

## PERBANDINGAN REGRESI GLOBAL DAN GEOGRAPHICAL WEIGHTED REGRESSION (GWR) PADA MODEL KASUS PREVALENSI PENYAKIT HEPATITIS

<sup>1</sup>Sugiarto, <sup>2</sup>Haiban Hajjid Arsyadana

<sup>1,2</sup> Sekolah Tinggi Ilmu Statistik, Jakarta

Alamat e-mail : <sup>1</sup>soegie@stis.ac.id

### ABSTRAK

Model OLS akan reliabel untuk semua observasi jika tidak ada *spatial non stationarity* (Heterogenitas Spasial). *Spatial non stationarity* bisa dilihat dari terlanggarnya asumsi homoskedastisitas atau non autokorelasi dari error pada model OLS. Untuk mengatasi *nonstationarity* model *Geographically Weighted Regression* (GWR) dapat digunakan untuk mengestimasi data yang memiliki heterogenitas spasial. Tujuan dari artikel ini adalah mendapatkan model analisis statistik yang sesuai untuk model variabel yang memengaruhi prevalensi hepatitis di Indonesia. Hasil penelitian menunjukkan Model analisis terbaik untuk variabel yang memengaruhi prevalensi hepatitis di Indonesia adalah model GWR. Dengan terbentuk 12 kelompok provinsi berdasarkan variabel prediktor yang memengaruhi.

**Kata Kunci** : *Ordinary Least Square, Regresi Global, Geographically Weighted Regression*(GWR), Hepatitis.

### PENDAHULUAN

Metode *Ordinary Least Square*(OLS) merupakan suatu metode yang digunakan untuk menyatakan pola hubungan antara satu variabel respon dan satu atau lebih variabel prediktor. Jika kurva regresi mengikuti pola linear maka model tersebut disebut sebagai model regresi linear [1]. Persamaan umum model regresi linier untuk  $n$  observasi ( $i = 1, 2, \dots, n$ ) dan  $p$  independen, dapat ditulis sebagai berikut:

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

dimana  $y_i$  adalah nilai variabel dependen pada pengamatan  $k$ -i,  $x_{ik}$  adalah nilai variabel dependen ke- $k$  pada pengamatan ke- $i$ ,  $\beta_0$  adalah konstanta/*intercept*,  $\beta_k$  adalah nilai fungsi variabel prediktor  $x_k$  pada pengamatan ke- $i$ , dan  $\varepsilon_i$  adalah

random error yang diasumsikan berdistribusi  $N(0, \sigma^2)$ . Model dengan menggunakan metode OLS mensyaratkan terpenuhinya asumsi klasik yaitu: berdistribusi normal, tidak ada multikolinearitas, tidak ada autokorelasi dan data bersifat homogen.

Model OLS atau model regresi linier yang umum digunakan dalam penelitian merupakan *global model* dimana estimasi parameternya bernilai sama pada semua lokasi penelitian atau hubungan antara variabel prediktor dan variabel respon dianggap stasioner di semua lokasi penelitian. Keadaan tersebut disebut homogenitas spasial. Jika parameter regresi yang diestimasi bervariasi pada tiap-tiap wilayah penelitian atau terdapat efek spasial akan terlihat dalam nilai signifikansi parameter dan residual dari model yang dibentuk, sehingga mengakibatkan estimasi parameter

menggunakan regresi klasik menjadi tidak efisien, yang artinya *standard error* dari parameter menjadi besar (heterogen) dan estimasi parameter tidak signifikan.

Model OLS akan *reliable* untuk semua observasi jika tidak ada *spatial non stationarity*. *Spatial non stationarity* bisa dilihat dari terlanggarnya asumsi homoskedastisitas atau nonautokorelasi dari *error* pada model OLS. Apabila terjadi heterogenitas spasial/*spatial nonstationarity* pada parameter regresi, maka informasi yang tidak dapat ditangani oleh metode OLS akan ditampung sebagai galat. Bila kasus semacam itu terjadi, regresi OLS menjadi kurang mampu dalam menjelaskan fenomena data yang sebenarnya. Maka dari itu untuk mengatasi *nonstationarity* model *Geographically Weighted Regression* (GWR) dapat digunakan untuk mengestimasi data yang memiliki heterogenitas spasial [1].

