

BAB III
MULTIVARIATE GEOGRAPHICALLY WEIGHTED REGRESSION
(MGWR)

3.1 Variabel Penelitian

Variabel penelitian yang digunakan dalam penelitian ini adalah data derajat kesehatan tahun 2013 pada 27 kabupaten dan kota di Jawa Barat.

3.1.1 Variabel Respon (Y)

Variabel respon (Y) yang digunakan adalah Angka Harapan Hidup (Y_1) dan persentase kurang gizi pada balita (Y_2) pada kabupaten dan kota di Jawa Barat pada tahun 2013. Variabel persentase kurang gizi pada balita diperoleh menggunakan *Microsoft Excel 2010* perhitungan manual dengan membandingkan jumlah balita kurang gizi dan jumlah balita di kabupaten dan kota di Jawa Barat.

3.1.2 Variabel Penjelas (X)

Variabel prediktor yang digunakan dalam penelitian ini adalah 6 variabel. Variabel yang di pilih adalah sebagai berikut:

- X_1 : Angka Melek Huruf
- X_2 : Rata-Rata Lama Sekolah
- X_3 : Indeks Kesejahteraan Rakyat
- X_4 : Persentase Penduduk Dengan Sumber Air Minum Ketempat Penampungan Kotoran > 10 Meter
- X_5 : Persentase Penduduk Dengan Sumber Air Minum Terlindungi PDAM
- X_6 : Pengeluaran Perkapita Penduduk Untuk Makan Sebulan

Definisi Variabel Penelitian

1. Angka Harapan Hidup merupakan rata-rata perkiraan banyak tahun yang dapat ditempuh oleh seseorang semasa masih hidup.
2. Persentase Balita Kekurangan Gizi merupakan persentase balita yang memiliki ukuran antropometri berat badan terhadap tinggi badan berkisar antara 71% s/d 80% .
3. Angka Melek Huruf didefinisikan sebagai proporsi penduduk usia 15 tahun ke atas yang mempunyai kemampuan membaca dan menulis huruf latin dan huruf lainnya, tanpa harus mengerti apa yang dibaca atau ditulisnya.
4. Rata-Rata Lama Sekolah didefinisikan sebagai jumlah tahun belajar penduduk usia 15 tahun keatas yang telah diselesaikan dalam pendidikan formal (tidak termasuk tahun yang mengulang).
5. Indeks Kesejahteraan Rakyat merupakan alat ukur ketersediaan akses terhadap pemenuhan hak-hak dasar rakyat
6. Persentase penduduk dengan sumber air minum terlindungi merupakan persentase sumber air minum keluarga yang bersumber dari sarana air bersih yang telah memenuhi persyaratan, baik kimia, biologis dan fisik. Sumber mata air PDAM, sumur gali, sumur pompa relatif lebih terlindungi dan memenuhi persyaratan. Sedangkan sumber mata air danau, sungai, relatif tidak terlindungi dan tidak memenuhi persyaratan kesehatan. Perhitungan persentase penduduk dengan sumber air minum terlindungi PDAM/pipa dihitung manual menggunakan *Microsoft Excel 2010* dengan perbandingan jumlah penduduk dengan sumber air minum PDAM/Pipa dan jumlah penduduk kabupaten dan kota.
7. Persentase penduduk dengan sumber air minum ketempat penampungan kotoran > 10 meter.
8. Pengeluaran Perkapita Penduduk Sebulan Untuk Makan adalah rata-rata biaya yang dikeluarkan rumah tangga sebulan untuk konsumsi semua anggota rumah tangga dibagi dengan banyaknya anggota rumah tangga. Pengerluaran perkapita untuk makan ditulis dalam ribu rupiah.

3.2 Metode *Multivariate Geographically Weighted Regression* (MGWR)

Model regresi linier global mengasumsikan bahwa nilai penaksiran parameter regresi bersifat umum dan berlaku sama di setiap wilayah yang diteliti. Apabila terjadi perbedaan wilayah, baik dari segi geografis, sosial-budaya maupun hal lain yang melatarbelakanginya, maka akan terjadi heterogenitas spasial yang menyebabkan nonstasioner spasial pada parameter regresi yaitu kondisi dimana parameter regresi bervariasi secara spasial. Sehingga regresi global tidak bisa digunakan sebagai alat analisis. Selain itu apabila menggunakan regresi global akan menyebabkan kesimpulan yang kurang tepat karena asumsi residual saling bebas dan asumsi homogenitas tidak dipenuhi. Metode *Multivariate Geographically Weighted Regression* (MGWR) merupakan pengembangan dari model linier spasial dengan penaksir bersifat lokal untuk setiap lokasi pengamatan.

