

# Geographically Weighted Regression (GWR) pada Data Jumlah Penderita Penyakit AIDS

Khalilah Nurfadilah

Univesitas Islam Negeri Alauddin, Khalilah.nurfadilah@uin-alauddin.ac.id

**ABSTRAK,** *Acquired Immune Deficiency Syndrome* (AIDS) merupakan salah satu penyakit yang menjadi penyebab kematian tertinggi di seluruh dunia. Oleh karenanya perlu dilakukan pendekatan model terhadap faktor-faktor penyebabnya. Pada artikel ini digunakan pendekatan regresi linear dan pendekatan GWR. Dari kedua pendekatan ini dipilih model yang representative berdasarkan nilai AIC dan SSE. Berdasarkan hasil analisis, diperoleh kesimpulan bahwa model GWR lebih representative dibandingkan model regresi linear. Hal ini ditunjukkan dengan nilai AIC dan SSE GWR yang lebih kecil dibandingkan dengan model regresi linear.

**Kata Kunci:** AIDS, GWR, Regresi linear

## 1. PENDAHULUAN

*Acquired Immune Deficiency Syndrome* (AIDS) merupakan salah satu penyakit yang menjadi penyebab kematian tertinggi di seluruh dunia, termasuk Indonesia. Data dari kementerian kesehatan RI menunjukkan bahwa pada tahun 2016, dari 7000 penderita AIDS, sebanyak 800 orang mengalami kematian. Artinya, lebih dari 10% penderita mengalami kematian pada tahun tersebut. Identifikasi dan antisipasi dilakukan oleh pemerintah untuk mencegah peningkatan jumlah kematian yang disebabkan oleh AIDS.

Dalam Statistika, identifikasi terhadap penyebab dapat dilakukan dengan memodelkan fenomena jumlah penderita AIDS ke dalam model regresi. Namun, untuk kasus jumlah penderita AIDS di setiap kabupaten/kota, memiliki efek kewilayahan yang bervariasi untuk setiap kabupaten/kota tersebut. Efek ini disebut juga sebagai efek spasial. Efek spasial yang terdapat dalam data kewilayahan terdapat dua, yaitu ketergantungan spasial dan keragaman spasial. Pada kasus jumlah penderita AIDS di setiap kabupaten/kota, maka efek spasial yang digunakan adalah *Geographically Weighed*

*Regression* (GWR), yaitu model regresi secara lokal.

Berbeda dengan model regresi biasa, atau regresi global, GWR menggambarkan model regresi lokal untuk setiap wilayahnya, sehingga akan terdapat banyak parameter taksiran yang dihasilkan.

Pada penelitian ini akan dilakukan pemodelan jumlah penderita AIDS dan faktor-faktor yang mempengaruhinya dengan menggunakan dua pendekatan, yaitu regresi linear dan GWR. Faktor yang dimodelkan adalah jumlah Puskesmas, persentase pengguna KB aktif, dan persentase rumah tangga berperilaku hidup bersih dan sehat. Dari analisis ini akan diketahui model yang representative terhadap jumlah penderita AIDS berdasarkan nilai AIC dan SSE.

## 2. TINJAUAN PUSTAKA

### Regresi Linear

Metode regresi linier merupakan metode yang memodelkan hubungan antara variabel respon  $y$  dan variabel prediktor  $x_1, x_2, \dots, x_p$ . Model regresi linier secara umum :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon$$

dimana  $\beta_0, \beta_1, \dots, \beta_p$  merupakan parameter model dan  $\varepsilon_i$  merupakan sisaan yang diasumsikan identik, independen, dan berdistribusi normal dengan mean nol dan varians konstan  $\sigma^2$  atau ( $\varepsilon_i \sim IIDN(0, \sigma^2)$ ). Pada model ini, hubungan antara variabel prediktor dan variabel respon dianggap konstan pada setiap lokasi geografis.