Model GWR merupakan pengembangan dari regresi global. Namun, berbeda dengan regresi global yang nilai parameter modelnya konstan, parameter model GWR berbeda-beda pada setiap lokasi penelitian. Parameter pada model GWR dihitung pada setiap lokasi pengamatan, sehingga setiap lokasi pengamatan mempunyai nilai parameter regresi yang berbeda-beda, sehingga akan terdapat sebanyak  $n(p+1)$  parameter yang diestimasi pada model GWR, dimana  $n$  adalah jumlah lokasi pengamatan dan  $(p+1)$  adalah jumlah parameter. Pada model GWR, masing-masing lokasi pengamatan bergeoreferensi, yaitu memiliki koordinat spasial. Koordinat spasial pada lokasi pengamatan ke- $i$  dilambangkan dengan  $(u_i, v_i)$ . Adapun model GWR dapat dituliskan sebagai berikut [1]:

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

$y_i$  : Nilai observasi variabel respon pada pengamatan ke- $i$

$\beta_0(u_i, v_i)$  : Nilai konstanta/intercept pada pengamatan ke- $i$

$\beta_k(u_i, v_i)$  : Koefisien regresi variabel prediktor ke- $k$  pada pengamatan ke- $i$

$x_{ik}$  : Nilai observasi variabel prediktor ke- $k$  pada pengamatan ke- $i$

$\varepsilon_i$  : *Error* ke- $i$  yang diasumsikan identik, independen, dan berdistribusi normal dengan rata-rata nol dan varian konstan  $\sigma^2$

$(u_i, v_i)$  : Menyatakan titik koordinat (*longitude, latitude*) lokasi pengamatan ke- $i$

Ada dua asumsi yang digunakan dalam model GWR [1]:

1. *Error term* adalah independen, identik, dan mengikuti distribusi normal dengan rata-rata nol dan varians konstan  $[(\varepsilon_i \sim \text{IIDN}(0, \sigma^2))]$
2.  $E(\hat{y}_i) = E(y_i)$  untuk semua  $i$

Asumsi ke-2 biasanya tidak selalu tepat untuk penaksir parameter lokal linier, kecuali hubungan linier antara variabel independen dan dependen berlaku secara menyeluruh (global)[2]. Metode GWR tidak mungkin menghasilkan estimasi parameter lokal yang unbiased, tetapi bisa meminimalisir estimasi yang bias [1].

Salah satu kasus data yang memiliki variasi berbeda-beda pada tiap daerah adalah data prevalensi penyebaran penyakit hepatitis. Menurut data WHO angka prevalensi penyakit hepatitis amat bervariasi dalam distribusi secara geografis. Letak geografis juga berpengaruh terhadap penyebaran penyakit ini. Biasanya daerah-daerah yang memiliki letak berdekatan akan memiliki prevalensi yang hampir sama, dibandingkan daerah-daerah yang letaknya lebih jauh.

Penyakit hepatitis merupakan penyakit menular dimana WHO (2010)