Pada model *Multivariate Geographically Weighted Regression* (MGWR) asumsi yang digunakan adalah vektor error (ε) berdistribusi normal multivariat dengan mean vektor nol dan matriks varian-kovarian Σ . 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 *Multivariate Geographically Weighted Regression* (MGWR) dapat dinyatakan sebagai berikut:

$$\begin{aligned}
 Y_{i1} &= \beta_{01}(u_i, v_i) + \beta_{11}(u_i, v_i)X_{11} + \dots + \beta_{p1}(u_i, v_i)X_{1p} + \varepsilon_{1q} \\
 Y_{i2} &= \beta_{02}(u_i, v_i) + \beta_{12}(u_i, v_i)X_{12} + \dots + \beta_{p1}(u_i, v_i)X_{2p} + \varepsilon_{2q} \\
 &\vdots \\
 Y_{iq} &= \beta_{0q}(u_i, v_i) + \beta_{1q}(u_i, v_i)X_{11} + \dots + \beta_{pm}(u_i, v_i)X_{ip} + \varepsilon_{iq} \quad (3.1)
 \end{aligned}$$

Dimana

Y_{iq} : nilai amatan ke- i untuk variabel respon ke- q ,

- X_{ip} : nilai amatan ke-i untuk prediktor ke p,
- u_i, v_i : koordinat titik lokasi pengamatan ke-i (*longitude, latitude*)
- $\beta_{0q}(u_i, v_i)$: konstanta/Intercept *Multivariate Geographically Weighted Regression* (MGWR)
- $\beta_{rm}(u_i, v_i)$: parameter model *Multivariate Geographically Weighted Regression* (MGWR)
- ε_{iq} : error pada titik lokasi ke-i untuk prediktor ke-q

3.2.1 Penaksiran Parameter

Penaksiran koefisien regresi pada metode *Multivariate Geographically Weighted Regression* (MGWR) dilakukan dengan menggunakan metode kuadrat terkecil terboboti atau *Weighted Least Square* (WLS), yaitu metode dengan meminimumkan kuadrat errornya, dimana pembobotan yang diberikan berbeda pada setiap titik lokasi pengamatan. Sehingga koefisien regresi pada lokasi pengamatan (u_i, v_i) ditaksir dengan penambahan bobot w_{ij} dan meminimumkan jumlah kuadrat errornya.

$$\sum w_{ij} E = \sum w_{ij} \hat{\varepsilon}^T \hat{\varepsilon} \quad (3.2)$$

$$= \sum w_{ij} (Y - X\hat{B})^T (Y - X\hat{B}) \quad (3.3)$$

$$= \sum w_{ij} (Y^T - 2\hat{B}^T X^T Y + \hat{B}^T X^T X \hat{B}) \quad (3.4)$$

$$= Y^T W(u_i, v_i) Y - 2\hat{\beta}(u_i, v_i) X^T W(u_i, v_i) + \hat{\beta}^T(u_i, v_i) X^T W(u_i, v_i) X \hat{\beta}(u_i, v_i) \quad (3.5)$$

Kuadrat error akan minimum dengan mendiferensialkan persamaan (3.5) terhadap $\hat{\beta}(u_i, v_i)$ sebagai berikut:

$$\frac{\partial}{\partial \hat{\beta}(u_i, v_i)} \sum w_{ij} \hat{\varepsilon}^T \hat{\varepsilon} = 0 - 2X^T W(u_i, v_i)Y + 2X^T W(u_i, v_i)X\hat{\beta}(u_i, v_i) \quad (3.6)$$

