

## BAB II TINJAUAN PUSTAKA

### 2.1 *Coronavirus Disease 2019*

*Coronavirus* merupakan virus RNA strain tunggal positif berkapsul dan tidak bersegmen yang tergolong kedalam ordo *Nidovirales* keluarga *Coronavirida*. Struktur *coronavirus* membentuk seperti kubus dengan protein S berlokasi di permukaan virus. Protein S atau spike protein merupakan salah satu protein antigen utama virus dan merupakan struktur utama untuk penulisan gen yang berperan dalam penempelan dan masuknya virus kedalam sel inang (Wang:2020). Covid-19 menyerang manusia ditandai dengan adanya perubahan di paru-paru dan rata-rata jumlah sel limfosit dan trombosit menunjukkan hasil yang lebih rendah dan disertai hipoksemia (Levani, dkk:2020).

Menurut WHO (*World Health Organization*) ada beberapa cara yang menjadi jalur penyebaran virus Corona yaitu:

1. Penyebaran virus corona melalui droplet

Penularan virus corona bisa terjadi melalui droplet saat seseorang batuk, bersin, bernyanyi, berbicara hingga bernapas. Saat melakukan hal-hal tersebut, udara yang keluar dari hidung dan mulut mengeluarkan partikel kecil atau aerosol dalam jarak dekat.

2. Penyebaran virus corona melalui udara

Virus corona bisa menyebar di udara melalui partikel-partikel kecil yang melayang di udara.

### 3. Penyebaran virus corona melalui permukaan yang terkontaminasi

Cara penularan virus corona ini terjadi saat seseorang menyentuh permukaan yang mungkin telah terkontaminasi virus dari orang yang batuk atau bersin. Lalu virus itu berpindah ke hidung, mulut, atau mata yang disentuh setelah menyentuh permukaan yang terkontaminasi tersebut.

Infeksi Covid-19 dapat menimbulkan gejala ringan, sedang dan berat. Menurut PDPI (Persatuan Dokter Paru Indonesia) tahun 2020 adapun kondisi yang dapat muncul jika terinfeksi virus Covid-19:

#### 1. Tidak Berkomplikasi

Kondisi ini merupakan kondisi ringan dan gejala yang muncul berupa gejala yang tidak spesifik. Gejala utama tetap muncul seperti demam, batuk, dapat disertai dengan nyeri tenggorokkan, kongesti hidung, malaise, sakit kepala dan nyeri otot. Pada pasien lanjut usia dan pasien *immunocompromises* presentasi gejala menjadi tidak khas atau atipikal.

#### 2. Pneumonia Ringan

Gejala utama yang dapat muncul seperti demam, batuk, dan sesak.

#### 3. Pneumonia Berat

Gejala yang muncul diantaranya demam dan infeksi saluran pernafasan. Tanda tanda yang muncul yaitu takipnea (frekuensi pernapasan  $> 30x$ /menit), distress pernapasan berat atau saturasi oksigen pasien  $<90\%$  udara luar.

Periode inkubasi untuk Covid-19 antara 3-14 hari. Ditandai dengan kadar leukosit dan limfosit yang masih normal atau sedikit menurun, serta pasien belum merasakan gejala. Selanjutnya, virus mulai menyebar melalui aliran darah, terutama menuju ke organ yang mengekspresikan ACE2 dan pasien mulai

merasakan gejala ringan. Empat sampai tujuh hari dari gejala awal, kondisi pasien mulai memburuk dengan ditandai oleh timbulnya sesak, menurunnya limfosit, dan perburukan lesi di paru. Jika fase ini tidak teratasi, maka dapat terjadi *Acute Respiratory Distress Syndrome* (ARSD), sepsis dan komplikasi lainnya. Tingkat keparahan klinis berhubungan dengan usia (diatas 70 tahun), komorbiditas seperti diabetes, penyakit paru bstruktif kronis (PPOK), hipertensi, dan obesitas (Gennaro dkk., 2020; Susilo dkk., 2020).