menyatakan penyakit hepatitis telah menjadi ancaman kesehatan di dunia. Infeksi virus hepatitis ini telah membuat banyak orang di dunia menjadi korban. Di Asia Pasifik angka kematian akibat infeksi virus Hepatitis tiga kali lebih tinggi dibandingkan HIV-AIDS dan sembilan kali lebih tinggi dibandingkan malaria [3]. Kementerian Kesehatan Indonesia periode 2013 melaporkan bahwa kasus infeksi virus Hepatitis di Indonesia sendiri terjadi peningkatan prevalensi Hepatitis pada semua umur dari 0,6 persen tahun 2007 menjadi 1,2 persen tahun 2013. Lima provinsi dengan prevalensi Hepatitis tertinggi adalah Nusa Tenggara Timur (4,3%), Papua (2,9%), Sulawesi Selatan (2,5%), Sulawesi Tengah (2,3%) dan Maluku (2,3%) [3]. Di Indonesia prevalensi infeksi virus Hepatitis ditemukan sangat bervariasi [4]. Variabel-variabel yang mempengaruhi prevalensi penyakit hepatitis akan berbeda pada masing-masing daerah, sehingga apabila dibuat model regresi linier berganda variabel yang memengaruhi prevalensi penyakit hepatitis di Indonesia akan menjadi kurang reliabel karena pengaruh variasi daerah dan efek spasial.

Dalam penelitian ini akan mencari model analisis statistik yang sesuai untuk model variabel yang memengaruhi prevalensi hepatitis di Indonesia tahun 2013, apakah model global ataukah model GWR. Selain itu apakah ada variasi spasial variabel prediktor di masing-masing provinsi dalam pengolahan data, peneliti akan menggunakan beberapa software statistik diantaranya SPSS versi 19, software geoda, software GWR4 dan ArcGIS. Untuk melihat model mana yang terbaik akan dilihat dari nilai *Akaike Information Criterion* (AIC), *Residual sum of square* (RSS), *Bayesian Information Criterion* (BIC) yang nilainya terkecil dan nilai *R-Square* yang terbesar.

## METODE PENELITIAN

### Sumber Data dan Variabel Penelitian

Penelitian ini menggunakan informasi yang dikumpulkan dari Riset Kesehatan Dasar tahun 2013 (RISKESDAS 2013) dimana survei ini dilaksanakan pada bulan Mei-Juni 2013 dan informasi yang dikumpulkan mengacu pada kejadian selama periode 2010-2013. Variabel-variabel yang digunakan dalam penelitian ini diambil dari kajian teori dari penelitian sebelumnya dan juga mengadopsi model proses persebaran penyakit dari model John Gordon. Adapun variabel yang digunakan antara lain:

HEPATITIS (Y) =prevalensi penduduk yang terpapar Hepatitis di provinsi -i  
 PENDIDIKAN ( $X_1$ ) = persentasi penduduk yang memiliki pendidikan terakhir belum pernah sekolah, tidak tamat SD/MI, tamat SD/MI di provinsi ke-i

UMUR ( $X_2$ ) =persentasi penduduk yang memiliki umur 45 tahun ke atas di provinsi ke-i

SANITASI ( $X_3$ ) =persentasi rumah tangga yang memiliki sanitasi yang tidak layak di provinsi ke-i

KEPADATANRT ( $X_4$ ) =persentasi rumah tangga yang kepadatan huniannya  $< 8 \text{ m}^2/\text{orang}$  di provinsi ke-i

SAMPAH ( $X_5$ ) =persentasi rumah tangga yang membuang sampah sembarangan di provinsi ke-i

AIRNONLAYAK ( $X_6$ ) =persentasi rumah tangga yang menggunakan sumber air tidak baik (berwarna, berbau, berasa, berbusa,keruh) di provinsi ke-i

### Metode Analisis

Tahapan penelitian yang dilakukan dalam mendapatkan pola hubungan terbaik dalam mencari variabel-variabel

yang memengaruhi privalensi penyakit hepatitis di Indonesia adalah sebagai berikut:

- a. Melakukan pemilihan peubah penjelas dan peubah respon yang digunakan dalam analisis berdasarkan studi literatur sebelumnya dan menduga parameter model regresi klasik menggunakan analisis regresi klasik berganda.
- b. Melakukan uji asumsi pada regresi klasik yaitu asumsi kenormalan menggunakan Uji Jarque-Bera, asumsi multikolinieritas dengan melihat nilai *Variance Inflation Factor* (VIF), asumsi autokolerasi dengan Uji Durbin-Watson, serta uji kehomogenan data dengan Uji Breusch-Pagan. Jika asumsi klasik terpenuhi maka regresi global yang terbaik, sedangkan jika tidak maka lanjut ke langkah berikutnya. Jika ada asumsi yang tidak terpenuhi maka akan dilakukan penanganan pada asumsi tersebut, seperti melakukan transformasi fungsi pada asumsi kenormalan dan melakukan seleksi peubah pada asumsi multikolinieritas.
- c. Memeriksa ketergantungan/efek spasial pada data menggunakan uji Indeks Moran [5].

