasial dan terdapat pencilan yang terdiri dari:
  - a. Menghitung jarak *Euclidean*
  - b. Menentukan bandwidth optimum dengan menggunakan metode *Absolute Cross-Validation*
  - c. Menghitung matriks pembobot dengan fungsi kernel *Adaptive Gaussian Kernel*
  - d. Melakukan estimasi parameter model *Robust GWR* dengan metode Mutlak Simpangan Terkecil
  - e. Menghitung ukuran ketepatan model dengan MAPE

## 4. HASIL DAN PEMBAHASAN

### 4.1 Deskripsi data

Berdasarkan Gambar 1, dapat diketahui bahwa kecamatan yang memiliki persentase kejadian diare tertinggi dan terendah di Kota Semarang. Kecamatan Tugu menjadi kecamatan yang memiliki kejadian diare tertinggi di Kota Semarang, yaitu sebesar 3,3862% . Persentase kejadian diare terendah berada di Kecamatan Ngaliyan, yaitu sebesar 0,7875%. Rata-rata persentase kejadian diare di Kota Semarang pada tahun 2015 adalah 1,4795%.



Gambar 1. Grafik Persentase Kejadian Diare di Kota Semarang Tahun 2015

#### 4.2. Regresi Linier Berganda

Langkah selanjutnya adalah membentuk model regresi linier dengan metode kuadrat terkecil (OLS). Variabel persentase kejadian diare di Kota Semarang tahun 2015 (Y) dihubungkan dengan 6 variabel bebas yang diduga mempengaruhinya. Maka, diperoleh model regresi sebagai berikut:

$$\hat{y} = 19,995 - 0,01502X_1 - 0,4057X_2 - 6 \times 10^{-3}X_3 - 0,1539X_4 + 0,2564X_5 - 0,02059X_6$$

Namun, dari 6 variabel bebas hanya ada 2 variabel bebas yang signifikan terhadap model, yaitu variabel  $X_4$  (persentase penduduk yang memiliki akses air minum layak) dan  $X_5$  (persentase rumah tangga tanpa fasilitas BAB). Sehingga diperoleh model yang signifikan sebagai berikut:

$$\hat{y} = 14,629537 - 0,135580X_4 + 0,172463X_5$$

#### 4.3. Uji Heterogenitas Spasial dan Pendeteksian Pencilan Spasial

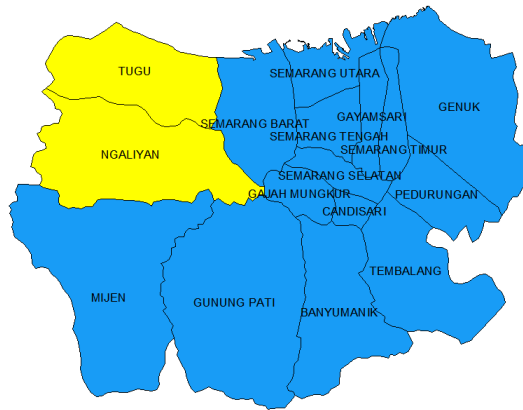
Tabel 2. Uji Heterogenitas Spasial

<i>Breusch-Pagan</i>	<i>df</i>	<i>P-Value</i>
11,089	2	0,003909

Karena nilai  $BP = 11,089 > \chi^2_{(0,05;2)} = 5,9915$  dan nilai  $p\text{-value} = 0,0000309$  lebih kecil dari  $\alpha = 0,05$  sehingga menunjukkan bahwa terdapat heterogenitas spasial atau terdapat karakteristik yang berbeda terhadap data kejadian diare di Kota Semarang pada tahun 2015 di setiap kecamatan.

Selain efek spasial, yang menjadi fokus perhatian pada model RGWR adalah keberadaan pencilan. Metode yang dapat digunakan untuk pendeteksian pencilan secara spasial adalah Statistik Uji Spasial Z. Dalam melakukan pendeteksian pencilan spasial terlebih dahulu menentukan tetangga terdekat dari suatu lokasi. Untuk menentukan tetangga terdekat dari suatu lokasi, maka dapat digunakan kaidah *Rook Contiguity* dimana ketetanggaan ditentukan berdasarkan sisi-sisi yang saling bersinggungan.