Infeksi dari virus mampu memproduksi reaksi imun yang berlebihan pada inang. Pada beberapa kasus, terjadi reaksi yang secara keseluruhan disebut “badai sitokin”. Badai sitokin merupakan peristiwa reaksi inflamasi berlebihan dimana terjadi produksi sitokin yang cepat dan dalam jumlah yang banyak sebagai respon dari suatu infeksi. Dalam kaitannya dengan Covid-19, ditemukan adanya penundaan sekresisitokin dan kemokin oleh sel imun innate dikarenakan blokade oleh protein non-struktural virus. Selanjutnya, hal ini menyebabkan terjadinya lonjakan sitokin proinflamasi dan kemokin (IL-6, TNF- $\alpha$ , IL-8, MCP-1, IL-1  $\beta$ , CCL2, CCL5, dan interferon) melalui aktivasi makrofag dan limfosit. Pelepasan sitokin ini memicu aktivasi sel imun adaptif seperti sel T, neutrofil, dan sel NK, bersamaan dengan terus terproduksinya sitokin proinflamasi. Lonjakan sitokin proinflamasi yang cepat ini memicu terjadinya infiltrasi inflamasi oleh jaringan paru yang menyebabkan kerusakan paru pada bagian epitel dan endotel. Kerusakan ini dapat berakibat pada terjadinya ARDS dan kegagalan multi organ yang dapat menyebabkan kematian dalam waktu singkat (Gennaro dkk., 2020; Lingeswaran dkk., 2020).

Berikut adalah beberapa pemeriksaan yang dapat digunakan untuk mengidentifikasi Covid-19 menurut PDPI.

1. Pemeriksaan radiologi

Berupa foto toraks, CT-scan toraks, dan USG toraks.

2. Pemeriksaan spesimen saluran nafas atas dan bawah.

a. Saluran napas atas dengan swab tenggorokan (nasofaring dan orofaring)

b. Saluran napas bawah (sputum, bilasan bronkos, BAL dan bila menggunakan endotrakeal tube dapat berupa aspirat endotrakeal).

3. Bronkoskopi

4. Pungsi pleura sesuai kondisi

5. Pemeriksaan kimia darah

6. Biakan mikroorganisme dan uji kepekaan dari bahan saluran napas

7. Pemeriksaan feses dan urin

**2.2 Faktor – Faktor yang Mempengaruhi Penyebaran Covid-19 di Indonesia**

Berikut adalah beberapa faktor yang mempengaruhi penyebaran Covid-19 di Indonesia:

1. Kepadatan Penduduk

Meningkatnya jumlah penduduk akibat aktivitas ekonomi yang terus berkembang mendorong bertambahnya daerah permukiman dan menyebabkan naiknya tingkat kepadatan penduduk. Menurut Hardianto (2020), kepadatan penduduk memiliki andil dalam penyebaran Covid-19 di Indonesia, hal ini merujuk pada kenyataan bahwa kawasan perkotaan yang memiliki tingkat

kepadatan penduduk tinggi akan menyebabkan transmisi penyakit lebih cepat dengan rantai penyebaran yang lebih kompak dan kompleks dibandingkan dengan daerah pinggiran. Parameter kepadatan penduduk terbukti mempengaruhi penyebaran Covid-19 di Negara Bagian India Selatan dimana pada penelitian yang dilakukan Arif dan Sengupta (2020), terdapat tiga negara bagian yang memiliki tingkat kepadatan penduduk tinggi dan memiliki penyebaran kasus yang tinggi yaitu Tamil Nadu, Karnataka dan Telangana. Selain itu terdapat penelitian lain yang menyatakan bahwa terdapat korelasi antara kepadatan penduduk terhadap wabah dari suatu penyakit (Li et all (2018).

## 2. Persentase Penduduk Lansia

Menurut Kementerian Kesehatan Republik Indonesia (2020), penduduk lanjut usia (lansia) sangat rentan terserang berbagai penyakit termasuk Covid-19 hal ini dikarenakan dengan bertambahnya usia, tubuh akan mengalami berbagai penurunan akibat proses penuan, hampir semua fungsi organ dan gerak menurun diikuti dengan menurunnya imunitas sebagai pelindung tubuh. Sistem imun yang sudah melemah ditambah adanya penyakit kronis dapat meningkatkan risiko COVID-19 pada lansia, baik risiko terjadinya infeksi virus Corona maupun risiko virus ini untuk menimbulkan gangguan yang parah, bahkan kematian.