$$I = \frac{n}{\sum_i \sum_j W_{ij}} \frac{\sum_i \sum_j W_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_i (y_i - \bar{y})^2} \quad (3)$$

- d. Menguji efek keragaman spasial dengan menggunakan uji *Breusch-Pagan* dengan hipotesis  $H_0 : \sigma^2(u_i, v_i) = \dots = \sigma^2(u_n, v_n) = \sigma^2$  (Data homogen) dan  $H_1 : \text{minimal ada satu } \sigma^2(u_i, v_i) \neq \sigma^2(u_j, v_j) \text{ untuk } i \neq j$ , dengan  $i, j = 1, 2, \dots, n$  atau (Data Heterogen). Statistik uji Breusch-Pagan sebagai berikut:

$$BP = \frac{1}{2} h' Z (Z' Z)^{-1} Z' h \text{ dimana } h = \left( \frac{e_i^2}{\sigma^2} - 1 \right) \quad (4)$$

Z adalah vektor amatan peubah respon  $y$  yang berukuran  $(n \times 1)$  dan sudah dibakukan untuk setiap

pengamatan dengan  $i = 1, 2, \dots, n$ ,  $e_i^2$  adalah kuadrat galat untuk pengamatan ke- $i$  dan  $\sigma^2$  merupakan ragam dari  $e_i$ . Jika tidak tolak  $H_0$  maka digunakan model regresi spasial, jika tolak  $H_0$  digunakan model GWR dan lanjut ke langkah berikutnya

- e. Menentukan nilai lebar jendela optimum dengan melihat Cross Validation (CV) yang minimum. Formula yang digunakan untuk menghitung CV adalah sebagai berikut:

$$CV = \sum_{i=1}^n [y_i - \hat{y}_{\neq i}(b)]^2 \quad (5)$$

Dimana  $\hat{y}_{\neq i}(b)$  merupakan nilai prediksi  $y_i$  dengan pengamatan ke- $i$  dikeluarkan dari proses estimasi.

- f. Menghitung matriks jarak dan matriks pembobot  $W(u_i, v_i)$  dengan fungsi kernel normal. Adapun formulanya sebagai berikut:

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

Dengan  $h_i(q)$  adalah lebar jendela adaptif yang menetapkan  $q$  sebagai jarak tetangga terdekat dari wilayah  $i$ .

- g. Menguji kebaikan model dengan melakukan uji F.

$$F_h = \frac{(RSS_{OLS} - RSS_{GWR})/V}{RSS_{GWR}/\delta 1} = \frac{GWR_{IMP}/V}{RSS_{GWR}/\delta 1} \quad (7)$$

Dimana  $v = \text{tr}(R_0 - R_1)$  dan  $\delta 1 = \text{tr}(R_1)$ .

Hipotesis yang diuji adalah  $H_0$  : model GWR dan OLS samadan  $H_1$ : model GWR lebih baik dibanding model OLS. Jika hasil uji F tidak signifikan maka tidak terdapat perbedaan antara model GWR, sedangkan jika  $H_0$  ditolak maka lanjut ke langkah berikutnya.

- h. Melihat kebaikan model dengan melihat nilai koefisien determinasi ( $R^2$ ), jumlah kuadrat galat (RSS), BIC dan AIC dari model regresi global dan GWR.

- i. Jika model GWR lebih baik, maka dilanjutkan dengan pengujian pendugaan parameter secara parsial di setiap Provinsi di Indonesia.