Dalam notasi matriks, model regresi dapat dituliskan menjadi :  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$  dengan :

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix},$$

$$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ M \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

**Model Geographically Weighted Regression (GWR)**

Model *Geographically Weighted Regression* (GWR) adalah pengembangan dari model regresi dimana setiap parameter dihitung pada setiap titik lokasi, sehingga setiap titik lokasi geografis mempunyai nilai parameter regresi yang berbeda-beda. Model GWR merupakan pengembangan dari model regresi global dimana ide dasarnya diambil dari regresi non parametrik. Variabel respon  $y$  dalam model GWR diprediksi dengan variabel prediktor yang masing-masing koefisien regresinya bergantung pada lokasi dimana data tersebut diamati. Model GWR dapat ditulis sebagai berikut ini :

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i)x_{ik} + \varepsilon_i$$

dengan :

- $y_i$  : Nilai observasi variabel respon untuk lokasi ke- $i$
- $(u_i, v_i)$  : Menyatakan titik koordinat (*longitude, latitude*) lokasi  $i$
- $\beta_k(u_i, v_i)$  : Koefisien regresi variabel prediktor ke- $k$  untuk lokasi ke- $i$

**Bentuk Estimasi Parameter model GWR**  
 bentuk dari estimasi parameter model untuk GWR dinyatakan sebagai berikut :

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

Denngan  $W(u_i, v_i)$  merupakan matriks pembobot.

Peran pembobot pada model GWR sangat penting karena nilai pembobot ini mewakili letak data observasi satu dengan lainnya. Fungsi kernel digunakan untuk mengestimasi paramater dalam model GWR jika fungsi jarak ( $w_j$ ) adalah fungsi yang kontinu dan monoton turun. Pembobot yang terbentuk dengan menggunakan fungsi kernel ini adalah fungsi jarak *Gaussian (Gaussian Distance Function)*, fungsi *Exponential*, fungsi *Bisquare*, dan fungsi kernel *Tricube*. Fungsi pembobot yang digunakan pada penelitian ini adalah fungsi jarak *Gaussian* :

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

dengan  $d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2}$  adalah jarak *eucliden* antara lokasi  $(u_i, v_i)$  ke lokasi  $(u_j, v_j)$  dan  $h$  adalah parameter non negatif yang diketahui dan biasanya disebut parameter penghalus (*bandwidth*).

*Bandwidth* dapat dianalogikan sebagai radius dari suatu lingkaran, sehingga sebuah titik yang berada di dalam radius lingkaran masih dianggap memiliki pengaruh. Di dalam pembentukan sebuah model GWR, *bandwidth* berperan sangat penting karena akan berpengaruh pada ketepatan model terhadap data, yaitu mengatur varians dan bias dari model.

Ada beberapa metode yang digunakan untuk memilih *bandwidth* optimum, salah satu diantaranya adalah metode *Cross Validation* atau CV yang secara matematis didefinisikan sebagai berikut:

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

dengan  $\hat{y}_{\neq i}(h)$  adalah nilai penaksir  $y_i$  dimana pengamatan di lokasi  $(u_i, v_i)$  dihilangkan dari proses estimasi. Untuk mendapatkan nilai  $h$  yang optimal maka diperoleh dari  $h$  yang menghasilkan nilai CV yang minimum.

**3. METODOLOGI**

**Variabel Penelitian**

Data yang digunakan pada penelitian ini bersumber dari publikasi Kementerian Kesehatan RI pada tahun 2016. Data tersebut menunjukkan jumlah penderita AIDS pada tahun 2015 pada setiap kabupaten/kota di Provinsi Sulawesi Selatan. Adapun variable yang berkaitan dalam pemodelan ini adalah sebagai berikut :

Variable	Definisi Operasional Variabel
Y	Jumlah penderita AIDS
X1	Persentase penduduk pengguna KB aktif
X2	Persentase Rumah Tangga Berperilaku Hidup Bersih dan Sehat (PHBS)
X3	Jumlah Puskesmas