Penyelesaian persamaan (3.6) sebagai berikut:

$$0 - 2X^T W(u_i, v_i)Y + 2X^T W(u_i, v_i)X\hat{\beta}(u_i, v_i) = 0 \quad (3.7)$$

$$2X^T W(u_i, v_i)X\hat{\beta}(u_i, v_i) = 2X^T W(u_i, v_i)Y \quad (3.8)$$

$$X^T W(u_i, v_i)X\hat{\beta}(u_i, v_i) = X^T W(u_i, v_i)Y \quad (3.9)$$

Untuk memperoleh taksiran koefisien regresi $\hat{\beta}(u_i, v_i)$, kedua ruas persamaan (3.9) dikalikan dengan $(X^T W(u_i, v_i)X)^{-1}$

Karena $(X^T W(u_i, v_i)X)^{-1}(X^T W(u_i, v_i)X) = I$, maka:

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

atau

$$\hat{\beta}(u_i, v_i) = (X^T W(u_i, v_i)X)^{-1}X^T W(u_i, v_i)Y \quad (3.10)$$

3.2.2 Bandwidth

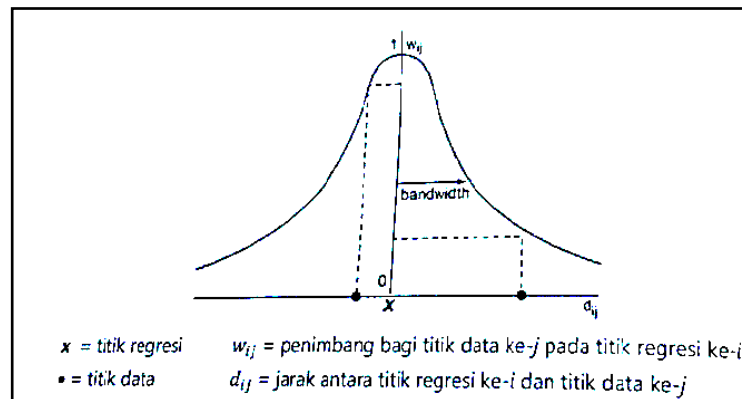
Bandwidth merupakan sebuah pengukuran yang bersifat *deistance-decay* atau nilai penimbang yang menunjukkan seberapa jauhnya jarak antarwilayah. Fungsi dari Bandwith adalah untuk menentukan bobot dari suatu lokasi terhadap lokasi pusat. Bandwith dapat dianalogikan sebagai radius lingkaran (b) sehingga suatu titik lokasi pengamatan yang berada dalam radius lingkaran dianggap berpengaruh dalam membentuk parameter di titik pengamatan lokasi ke-i. Pemilihan *bandwidth* optimum sangat penting karena mempengaruhi ketepatan model pada data.

Apabila nilai *bandwidth* sangat kecil akan menyebabkan penaksiran parameter di lokasi pengamatan ke-i semakin bergantung pada titik lokasi pengamatan lain yang memiliki jarak terdekat dengan lokasi pengamatan ke-i. Sehingga varians yang dihasilkan semakin besar. Apabila nilai *bandwidth* sangat

besar menyebabkan bias yang sangat besar sehingga model yang diperoleh terlalu halus.

Ilustrasi

Gambar 3.1 Ilustrasi *bandwidth* (Badan Pusat Stat)



Terdapat beberapa metode yang dapat digunakan dalam menentukan nilai *bandwidth* optimum, antara lain *Cross Validation* (CV), *Generalised Cross Validation* (GCV), *Akaike Information Criterion* (AIC), dan *Bayesian Information Criterion* (BIC)

Cross Validation (CV) atau metode validasi silang merupakan metode yang digunakan untuk menentukan *bandwidth* optimum. Metode *Cross Validation* dapat ditulis sebagai berikut:

$$CV = \sum_{i=1}^n [y_i - \hat{y}_{i \neq 1}(b)]^2 \quad (3.11)$$

Dengan $\hat{y}_{i \neq 1}$ merupakan nilai penaksir y_i dimana pengamatan pada titik lokasi ke- i dihilangkan dari proses penaksiran. Nilai *bandwidth* optimum diperoleh saat nilai CV minimum (Fotheringham, Brunson, & Charlton, 2002). Metode lain untuk menentukan *bandwidth* optimum adalah *Generalised Cross Validation* (GCV). Metode *Generalised Cross Validation* (GCV) dapat dirumuskan sebagai berikut:

$$GCV = n \sum_{i=1}^n \frac{[y_i - \hat{y}_i(b)]^2}{(n - v_i)^2} \quad (3.12)$$

Dimana v_i merupakan banyak parameter yang terdapat dalam model.