## 3. Persentase Tingkat Kemiskinan

Hanoatubun (2020) mengemukakan bahwa salah satu skenario penanganan Pandemi Covid-19 yaitu skenario intervensi kuat, di mana dari hasil analisis sederhana dikatakan bahwa intervensi kuat untuk meminimalisasi

penyebaran kasus coronavirus dapat menurunkan pertumbuhan ekonomi yang lebih parah. Salah satu bentuk dari intervensi kuat ini adalah meningkat angka kemiskinan. Ahmed, Ahmed, Pissarides, & Stiglitz (2020) mengatakan bahwa angka kemiskinan memiliki pengaruh signifikan terhadap total kasus Covid-19. Hal ini terjadi karena individu atau populasi penduduk miskin tidak memiliki akses ke layanan kesehatan sehingga menempatkan mereka pada kriteria kondisi resiko terpapar Covid-19 lebih tinggi. Selain itu, Petet et al (2020) menjelaskan beberapa alasan utama penduduk miskin rentan terpapar Covid-19 adalah pekerjaan yang tidak fleksibel dan kondisi kerja yang tidak stabil .

#### 4. Sumber Air Bersih layak

Menurut Kodoatie (2003), air bersih adalah air yang dipakai sehari-hari untuk keperluan mencuci, mandi, memasak, dan dapat diminum setelah dimasak. Menurut Suripin (2002), yang dimaksud air bersih yaitu air yang aman (sehat) dan baik untuk diminum, tidak berwarna, tidak berbau, dengan rasa yang segar. Berdasarkan kedua pendapat tersebut, air bersih terdiri dari air yang dapat dikonsumsi (air minum) dan juga air yang dapat digunakan untuk keperluan lainnya dalam kegiatan rumah tangga. WHO (2020) menetapkan protokol kesehatan yang sebagian besar berisi anjuran untuk menjaga kebersihan dan kesehatan. Protokol kesehatan saat pandemi Covid-19 dan harus tetap dilaksanakan saat *new normal* antara lain mencuci tangan dengan air bersih dan juga mandi apabila pulang bepergian sebagai wujud perilaku hidup bersih dan sehat (PHBS). Kedua aktivitas tersebut tentu membutuhkan air bersih dalam jumlah yang mencukupi.

### 2.3 Pendeteksian Efek Spasial

Analisis regresi adalah salah satu analisis yang sering digunakan untuk mengkaji hubungan antara variable dependen dengan satu atau dengan beberapa variable independen. Secara umum dapat model regresi dapat ditulis:

$$y_i = \beta_0 + \sum_{k=1}^k \beta_k x_{ik} + \varepsilon_i \quad i = 1, 2, \dots, n \quad (2.1)$$

Dengan

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

$\beta_0, \beta_k$  = Intersep dan koefisien regresi dari variable penjelas ke- $k$

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

$\varepsilon_i$  = Komponen sisaan pada observasi ke- $i$  yang diasumsikan berdistribusi normal

Untuk menaksir koefisien regresi  $\beta_1, \beta_2, \beta_3, \dots, \beta_n$  kita dapat menggunakan metode kuadrat terkecil atau *Ordinary Least Square* (OLS). Prinsip dasar metode kuadrat terkecil adalah meminimumkan jumlah kuadrat residual yaitu meminimumkan  $\sum \varepsilon_i^2$  (Suryanto: 1998).

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_t - \beta_0 - \sum_{i=1}^n \beta_n x_{in})^2 \quad (2.2)$$

Model tersebut merupakan model regresi global dimana semua parameternya berlaku untuk semua lokasi penelitian (Fotheringham, dkk:2002). Artinya hubungan antara variable respon dan variable penjelas diasumsikan homogen untuk semua lokasi data tersebut diamati. Mengestimasi model regresi dengan OLS, sangat bergantung pada beberapa asumsi yang harus dipenuhi, yaitu normalitas, nonautokorelasi, homoskedastisitas dan multikolinieritas. Apabila semua asumsi tersebut terpenuhi, maka akan dihasilkan estimator yang bersifat

*Best Linear Unbiased Estimator* (BLUE). Seandainya ada salah satu asumsi tidak terpenuhi, maka estimator yang paling baik tidak akan diperoleh.