**Tahapan Analisis**

Tahapan analisis yang digunakan dalam memodelkan jumlah penderita AIDS dengan model GWR adalah sebagai berikut :

1. Mendeskripsikan tentang faktor-faktor yang mempengaruhi jumlah penderita AIDS untuk setiap kabupaten/kota
2. Pengujian pengaruh spasial dengan menggunakan uji breusch-pagan (BP).
3. Menduga parameter model regresi global dengan menggunakan OLS.
4. Menentukan bandwidth (h) optimum untuk semua lokasi pengamatan dengan menggunakan CV.
5. Pemilihan model terbaik antara model regresi global dan model GWR
6. Menginterpretasi dan menyimpulkan hasil yang diperoleh.

**4. PEMBAHASAN**

**Analisis Deskriptif**

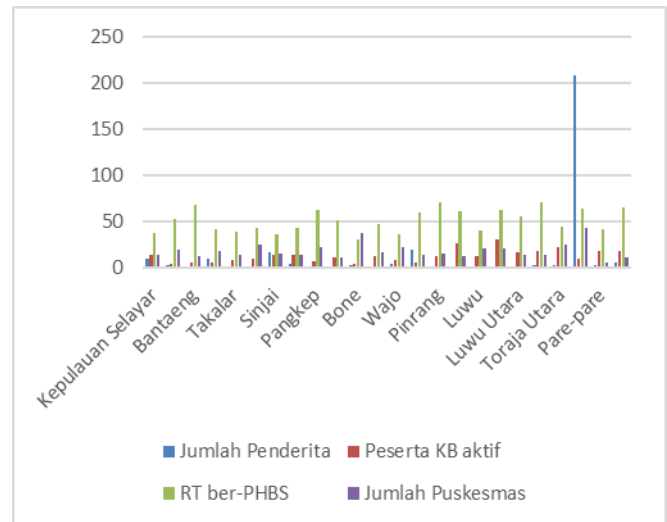
Gambaran umum jumlah penderita AIDS serta faktor-faktor yang mempengaruhi.

Table 1. Nilai minimum dan maksimum untuk semua variabel

Variable	Min	Maks
Jumlah penderita	0	208
Persentase penduduk pengguna KB aktif	4.92	31.72
Persentase Rumah tangga berperilaku hidup bersih dan sehat	31.22	71.27
<b>Jumlah Puskesmas</b>	6	43

Berdasarkan Table 1 di atas, terlihat bahwa untuk jumlah penderita AIDS, yang terendah adalah 0, atau tidak ada pendeita, sedangkan yang tertinggi adalah 208 orang. Untuk persentase penduduk pengguna KB aktif, nilai minimum adalah sebesar 4.92% persen penduduk, sedangkan maksimum sebesar 31.72%. Persentase rumah tangga berperilaku hidup bersih dan sehat, nilai minimum adalah 31.22% dan maksimum 71.27%. Adapun secara keseluruhan perbandingan nilai-nilai dari semua variable dapat dilihat dari grafik berikut :

Gambar 1. Grafik nilai jumlah penderita AIDS, peserta KB aktif, RT ber-PHBS dan jumlah puskesmas di setiap kabupaten/kota



**Model Regresi Linear dan GWR**

**Model Regresi Linear**

Estimasi parameter model regresi linear ditampilkan pada table berikut :

Tabel 2. Hasil estimasi parameter model regresi linear

VARIABEL	KOEFISIEN
<b>INTERSEP</b>	-96.6813
<b>X1</b>	-0.0735
<b>X2</b>	0.9479
<b>X3</b>	3.3072

Berdasarkan hasil estimasi parameter tersebut, maka model regresi linear yang terbentuk adalah :

$$\hat{y} = -96.6813 - 0.0735 X_1 + 0.9479X_2 + 3.3072X_3$$

**Uji Breusch Pagan**

Setelah dilakukan pemodelan dengan regresi linear, selanjutnya dilakukan uji asumsi klasik. Salah satunya adalah uji asumsi kehomogenan ragam. Uji yang digunakan adalah uji Breusch Pagan.