Bentuk hipotesisnya adalah sebagai berikut [7]:

$$H_0 : \beta_k(ui, vi) = 0$$

$$H_1 : \beta_k(ui, vi) \neq 0$$

dengan  $i = 1, 2, \dots, n; k = 1, 2, \dots, p$

$$T_{hit} = \frac{\hat{\beta}(ui, vi)}{\sigma \sqrt{C_{kk}}} \quad (8)$$

$\hat{\beta}(ui, vi)$  adalah estimasi parameter (koefisien GWR) sedangkan  $C_{kk}$  adalah elemen diagonal ke-k dari matriks pembobotnya. Hipotesis nol akan ditolak jika  $|T_{hit}| > t_{\alpha/2, df}$  dimana  $df = \frac{\delta_1^2}{\delta_2}$ .

### HASIL PENELITIAN

Tahap awal yang dilakukan dalam penelitian ini adalah mencari model regresi klasik berganda berdasarkan variabel-variabel yang diperoleh dari kajian teori penelitian sebelumnya. Dalam penelitian ini pengolahan data menggunakan SPSS 19 dan diperoleh hasil sebagai berikut:

**Tabel 1.** Nilai Koefisien dan *p-value* Regresi Global

Variabel	Koefisien	t-stat	p-value
Intersep	-0.221	-0.216	0.830
Pendidikan	-0.030	-1.439	0.162
Umur	0.039	2.145	0.042
Sanitasi	0.022	1.701	0.101
KepadatanRT	0.014	1.196	0.243
Sampah	0.051	3.663	0.001
AirNonlayak	0.057	1.861	0.074
F-Statistik=12.939			
P-Value (F-Stat) = 0.000			
R-Square = 0.749			

Sesuai output diatas diperoleh hasil uji simultan (Uji F-Stat) bahwa model global yang dibentuk sudah cukup baik (dengan *p-value*  $0,0000 < 0,10$ ). Adapun model regresi global yang terbentuk sebagai berikut:

$$\text{HEPATITIS} = -\beta_0 - \beta_1 \text{PENDIDIKAN} + \beta_2 \text{UMUR} + \beta_3 \text{SANITASI} + \beta_4 \text{KEPADATANRT} + \beta_5 \text{SAMPAH} + \beta_6 \text{AIRNONLAYAK} + \varepsilon_i$$

Kemudian kita lakukan uji asumsi pada sisaannya yaitu harus berdistribusi normal, tidak ada multikolinearitas, tidak ada autokorelasi dan harus homogen. Dari output diperoleh hasil.

REGRESSION DIAGNOSTICS  
 MULTICOLLINEARITY CONDITION NUMBER 42.533185  
 TEST ON NORMALITY OF ERRORS  
 TEST DF VALUE PROB KEPUTUSAN  
 Jarque-Bera 2 0.9284 0.6286 NORMAL  
 (Data Berdistribusi Normal)

**Tabel 2.** Nilai VIF Variabel Prediktor

Variabel	VIF	KEPUTUSAN
Pendidikan	2.044	Nilai VIF < 10 berarti tidak ada mukltikolinearitas
Umur	1.304	
Sanitasi	3.709	
KepadatanRT	1.784	
Sampah	1.924	
AirNonlayak	2.263	
(Datatidak ada multikolinearitas)		
Test Durbin-Watson = 1.933 (Datatidak ada autokolerasi)		

DIAGNOSTICS FOR HETEROSKEDASTICITY  
 RANDOM COEFFICIENTS  
 TEST DF VALUE PROB KEPUTUSAN  
 Breusch-Pagan 6 14.5622 0.0240 HETEROGEN  
 (Data Tidak Homogen)

Dari uji asumsi sisaan diperoleh hasil bahwa dari 4 (empat) persyaratan asumsi yang diwajibkan, ada satu asumsi yang terlanggar yaitu sisaan tidak homogen (bersifat heterogen). Sehingga regresi global melanggar satu asumsi klasik.