Berdasarkan Statistik Uji Spasial Z dapat disimpulkan terdapat dua residual dari model regresi OLS yang merupakan pencilan spasial, yaitu pada lokasi 16 (Kecamatan Ngaliyan) dan lokasi 15 (Kecamatan Tugu). Lokasi yang merupakan pencilan dapat dilihat pada gambar:



Gambar 2. Kecamatan yang Merupakan Pencilan Spasial

#### 4.4. Model GWR

Sebelum memodelkan kejadian diare di Kota Semarang menggunakan RGWR, terlebih dahulu akan dimodelkan dengan GWR. Berikut merupakan ringkasan parameter model GWR dari variabel yang berpengaruh:

Tabel 3. Ringkasan Estimasi Parameter Model GWR untuk Variabel Signifikan

Parameter	Minimum	Maksimum	Rata-Rata	Median	Standar Deviasi
$\hat{\beta}_0$	9,8417	10,8169	10,1706	10,1054	0,2530
$\hat{\beta}_4$	-0,1017	-0,0916	-0,0950	-0,0942	0,0027
$\hat{\beta}_5$	0,1881	0,2122	0,2040	0,2054	0,0071

Berdasarkan Tabel 3 terlihat bahwa variabel-variabel penjelas pada setiap kecamatan memiliki pengaruh yang berbeda-beda terhadap persentase kejadian diare.

#### 4.5. Model RGWR

Berdasarkan kriteria ACV, maka dihasilkan *bandwidth Adaptive Gaussian Kernel* untuk setiap lokasi sebagai berikut:

Tabel 4. Nilai *Adaptive Bandwidth* Setiap Lokasi Model RGWR

Lokasi	Bandwidth	Lokasi	Bandwidth
Mijen	0,2256	Genuk	0,2256
Gunung Pati	0,1903	Gayamsari	0,1803
Banyumanik	0,1368	Semarang Timur	0,1782
Gajah Mungkur	0,1387	Semarang Utara	0,1668
Semarang Selatan	0,1669	Semarang Tengah	0,1625
Candisari	0,1450	Semarang Barat	0,1392
Tembalang	0,1567	Tugu	0,1429
Pedurungan	0,1908	Ngaliyan	0,1442

Kemudian menghitung matriks pembobot untuk setiap lokasi. Sebagai contohnya adalah matriks pembobot untuk Kecamatan Tugu:

$$W(u_i, v_i) = \begin{bmatrix} 0,9101 & 0 & \dots & 0 \\ 0 & 0,8073 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0,9968 \end{bmatrix}$$

Ringkasan dari nilai estimasi parameter untuk model RGWR dengan metode mutlak simpangan terkecil yang diperoleh dengan menggunakan *software* R adalah sebagai berikut:



Tabel 5. Ringkasan Estimasi Parameter Model RGWR

Parameter	Minimum	Maksimum	Rata-Rata	Median	Standar Deviasi
$\hat{\beta}_0$	1,0168	12,9642	6,0966	2,9351	5,5497
$\hat{\beta}_4$	-0,1205	0,0048	-0,0498	-0,0194	0,0575
$\hat{\beta}_5$	-0,1314	0,2839	0,0747	0,0296	0,1798

Berdasarkan Tabel 5 terlihat bahwa variabel-variabel penjelas pada setiap kecamatan memiliki pengaruh yang berbeda-beda terhadap persentase kejadian diare. Variabel penjelas yang digunakan adalah variabel penjelas yang signifikan pada model GWR. Sebagai contoh model RGWR yang dihasilkan untuk Kecamatan Tugu adalah

$$\hat{y} = 12,9642 - 0,1205X_4 + 0,2839X_5$$

Secara garis besar, variabel penjelas yang memiliki koefisien regresi terbesar adalah pada variabel  $X_5$  (persentase rumah tangga tanpa fasilitas BAB). Jika koefisien variabel penjelas lain dianggap tetap, maka persentase kejadian diare di Kecamatan Tugu akan meningkat sebesar 0,2839 apabila persentase rumah tangga tanpa fasilitas BAB bertambah 1%. Kemudian pada variabel  $X_4$  (persentase penduduk yang memiliki akses air minum layak), jika persentase penduduk yang memiliki akses air minum layak bertambah 1%, maka kejadian diare di Kecamatan Tugu akan menurun sebesar -0,205 dengan asumsi variabel lain tetap.