Pada data spasial, asumsi nonautokorelasi dan homoskedastisitas akan sangat sulit terpenuhi karena adanya efek spasial pada jenis data tersebut. Efek ketergantungan spasial akan melanggar asumsi nonautokorelasi pada model dengan OLS. Suatu pengamatan (dalam hal ini lokasi) akan mempengaruhi pengamatan yang lain. Efek yang lain yaitu heterogenitas spasial akan membuat asumsi homoskedastisitas tidak terpenuhi. Adanya efek ini akan membuat varian pada model regresi tidak lagi konstan, tetapi berbeda-beda sesuai dengan amatannya. Pendeteksian ketergantungan spasial dapat dilakukan menggunakan Indeks Moran. Indeks Moran membandingkan nilai suatu variabel pada suatu lokasi dengan variabel yang sama pada lokasi lainnya. Hipotesis yang digunakan yaitu:

$H_0$ : Tidak terdapat autokorelasi spasial

$H_1$ : Terdapat autokorelasi spasial

Statistik uji Indeks Moran (Anselin, 1998) yaitu:

$$Z(I) = \frac{1-E(I)}{\sqrt{Var(I)}} \approx N(0,1) \quad (2.3)$$

Dengan  $E(I) = -\frac{1}{n-1}$ ,  $Var(I) = \frac{n^2.S_1 - n.S_2 + 3.S_0^2}{(n^2-1)S_0^2} - [E(I)]^2$

$$S_0 = \sum_{i=1}^n \sum_{j=1}^n W_{ij} \quad S_1 = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (W_{ij} + W_{ji})^2 \quad S_2 = \sum_{i=1}^n (\sum_{j=1}^n W_{ij} + \sum_{j=1}^n W_{ji})^2$$

dimana  $i$  dan  $j$  merupakan indeks untuk lokasi ke- $i$  dan ke- $j$  dari total  $n$  lokasi.  $X$  merupakan variabel yang akan diuji, serta  $W_{ij}$  merupakan elemen dari matrik pembobot spasial antara lokasi ke- $i$  dan ke- $j$ . Sedangkan untuk mendeteksi adanya



efek heterogenitas spasial, prosedur yang dapat dilakukan adalah dengan uji *Breusch-Pagan* (Anselin:1988). Hipotesis yang digunakan dalam pengujian heterogenitas spasial dengan uji *Breusch-Pagan* adalah:

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

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

Statistik Uji *Breusch – Pagan* adalah :

$$BP = \frac{1}{2} \mathbf{f}^t \mathbf{z} (\mathbf{z}^T \mathbf{z})^{-1} \mathbf{z}^T \mathbf{f} \quad (2.4)$$

$$\mathbf{f} = (f_1, f_1, \dots, f_n)^T \text{ dengan } f_i = \left( \frac{e_i^2}{\sigma^2} - 1 \right).$$

Matriks  $\mathbf{Z}$  merupakan matriks berukuran  $n \times (p+1)$  yang berisi variabel penjelas yang menyebabkan heteroskedastisitas telah distandarkan. Nilai BP merupakan sebuah nilai scalar yang mengikuti distribusi *Chi-Square* berderajat bebas  $p$ . Tolak  $H_0$  jika  $BP > \chi_{(\alpha;p)}^2$  dengan  $p$  adalah banyaknya variabel independen atau  $p\text{-value} < \alpha$ . 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

Metode *Ordinary Least Squares (OLS)* merupakan metode yang banyak digunakan untuk mengestimasi nilai koefisien regresi. Metode OLS ini mengestimasi nilai koefisien regresi dengan meminimumkan jumlah kuadrat error. Akan tetapi, metode OLS ini mempunyai kelemahan ketika terdapat pencilan pada data. Keberadaan pencilan akan menghasilkan residual yang besar. Salah satu metode yang umum digunakan untuk mendeteksi keberadaan pencilan adalah dengan *Boxplot*. Metode *Boxplot* menggunakan nilai kuartil dan jangkauan

interkuartil. Data pengamatan yang akan dibagi berdasarkan kuartil 1, 2, dan 3. Selanjutnya nilai interkuartil (IQR) dihitung dengan rumus:

$$IQR = Q_3 - Q_1 \quad (2.5)$$

Dimana IQR merupakan jangkauan (interquartile),  $Q_1$  merupakan kuartil ke-1 dan  $Q_3$  merupakan kuartil ke-3. Suatu pengamatan dikatakan sebagai pencilan jika memenuhi kondisi:

1. Nilainya lebih kecil dari  $\{Q_1 - (1.5 \times IQR)\}$ , atau
2. Nilainya lebih besar dari  $\{Q_3 - (1.5 \times IQR)\}$ ,

Dalam mengeksplorasi analisis data spasial, anselin (2005) metransformasi bentuk *boxplot* ini dalam ke dalam bentuk peta yang disebut *boxmap*. *Boxmap* ini memperlihatkan data yang menjadi secara spasial.