Selain itu metode lain yang dapat digunakan adalah metode *Akaike Information Criterion* (AIC). Perumusan metode Akaike Information Criterion (AIC) adalah sebagai berikut:

$$AIC_c = 2n \log_e(\hat{\sigma}) + n \log_e(2\pi) + n \left\{ \frac{n+tr(S)}{n-2-tr(S)} \right\} \quad (3.13)$$

Dimana n menunjukkan banyak sampel, $(\hat{\sigma})$ merupakan penaksir nilai kekeliruan baku dan $tr(S)$ adalah trace dari *hat matrix*. Semakin kecil nilai AIC maka model taksiran yang dilakukan semakin baik.

Metode lainnya yang dapat digunakan untuk menghitung bandwidth optimal adalah *Bayesian Information Criterion* (BIC). Perumusan untuk metode Bayesian Information Criterion (BIC) adalah sebagai berikut:

$$BIC = -2 \log_e(L) + k \log_e(n) \quad (3.14)$$

Dengan L merupakan nilai fungsi kemungkinan dalam model, k merupakan banyak parameter dalam model dan n merupakan ukuran sampel. (Stat, B. P)

3.2.3 Pembobotan

Fungsi pembobotan lokasi mempunyai peranan yang sangat penting dalam *Geographically Weighted Regression*. Pembobotan pada *Geographically Weighted Regression* 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. Matriks diagonal pembobotan titik pada lokasi pengamatan ke- i dapat ditulis sebagai berikut

$$W(u_i, v_i) = \begin{bmatrix} w_{i1} & 0 & \dots & 0 \\ 0 & w_{i2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_{in} \end{bmatrix} \quad (3.15)$$

Dimana

$W(u_i, v_i)$: matriks diagonal ($n \times n$)

w_{ij} : pembobotan untuk masing masing tiap titik lokasi pengamatan (u_i, v_i)

Titik pengamatan lokasi lain yang memiliki jarak paling dekat dengan titik pengamatan lokasi ke-i diasumsikan memiliki pengaruh paling besar terhadap penaksiran parameter di titik lokasi pengamatan ke-i. Maka apabila jarak antara titik pengamatan semakin dekat, matriks pembobotan $W(u_i, v_i)$ akan semakin besar (Fotheringham, Brunson, & Charlton, 2002).

Terdapat beberapa cara untuk melakukan pembobotan. Cara yang paling sederhana dengan memberikan bobot sebesar 1 untuk setiap titik lokasi pengamatan i dan j sebagai berikut:

$$w_{ij} = 1, \forall i \text{ dan } j \quad (3.16)$$

Pembobotan dengan menggunakan fungsi invers jarak adalah sebagai berikut:

$$W_{ij} = \begin{cases} 1 & , \text{saat } d_{ij} < b \\ 0 & , \text{saat } d_{ij} \geq b \end{cases} \quad (3.17)$$

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

Dimana

d_{ij} : jarak euclidean antara titik lokasi pengamatan ke-i dengan titik lokasi pengamatan ke-j

b : bandwidth atau lebar jendela yang dianalogikan sebagai radius lingkaran

Pembobotan matriks juga dapat ditentukan dengan menggunakan fungsi Kernel. Fungsi pembobotan Kernel memberikan pembobotan yang sesuai dengan nilai bandwidth optimum bergantung pada kondisi data. Menurut Fotheringham et.al (2002) terdapat beberapa fungsi pembobotan yang dapat digunakan pada model *geographically weighted regression*, diantaranya:

1. Fixed kernel

Fungsi Fixed Kernel memiliki *bandwidth* yang sama pada setiap titik lokasi pengamatan. Terdapat dua jenis Fungsi Fixed Kernel yaitu:

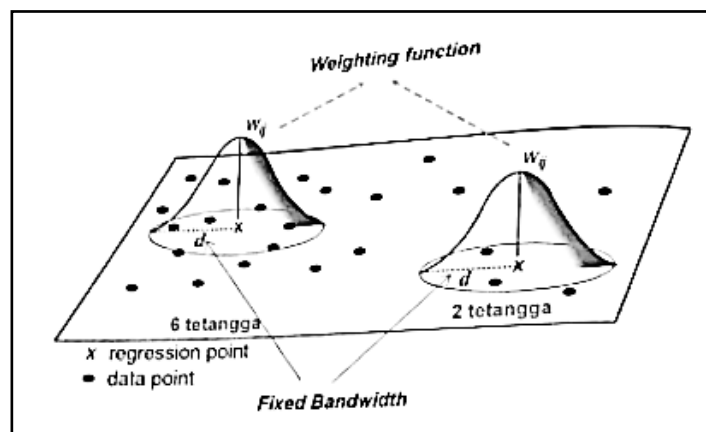
a. Fungsi kernel Gaussian

$$w_{ij} = e^{-0,5\left(\frac{d_{ij}}{b}\right)^2} \quad (3.18)$$

b. Fungsi kernel Bi-Square

$$w_{ij} = \begin{cases} \left[1 - \left(\frac{d_{ij}}{b}\right)^2\right]^2, & \text{jika } d_{ij} < b \\ 0 & \text{jika } d_{ij} \geq b \end{cases} \quad (3.19)$$

Gambar 3.2 Ilustrasi Fixed Bandwidth (Badan Pusat Stat)



2. Adaptive Kernel

Fungsi Adaptive Kernel memiliki *bandwidth* yang berbeda pada setiap lokasi pengamatan. Terdapat dua jenis fungsi Adaptive Kernel, yaitu:

a. Fungsi Kernel Adaptive Gaussian

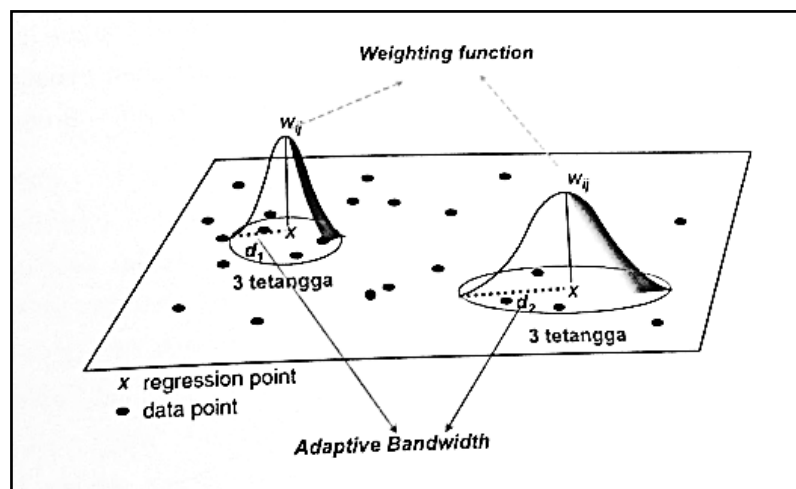
$$w_{ij} = e^{-0,5\left(\frac{d_{ij}}{b_{i(q)}}\right)^2} \quad (3.19)$$

b. Fungsi Kernel Adaptive Bi-square

$$w_{ij} = \begin{cases} \left[1 - \left(\frac{d_{ij}}{b_{i(q)}} \right)^2 \right]^2, & \text{jika } d_{ij} < b \\ 0 & \text{jika } d_{ij} \geq b \end{cases} \quad (3.20)$$

Dengan $b_{i(q)}$ merupakan bandwidth adaptif yang menetapkan q sebagai jarak lokasi pengamatan lain terdekat dari titik kolasi pengamatan ke- i .

Gambar 3.3 Ilustrasi Adaptive Bandwidth (badan Pusat Stat)



3.2.4 Pengujian Hipotesis Kesesuaian Model MGWR

Pengujian keberartian model MGWR dilakukan untuk menentukan model MGWR lebih baik secara signifikan dalam memodelkan data. Perumusan hipotesisnya adalah sebagai berikut:

$$H_0: \beta_{kh}(u_i, v_i) = \beta_k \text{ untuk setiap } k = 1, 2, \dots, p \text{ dan } i = 1, 2, \dots, n$$