Langkah selanjutnya adalah mengecek apakah ada pengaruh spasial pada sisaan, karena diperoleh sisaan model yang bersifat heterogen, sehingga diduga ada pengaruh spasial pada data. Uji yang digunakan untuk melihat ada tidaknya efek spasial pada data yaitu uji Indeks Moran's I, dimana hipotesisnya:

$H_0$ : Tidak ada autokorelasi spasial pada variabel kejadian Hepatitis

$H_1$ : Terdapat autokorelasi spasial pada variabel kejadian Hepatitis

Dari output menggunakan software geoda 1.6.7 dengan matriks pembobot

menggunakan metode *k-nearest neighbors* diperoleh hasil nilai Indeks Moran's sebesar 0,3018 dan nilai Z Indeks Moran's sebagai berikut:

DIAGNOSTICS FOR SPATIAL DEPENDENCE  
FOR WEIGHT MATRIX : queen.gal  
(row-standardized weights)

TEST	MI/DF	VALUE	P-Value
Moran's I (error)	0,1014	1.8388	0.0659

Dengan tingkat signifikansi 10 persen dapat dikatakan bahwa ada efek spasial pada data prevalensi penyebaran hepatitis hal ini terlihat dari nilai P-value yang lebih kecil dari tingkat signifikansi 0,10. Dari sini diduga model global variabel prediktor

Langkah selanjutnya mencari lebar jendela optimum dengan melihat nilai CV yang terkecil. Dalam penelitian ini dipakai software GWR4 untuk mencari nilai CV yang minimum. Fungsi Kernel yang digunakan adalah metode adaptif Gaussian. Alasan dipakainya metode gaussian karena memberikan bandwidth yang optimum tidak terlalu besar dan tidak terlalu kecil selain itu gaussian merupakan fungsi kernel yang baik karena memberikan nilai pembobot mengikuti kurva normal. Dari metode ini diperoleh nilai CV dan bandwidth terbaik sebagai berikut:

Tabel 3. Nilai CV dan Bandwidth Terbaik

Indikator	Adaptive Gaussian
(1)	(2)
Best Bandwidth Size	10
Minimum CV	0.299

Dari output diatas diperoleh bandwidth terbaik dengan metode penimbang kernel *Adaptive Gaussian* adalah sebesar 10. Artinya, terdapat 10 tetangga (provinsi) terdekat yang signifikan mempengaruhi suatu provinsi.

Hasil matriks jarak dan matriks pembobot  $W(u_i, v_i)$  sebagai berikut:

Tabel 4. Nilai Matriks pembobot Variabel Prediktor

Model GWR menghasilkan estimasi parameter yang berbeda-beda untuk

Variabel	Min	Q1	Med	Q3	Max	OLS
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Intersep	0.959	1.098	1.296	1.402	1.482	-0.221
Pendidikan	-0.313	-0.157	-0.117	-0.010	0.059	-0.030
Umur	0.063	0.082	0.187	0.282	0.344	0.039
Sanitasi	0.031	0.058	0.142	0.207	0.314	0.022
KepadatanRT	0.064	0.084	0.122	0.208	0.356	0.014
Sampah	-0.054	0.092	0.409	0.452	0.515	0.051
AirNonlayak	0.091	0.146	0.259	0.341	0.492	0.057

setiap lokasi. Sedangkan model OLS menghasilkan nilai estimasi parameter yang sama untuk semua lokasi penelitian yaitu yang ditunjukkan oleh kolom (7). Sementara kolom (2) sampai (6) merupakan statistik deskriptif dari estimasi koefisien model GWR yang meliputi nilai minimum, kuartil pertama, median, kuartil ketiga, dan nilai maksimum. Model GWR memungkinkan suatu variabel prediktor yang sama memiliki hubungan dengan variabel respon yang positif di suatu lokasi dan nilai negatif di lokasi lainnya. Variabel Pendidikan dan Sampah memiliki hubungan yang positif dan negatif terhadap prevalensi Hepatitis. Variabel yang memiliki pengaruh positif terhadap prevalensi Hepatitis di semua lokasi adalah Sanitasi, sedangkan variabel yang memiliki pengaruh negatif terhadap prevalensi Hepatitis di semua lokasi adalah penggunaan jarum suntik.