Pencilan dapat terjadi karena beberapa alasan, diantaranya akibat kesalahan dalam pencatatan hasil pengamatan atau kesalahan ketika menyiapkan peralatan yang digunakan dalam mengumpulkan data. Pencilan yang diakibatkan oleh kesalahan seperti ini dapat dikeluarkan dari kumpulan data yang akan dianalisis. Namun ada pencilan yang memang menunjukkan informasi yang tidak dapat diberikan oleh data lainnya yang akan berpengaruh besar terhadap model yang akan dihasilkan. Pencilan yang seperti ini, akan tidak bijak jika dikeluarkan dari proses pengolahan model. Oleh sebab itu, data yang mengandung pencilan tersebut akan lebih baik apabila diestimasi dengan metode yang *robust*, yang tidak terlalu terpengaruh oleh pencilan tetapi juga tidak menghilangkan informasi yang dibawa oleh data tersebut.

## 2.5 Geographically Weighted Regression

*Geographically Weighted Regression* (GWR) merupakan pengembangan dari model regresi dengan parameter dihitung pada setiap lokasi pengamatan. Model GWR digunakan untuk variabel respon yang bersifat kontinu yang akan menghasilkan dugaan parameter model regresi secara lokal untuk setiap lokasi. GWR digunakan untuk data yang non stasioner secara spasial atau memiliki *spatial heterogeneity* (Fotheringham, dkk: 1998). Setiap lokasi pengamatan akan memiliki nilai dugaan parameter yang berbeda-beda. Parameter dalam GWR akan berjumlah sebanyak dimana adalah banyaknya lokasi pengamatan dan adalah banyaknya parameter pada masing-masing lokasi. Dalam model GWR variabel respon  $y$  diprediksi oleh variabel prediktor, dimana masing-masing koefisien regresinya bergantung pada lokasi dimana data tersebut diamati. Model GWR dapat ditulis sebagai berikut (Fotheringham, dkk: 2002):

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^k \beta_k(u_i, v_i) x_{ik} + \varepsilon_i, \quad i=1,2,\dots,n \quad (2.6)$$

Dimana

$y_t$  = Variabel respon pada lokasi ke- $i$

$(u_i, v_i)$  = Titik koordinat longitude dan latitude lokasi ke- $i$

$\beta_k(u_i, v_i)$  = koefisien regresi variable penjelas ke- $k$  pada lokasi ke- $i$

$x_{ik}$  = variabel penjelas ke- $k$  pada lokasi ke- $i$

$\varepsilon_i$  = komponen sisaan pada lokasi ke- $i$  yang diasumsikan berdistribusi normal

### 1. Pembobot Model GWR

Pada model GWR, matriks pembobot menggambarkan pentingnya lokasi sekitar untuk mengestimasi parameter pada lokasi yang diamati. Matrik

pembobot merupakan matrik diagonal dari pembobot masing-masing lokasi disekitar lokasi amatan. Matrik pembobot dapat digambarkan dengan bentuk:

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

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- $i$ . Besarnya nilai pembobot pada setiap lokasi dapat dihitung dengan menggunakan beberapa metode, diantaranya:

a. Fungsi Jarak *Euclidean*

Metode ini menghitung nilai pembobot pada setiap lokasi dengan menggunakan jarak dari lokasi tersebut dengan lokasi sekitarnya. Misalnya diketahui  $d_{ij}$  merupakan fungsi jarak yang mewakili pembobot antara lokasi ke- $i$  dan lokasi ke- $j$  dimana  $d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2}$  merupakan jarak *Euclidean* antara lokasi  $(u_i, v_i)$  dan  $(u_j, v_j)$ . apabila terdapat lokasi ke- $j$  yang jaraknya berada diluar radius ke- $i$ , maka pembobot lokasi tersebut diberikan nol. Pembobot ini dapat ditulis:

$$w_j(u_i, v_i) = \begin{cases} 1, & \text{jika } d_{ij} \leq r \\ 0, & \text{jika } d_{ij} > r \end{cases}$$

b. Fungsi *Kernel*

Salah satu metode yang sering digunakan adalah dengan menggunakan fungsi *Kernel*. Ada beberapa fungsi *Kernel* yang dapat digunakan, diantaranya yaitu:

1) Fungsi *Kernel Gaussian*

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

(2.7)

2) Fungsi *Kernel Bi-square*

$$w_j(u_i, v_i) = \begin{cases} (1 - (d_{ij}/h)^2)^2, & d_{ij} \leq h \\ 0, & d_{ij} > h \end{cases} \quad (2.8)$$

$h$  adalah parameter non negatif yang bisa disebut dengan parameter penghalus (*bandwith*). Apabila nilai  $i=j$ , maka pembobot pada lokasi tersebut adalah 1. Nilai pembobot akan semakin kecil seiring dengan semakin bertambah besarnya jarak antara lokasi  $i$  dan  $j$  sesuai dengan bentuk kurva *Gaussian*.

Kedua fungsi kernel di atas menggunakan nilai *bandwidth* yang tetap. Ketika titik lokasi yang diamati tersebar tidak merata, atau dengan kata lain pada sebagian wilayah lokasi-lokasi pengamatan berdekatan sedangkan pada sebagian wilayah yang lain lokasi-lokasi amatan berjauhan, maka fungsi kernel dengan *bandwidth* tetap tidak lagi akurat. Oleh sebab itu akan lebih bijak apabila nilai *bandwidth* menyesuaikan dengan sebaran titik lokasi atau sering disebut *adaptive bandwidth*. Fungsi kernel dengan *bandwidth* berubah-ubah diantaranya adalah sebagai berikut:

1) Fungsi *Adaptif kernel Gaussian*

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

2) Fungsi *Adaptif Kernel Bi-Square*

$$w_j(u_i, v_i) = \begin{cases} \left(1 - \left(\left(d_{ij}/h_i\right)^2\right)^2\right), & d_{ij} \leq h \\ 0, & d_{ij} > h \end{cases}$$

## 2. Pemilihan Bandwith Optimum

Besarnya *bandwidth* yang digunakan sangat penting dalam proses estimasi parameter pada GWR. *Bandwidth* merupakan pengontrol keseimbangan antara kesesuaian kurva terhadap data dengan kemulusan data. Semakin besar *bandwidth*, semakin halus model yang akan terbentuk. Model yang terlalu halus juga menyebabkan model yang terbentuk akan semakin mendekati model global. Sebaliknya, ketika *bandwidth* yang digunakan terlalu kecil, model yang terbentuk juga akan kasar. Agar menghasilkan model regresi yang lebih tepat, maka *bandwidth* yang digunakan haruslah *bandwidth* yang optimum (Fotheringham, dkk: 2002). Kriteria yang sering digunakan untuk mencari nilai (*bandwidth*) yang optimum adalah *Cross Validation* (CV), dengan rumus:

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

$\hat{y}_{\neq i}(h)$  merupakan nilai estimasi dari  $y_i$  dimana pengamatan dilokasi ke-dihilangkan dari proses estimasi. *Bandwidth* yang optimum ditunjukkan oleh nilai CV yang minimum (Fotheringham, dkk : 2002).

## 3. Estimasi Parameter Model GWR

Pada prinsipnya, proses estimasi parameter pada GWR sama dengan proses pada model regresi klasik yang dihitung untuk setiap lokasi. Hanya saja, dalam mengestimasi parameter pada masing-masing lokasi, tidak hanya menggunakan informasi yang dimiliki lokasi tersebut tetapi juga menggunakan informasi yang dimiliki oleh lokasi disekitarnya. Semakin dekat jarak suatu lokasi dengan lokasi amatan, makin besar kontribusi lokasi tersebut dalam menentukan estimasi

parameter pada lokasi amatan. Besarnya kontribusi dari lokasi-lokasi sekitar ini dituangkan dengan pembobot. Oleh karena itu, penaksiran parameter model GWR menggunakan metode *Weighted Least Squares* (WLS) yaitu dengan memberikan pembobot yang berbeda untuk setiap lokasi dimana data diamati.