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

Atau

$$H_0: \text{tidak ada pengaruh faktor geografis pada model}$$

$$H_1: \text{ada pengaruh faktor geografis pada model}$$

Statistik uji

$$F_{hitung} = \frac{|(Y^T(I-M)Y) - (Y^T(I-S)Y)|}{\frac{n-p-r_1}{|Y^T(I-S)^T(I-S)Y|} \left(\frac{r_1^2}{r_2}\right)} \quad (3.21)$$

Dimana

$$M = X(X^T X)^{-1} X^T$$

$$S_{(n \times n)} = \begin{bmatrix} X_1^T (X^T W(u_1, v_1) X)^{-1} X^T W(u_1, v_1) \\ X_2^T (X^T W(u_2, v_2) X)^{-1} X^T W(u_2, v_2) \\ \vdots \\ X_n^T (X^T W(u_n, v_n) X)^{-1} X^T W(u_n, v_n) \end{bmatrix}$$

$$r_h = \text{tr}([(I-S)(I-S)]^i) ; h: 1,2$$

Kriteria pengujian, tolak H_0 jika $F_{hitung} > F\left(\alpha, n-p-r_1, \left(\frac{r_1^2}{r_2}\right)\right)$

3.2.5 Pengujian Serentak Model MGWR

Setelah didapatkan hasil pengujian hipotesis kemudian dilakukan uji serentak pada model MGWR dengan hipotesis sebagai berikut:

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

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

Statisik uji

$$F = \frac{\frac{|Y^T(I-S_\omega)^T(I-S_\omega)Y|}{\left(\frac{r_{1\omega}^2}{r_{2\omega}}\right)}}{\frac{|Y^T(I-S)^T(I-S)Y|}{\left(\frac{r_1^2}{r_2}\right)}} \quad (3.22)$$

Kriteria penujian, tolak H_0 jika $F > F\left(\alpha, \frac{r_{1\omega}^2}{r_{2\omega}}, \frac{r_1^2}{r_2}\right)$

3.2.6 Pengujian Keberartian Parameter MGWR

Pengujian parameter model MGWR dilakukan untuk mengetahui faktor-faktor yang mempengaruhi kasus derajat kesehatan di Jawa barat. Perumusan hipotesis yang digunakan sebagai berikut:

$$H_0: \beta_{kh}(u_i, v_i) = 0 ; k = 1, 2, \dots, p; h = 1, 2, \dots, q; i = 1, 2, \dots, n$$

$$H_1: \beta_{kh}(u_i, v_i) \neq 0 ; k = 1, 2, \dots, p; h = 1, 2, \dots, q; i = 1, 2, \dots, n$$

Statistik uji

$$t = \frac{\hat{\beta}_{kh}(u_i, v_i)}{SE(\hat{\beta}_{kh}(u_i, v_i))} \quad (3.23)$$

Dimana $SE(\hat{\beta}_{kh}(u_i, v_i))$ merupakan akar dari varians $\hat{\beta}_{kh}(u_i, v_i)$

Kriteria pengujian, tolak H_0 jika $|t| > t_{\left(\frac{\alpha}{2}, (n-p-1)\right)}$

Dengan $n-p-1$ merupakan derajat kebebasan Tabel *Distribusi t-Student*.

3.3 Langkah-langkah Analisis

Adapun langkah-langkah analisis yang dilakukan untuk menentukan faktor-faktor yang mempengaruhi kasus kesehatan di Jawa Barat dengan menggunakan *Multivariate Geographically Weighted Regression* (MGWR) adalah:

1. Mendeskripsikan variabel respon dan variabel-variabel prediktor kasus kesehatan di Jawa Barat.
2. Menganalisis model *Multivariate Geographically Weighted Regression* (MGWR) untuk kasus kesehatan di Jawa Barat. Dengan langkah langkah sebagai berikut:
 - a. Menentukan koordinat *longitude latitude* tiap kabupaten dan kota di Jawa Barat
 - b. Menghitung jarak *euclidean* antar kabupaten dan kota di Jawa Barat.
 - c. Menentukan bandwidth berdasarkan kriteria CV minimum.

- d. Menghitung matriks pembobotan tiap kabupaten dan kota di Jawa Barat dengan fungsi pembobotan fixed kernel Bi-square
 - e. Menaksir parameter *Multivariate Geographically Weighted Regression* (MGWR) dengan menggunakan *bandwidth* optimum
3. Menginterpretasi dan menyimpulkan hasil yang diperoleh.