Berdasarkan hasil estimasi yang diperoleh, dihasilkan model RGWR dengan MAPE sebesar 16,3396%. Artinya kinerja model yang diperoleh dari estimasi parameter menggunakan RGWR dengan metode mutlak simpangan terkecil adalah baik karena berada pada rentang 10% - 20%. Sedangkan kinerja model yang diperoleh dari estimasi parameter menggunakan GWR adalah 35,8665% yang artinya kinerja model masih dalam batas wajar karena berada pada rentang 20% - 50%.

## 5. KESIMPULAN

Pemodelan kejadian diare yang dihasilkan menggunakan RGWR berbeda-beda pada setiap lokasi. Secara garis besar, variabel penjelas yang memiliki koefisien regresi terbesar adalah pada variabel  $X_5$  (persentase rumah tangga tanpa fasilitas BAB).

Model RGWR mampu menghasilkan estimasi parameter yang lebih baik dibandingkan model GWR. Hal ini ditunjukkan oleh nilai MAPE yang dihasilkan model RGWR sebesar 16,3396%, sedangkan model GWR sebesar 35,8665%. Hal ini menunjukkan bahwa penggunaan teknik yang *robust* untuk data spasial yang mengandung pencilan memiliki performa yang baik.

## DAFTAR PUSTAKA

- Afifah, R., Andriyana, Y., & Jaya, I.G.N. 2017. Robust Geographically Weighted Regreession with Least Absolute Deviation Method in Case of Poverty in Java Island. *AIP Conference Proceeding*, Volume 1827, Issue 1.
- Anselin, L. 1988. *Spatial Econometrics: Method and Models*. Kluwer Academic Publisher: The Netherlands.
- Ariani, A.P. 2016. *Diare: Pencegahan dan Pengobatannya*. Yogyakarta: Nuha Medika.
- Bekti, R.D. 2014. Spatial Pattern of Diarrhea Based on Regional Economic and Environment by Spatial Autoregressive Model. *AIP Conference Proceeding*, Volume 1621, Issue 1.
- Chang, P, C., Wang, Y, W., & Liu, C, H. 2007. The Development of a Weighted Evolving Fuzzy Neural Network for PCB Sales Forecasting. *Expert System with Application*, Volume 32, page: 86 – 96.

- Fotheringham, A.S., Charlton, M., & Brunsdon, C. 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester. UK : John Wiley and Sons.
- Gujarati, D. 2006. *Ekonometrika Dasar*. Jakarta: Erlangga. Terjemahan dari *Basic Econometrics*.
- Hariyanti, L. 2010. *Spatial Autoregressive Structural Equational Model pada Prevalensi Diare (Studi Kasus di Pulau Jawa dan Madura)* (Tesis). Bandung: Universitas Padjajaran.
- Kementerian Kesehatan Republik Indonesia. 2015. *Buku Profil Kesehatan Kota Semarang 2015*. Jakarta : Kementerian Kesehatan RI.
- Larson, R. & Falvo, D. C. 2009. *Elementary Linear Algebra*. Boston: Houghton Mifflin Harcourt Publishing Company.
- Maimon, O. & Rokach, L. 2010. *Data Mining and Knowledge Discovery Handbook Second Edition*. New York: Springer.
- Montgomery, C. D. & Runger, G.C. 2003. *Applied Statistics and Probability for Engineers Third Edition*. New York: John Wiley and Sons, Inc.
- Myttenaere, D.A., Golden, B., Grand, B. L., & Rossi, F. 2016. Mean Absolute Percentage Error for Regression Models. *Neurocomputing, Elsevier, 2016, 23rd European Symposium on Artificial Neural Networks (ESANN 2015)*, Volume 192, page: 38 – 48.
- Wagner, H.M. 1959. Linear Programming Techniques for Regression Analysis. *Journal of the American Statistical Association*, Vol. 54, No. 285, page: 206 – 212.
- Wang, F.T. & Scott, D.W. 1994. The L1 Method for Robust Non Parametric Regression. *Journal of the American Statistical Association*, Vol. 89, No. 425, page: 65 – 76.
- Zhang, H. & Mei, C. 2011. Local Least Absolute Deviation Estimation of Spatially Varying Coefficient Models: Robust Geographically Weighted Regression Approaches. *International Journal of Geographical Information Science*, Volume 25, No. 9, page: 1467 – 1489.