Seperti halnya pada model regresi, penaksiran parameter masing-masing lokasi pada model GWR dengan metode WLS adalah dengan meminimumkan jumlah kuadrat error berikut :

$$\sum_{j=1}^n w_j(u_i, v_i) \varepsilon_j^2 = \sum_{j=1}^n w_j(u_i, v_i) [y_j - \beta_0(u_i, v_i) - \sum_{k=1}^k \beta_k(u_i, v_i) x_{ik}]^2 \quad (2.10)$$

Dalam bentuk matrik menjadi

$$\boldsymbol{\varepsilon}^t \mathbf{W}(u_i, v_i) \boldsymbol{\varepsilon} = \mathbf{y}^t \mathbf{W}(u_i, v_i) \mathbf{y} - 2\boldsymbol{\beta}^t(u_i, v_i) \mathbf{X}^t \mathbf{W}(u_i, v_i) \mathbf{y} + \boldsymbol{\beta}^t(u_i, v_i) \mathbf{X}^t \mathbf{W}(u_i, v_i) \mathbf{X} \boldsymbol{\beta}(u_i, v_i)$$

Koefisien regresi akan dtaksir dengan formula berikut:

$$\hat{\boldsymbol{\beta}}(u_i, v_i) = (\mathbf{X}^t \mathbf{W}(u_i, v_i) \mathbf{X})^{-1} \mathbf{X}^t \mathbf{W}(u_i, v_i) \mathbf{y}$$

Dimana

- $\mathbf{W}(u_i, v_i)$  merupakan matrik diagonal berukuran  $n \times n$  yang berisi nilai pembobot spasial setiap lokasi terhadap lokasi ke- $i$ .
- $\mathbf{X}$  merupakan matrik berukuran  $n \times p$  yang berisi 1 pada kolom pertama serta nilai  $K$  variabel penjelas pada  $n$  lokasi.

- $\hat{\boldsymbol{\beta}}(u_i, v_i) = \begin{bmatrix} \hat{\beta}_0(u_i, v_i) \\ \hat{\beta}_1(u_i, v_i) \\ \vdots \\ \hat{\beta}_k(u_i, v_i) \end{bmatrix}$  merupakan vektor taksiran parameter pada

lokasi ke- $i$  yang berukuran  $p \times 1$

Ketika ada sebanyak  $n$  lokasi, maka persamaan matriks diatas adalah penaksi dari setiap baris pada matrik parameter lokal dari seluruh lokasi berikut:

$$\beta(u_i, v_i) = \begin{bmatrix} \beta_0(u_1, v_1) & \beta_1(u_1, v_1) & \beta_2(u_1, v_1) & \cdots & \beta_k(u_1, v_1) \\ \beta_0(u_2, v_2) & \beta_1(u_2, v_2) & \beta_2(u_2, v_2) & \cdots & \beta_k(u_2, v_2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \beta_0(u_n, v_n) & \beta_1(u_n, v_n) & \beta_2(u_n, v_n) & \cdots & \beta_k(u_n, v_n) \end{bmatrix}$$

## 2.6 Robust Geographically Weighted Regression (RGWR)

Data yang mengandung pencilan yang memiliki informasi dan tidak dapat ditunjukkan oleh data yang lain membutuhkan teknik khusus agar informasi yang dimiliki tetap dapat dipertahankan sekaligus model yang dihasilkan juga tidak didominasi oleh pencilan tersebut. Keberadaan pencilan spasial seringkali susah untuk dideteksi pada model GWR. Teknik yang dapat diterapkan yaitu dengan mengestimasi model menggunakan metode yang *robust*, yaitu sebuah metode yang “kuat” dan tidak terlalu terpengaruh oleh keberadaan pencilan. Banyak metode yang *robust* yang dapat digunakan untuk mengatasi pencilan, seperti *M-estimators*, *Least Median of Squares (LMS)*, *Least Absolute Deviation (LAD)* dll. Akan tetapi menurut Wang dan Scott (1994) dalam Zhang dan Mei (2011), LAD merupakan metode yang paling sederhana pada regresi non-prameterik untuk membuat fungsi kernel menjadi lebih *robust*.