Langkah berikutnya adalah menguji model GWR yang dibentuk. Untuk melihat apakah model GWR lebih baik dibanding model regresi global, maka dilakukan *goodness of fit test*. Pengujian ini bertujuan untuk mengidentifikasi model GWR terbentuk mampu secara signifikan menjelaskan hubungan variabel respon dan variabel penjelasnya.

Hipotesis:

$H_0$ : model GWR dan OLS sama

$H_1$ : model GWR lebih baik dibanding model OLS

Berikut tabel Goodness of Fit hasil penelitian:

**Tabel 5.** Nilai ANOVA Model GWR

Source	SS	DF	MS	F hit	Ftabel
(1)	(2)	(3)	(4)	(5)	(6)
<i>Global Residual</i>	5.246	26.000			
<i>GWR Improvement</i>	3.143	10.209	0.308		
<i>GWR Residual</i>	2.103	15.791	0.133	2.312	2.059

kolom (1) merupakan sumber variasi dimana *GWR Improvement* adalah penurunan residual dari metode OLS (global) akibat penggunaan metode GWR. Sedangkan nilai *sum of square* (SS) pada kolom (2) dari *GWR Improvement* merupakan selisih antara SS *Global Residuals* dan SS *GWR Residuals*. Kolom (3) merupakan derajat bebas (DF) dimana baris kedua dan ketiga berturut-turut adalah  $v_1$  dan  $\delta_1$ , kolom (4) merupakan *mean of square* (MS) yang diperoleh dari SS dibagi Derajat bebas (DF), kolom (5) merupakan statistik uji F yang diperoleh dari pembagian MS *GWR Improvement* dengan MS *GWR Residuals*, dan kolom (6) merupakan nilai Ftabel dari distribusi F dengan derajat bebas  $v_1^2/v_2$ ,  $\delta_1^2/\delta_2$ . Hipotesis nol dari uji ini adalah tidak ada perbedaan antara model regresi global dan model GWR dalam menjelaskan hubungan antar variabel. Hipotesis nol ditolak jika  $F_{hitung} > F_{tabel}$  ( $0,1)(v_1^2/v_2, \delta_1^2/\delta_2$ ). Dari tabel 7 diketahui bahwa nilai  $F_{hitung}$  yang dihasilkan lebih besar dari nilai  $F_{tabel}$  ( $2,312 > 2,059$ ), sehingga hipotesis nol dapat ditolak. Artinya dapat disimpulkan bahwa dengan tingkat kepercayaan 90 persen, model GWR dapat menjelaskan hubungan antara variabel prediktor dan variabel respon lebih baik dibandingkan model regresi global.

Beberapa indikator yang dapat digunakan untuk membandingkan model

GWR dan model regresi global adalah  $R^2$ , AIC, dan *residual sum of square*.

**Tabel 6.** Nilai R-Square, AIC, RSS, dan BIC Model GWR dan Global

Pembanding	GWR	Regresi Global
<i>R-Square</i>	0.899	0.749
<i>AIC</i>	32.987	48.961
<i>RSS</i>	2.102	5.246
<i>BIC/MDL</i>	55.580	60.933

Berdasarkan empat alternatif ukuran yang mempertimbangkan tiga indikator ( $R^2$ , AIC, RSS, dan BIC/MDL), dapat dilihat bahwa model GWR dengan fungsi kernel *Adaptive Gaussian* lebih baik dibandingkan model regresi global memiliki nilai AIC, RSS, dan BIC/MDL yang paling kecil serta *R-Square* yang paling besar.