Table 3. Hasil Uji Breusch Pagan

Studentized Breusch-Pagan test BP = 13.945, df = 3, p-value = 0.002981
---

Berdasarkan hasil dari Tabel 2, dengan menggunakan  $\alpha = 0.05$ , terlihat bahwa uji Breusch Pagan signifikan pada taraf tersebut. Dengan demikian disimpulkan bahwa sisaan pada model regresi jumlah penderita AIDS terjadi keragaman, atau ada pelanggaran

homoskedastisitas, maka disimpulkan untuk melanjutkan analisis lanjutan dengan model GWR.

### Menentukan Bandwith Optimum

Table nilai Bandwith dan CV ditampilkan sebagai berikut :

BANDWITH	CV
1.643331	60156.96
2.656311	62634.77
1.017275	58908.62
0.6303512	59148.9
0.9440911	59079.67
1.256407	58952.58
1.125455**	58811.61**
1.125703	58811.62
1.122498	58811.53
1.12245	58811.53
1.12241	58811.53

Berdasarkan nilai CV terendah, maka diperoleh bandwith optimum adalah 1.12245

### Model GWR

Pada pengujian kehomogenan ragam, terlihat bahwa terjadi pelanggaran terhadap asumsi tersebut. Maka, langkah selanjutnya adalah melakukan pemodelan dengan GWR. Adapun hasil estimasi parameter dari model GWR adalah sebagai berikut :

Table 4. Hasil estimasi parameter model GWR

VARIABLE	KOEFSIEN	
	Minimum	Maksimum
INTERSEP	-139.896377	-20.065139
X1	-0.571512	2.675308
X2	0.411093	0.810727
X3	0.432428	5.482230

### Uji kebaikan model

Setelah melakukan pemodelan dengan dua pendekatan, selanjutnya adalah menguji kebaikan model. Statistic yang digunakan adalah nilai AIC dan nilai Sum Square dari kedua model. Nilai tersebut ditunjukkan pada table berikut :

Table 5. nilai AIC dan Sum Square

MODEL	AIC	SUM SQUARE SISAANR (SSE)
REGRESI LINEAR	318.83	22118.45
GWR	221.0196	10529.36

Berdasarkan Tabel 5 di atas, terlihat bahwa untuk AIC dan Sum Square Sisaan (SSE) terendah ada pada model GWR. Dengan demikian disimpulkan bahwa model GWR lebih baik dibandingkan model regresi linear biasa.

## 5. KESIMPULAN

Dari hasil analisis yang telah dilakukan, telah dilakukan dua pendekatan untuk data jumlah penderita AIDS di setiap kabupaten/kota di provinsi Sulawesi Selatan. Hasilnya diperoleh bahwa berdasarkan hasil uji kebaikan model, baik dengan nilai AIC maupun Sum Square Error (SSE), diketahui bahwa model terbaik adalah model GWR. Dengan nilai AIC dan SSE secara berturut-turut adalah 221.0196 dan 10529.36.

## 6. DAFTAR PUSTAKA

- [1] Agresti A. 2002. *Categorical Data Analysis*. New York : John Willey and Sons.
- [2] Anselin L. 1988. *Spatial Econometrics: Methods and Models*. Dordrecht : Kluwer Academic Publishers.
- [3] Anselin L, Getis A. 1992. Spatial Statistical Analysis and Geographic Information Systems. *The Annals of Regional Science* 26(1): 1992.
- [4] Fotheringham AS, Brunson C, Charlton M. 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. England: John Wiley & Sons, Ltd., West Sussex.
- [5] Nakaya T, Fotheringham AS, Brunson C, Charlton M. 2005. Geographically Weighted Poisson Regression for Disease Association Mapping. *Statistics in Medicine*, Volume 24 Issue 17, pages 2695-2717.