Metode LAD mulai diperkenalkan oleh Roger Joseph Boscovich tahun 1757 (Birkes and Dodge, 1993). Metode ini kemudian mulai digunakan 30 tahun kemudian. Akan tetapi, metode ini kehilangan pamornya dikalangan peneliti dibandingkan metode LS yang lebih sederhana proses komputasi pengolahannya. Akan tetapi, masalah komputasi pada saat ini tidak menjadi masalah lagi dalam menerapkan metode LAD. Sekarang, metode LAD sudah diterapkan dalam berbagai bidang penelitian, sebagai alternatif metode untuk menaksir model regresi.



Model yang digunakan pada RGWR sama dengan model yang digunakan pada model GWR. Begitu pula dengan matriks pembobot yang digunakan yang membedakan adalah kriteria yang digunakan pada pemilihan bandwith yang optimum. Kriteria pemilihan bandwith 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 CV. Kriteria ACV menggunakan nilai mutlak yang selisih variabel respon  $y_i$  yang dirumuskan sebagai berikut:

$$ACV(h) = \sum_{i=1}^n |y_i - \hat{y}_{\neq i}(h)| \quad (2.11)$$

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

### 1. Estimasi Parameter dengan *Least Absolute Deviation* (LAD)

Mengestimasi model regresi seperti pada persamaan (2.2) dengan metode LAD dilakukan dengan meminimumkan jumlah dari nilai mutlak residual, meminimumkan

$$\sum_{i=1}^n w_i(u_i v_i) |\varepsilon_i| = \sum_{i=1}^n |y_i - \beta_0(u_i v_i) - \sum_{k=1}^K \beta_k(u_i v_i) X_{ik}| \quad (2.12)$$

Solusi untuk menghasilkan parameter regresi  $(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k)$  tidak dapat dilakukan dengan proses diferensiasi seperti halnya metode LS, karena bentuk persamaan (2.12) di atas merupakan bentuk yang tidak dapat diturunkan pada semua titik. Wagner (1959) menyatakan bahwa solusi  $(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k)$  dapat diperoleh seperti halnya menyelesaikan masalah program linier (Birkes dan Dodge, 1993).

Misal didefinisikan

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

Dan dilakukan dekomposisi terhadap residual sebagai berikut:

$$\begin{aligned} \varepsilon_i &= \varepsilon_i^+ + \varepsilon_i^- \\ |\varepsilon_i| &= \varepsilon_i^+ + \varepsilon_i^- \end{aligned} \quad (2.13)$$

Dengan  $\varepsilon_i^- = |\varepsilon_i| I(\varepsilon_i < 0)$  dan  $\varepsilon_i^+ = |\varepsilon_i| I(\varepsilon_i > 0)$

berdasarkan persamaan (2.12) dan (2.13), prosedur estimasi nilai koefisien regresi dengan metode *least absolute deviation* dapat dibentuk menjadi:

Meminimumkan  $\sum_{i=1}^n (\varepsilon_i^+ + \varepsilon_i^-)$ , maka

$$\min \sum_i^n W_i(u_i v_i) |\varepsilon_i| = \min \left\{ \sum_{i=1, \varepsilon_i \geq 0}^n W_i(u_i v_i) (y_i - \mathbf{X}_i^T \hat{\beta}_i) + \sum_{i=1, \varepsilon_i < 0}^n W_i(u_i v_i) (y_i - \mathbf{X}_i^T \hat{\beta}_i) \right\} \quad 2.14$$

sehingga persamaan 2.14 merupakan solusi dari model RGWR dengan metode least absolute deviation pada lokasi pengamatan ke  $i$ . sehingga estimasi parameter model RGWR dengan metode *least absolute deviation* pada lokasi ke  $i$  dinyatakan dalam bentuk pemrograman linier sebagai berikut:

$$\text{minimalkan } \sum_{i=1}^n (\varepsilon_i^+ + \varepsilon_i^-) w_i(u_i v_i)$$

dengan kendala

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

$\varepsilon_i^+$  dan  $\varepsilon_i^-$  masing-masing merupakan deviasi positif dan negatif pada observasi ke- $i$ .

## 2.7 Ukuran Ketepatan Model

*Mean Absolute Percentage Error* (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):

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

Dengan

$y_t$  : nilai y aktual

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

Semakin kecil nilai MAPE maka model yang didapatkan semakin baik.