Setelah didapatkan bahwa model GWR adalah model yang lebih baik daripada model regresi global selanjutnya masing-masing variabel akan di uji secara parsial. Seperti yang telah dijelaskan sebelumnya, metode GWR menghasilkan model lokal yang berbeda-beda untuk setiap lokasi. Untuk mengetahui variabel prediktor yang signifikan memengaruhi variabel respon untuk setiap provinsi digunakan uji parsial menggunakan rumus (8). Dari pengujian parsial dengan menggunakan tingkat signifikansi 10% diperoleh variabel yang signifikan dimasing-masing provinsi sebagai berikut.

**Tabel 7.** Variabel prediktor yang signifikan memengaruhi prevalensi Hepatitis di Indonesia

No	PROVINSI	Variabel yang signifikan
(1)	(2)	(3)
1	Sumatera Utara, Riau	AirNonlayak
2	Kalimantan Barat	Sampah
3	Aceh	Sampah, AirNonlayak
4	Jawa Barat	Sanitasi, Sampah
5	Jawa Timur, Kalimantan Tengah, Kalimantan Selatan	Umur, Sampah, AirNonlayak

No	PROVINSI	Variabel yang signifikan
6	Jawa Tengah	Umur, Sanitasi, Sampah
7	DI Yogyakarta	Umur, Sanitasi, Sampah, AirNonLayak
8	Kalimantan Timur, Sulawesi Barat, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Sulawesi Tengah, Sulawesi Tenggara, Sulawesi Selatan	Umur, KepadatanRT, Sampah, AirNonLayak
9	Papua, Papua Barat	Pendidikan, Sanitasi, Umur, Sampah
10	Maluku Utara, Sulawesi Utara, Gorontalo	Pendidikan, Umur, KepadatanRT, Sampah, AirNonLayak
11	Maluku	Pendidikan, Sanitasi, Umur, KepadatanRT, Sampah, AirNonlayak
12	Sumatera Barat, Jambi, Sumatera Selatan, Bengkulu, Lampung, Bangka Belitung, Kepulauan Riau, DKI Jakarta, Banten	Tidak ada vaiabel yang signifikan

Berdasarkan Tabel 7, terbentuk 12 kelompok provinsi yang memiliki kesamaan variabel yang signifikan berpengaruh terhadap prevalensi Hepatitis di Indonesia Tahun 2013. Walaupun beberapa variabel prediktor di model global sebelumnya ada yang tidak signifikan, namun pada model GWR variabel-variabel tersebut signifikan di beberapa propinsi. Ini menandakan bahwa belum tentu variabel yang tidak signifikan di model global, tidak signifikan juga di model GWR. Kelebihan model GWR adalah data bisa ditampilkan dalam bentuk peta tematik. Dengan menggunakan software ArcGIS dibuatlah peta variabel preditor yang berpengaruh di masing-masing propinsi.

Lebih lengkapnya dapat dilihat pada peta berikut ini.



Gambar 1. Peta tematik variabel yang mempengaruhi di tiap provinsi di Indonesia 2013

Dengan model GWR bisa dilihat variasi masing-masing parameter variabel prediktornya yaitu dengan membandingkan model GWR dengan model GWR pengganti dimana variabel ke- $k$  dianggap konstan dan variabel lainnya bervariasi. *DIFF of Criterion* merupakan nilai dari perbedaan kriteria dalam hal ini AIC antara model GWR yang asli dengan model GWR pengganti. Apabila *DIFF of Criterion* bernilai positif, maka mengindikasikan tidak adanya variasi spasial. Dengan kata lain estimasi koefisien dari variabel tersebut tidak signifikan berbeda pada setiap wilayah atau nilainya dianggap konstan untuk seluruh provinsi. Sedangkan jika nilai dari *DIFF of Criterion* adalah negatif, maka terjadi *spatial nonstationarity* pada variabel tersebut atau nilai koefisien regresi dari variabel tersebut signifikan berbeda-beda untuk setiap provinsi [1].

Tabel 8 Nilai DiFF of Criterion Variabel

Variabel	F	DOF for F test	DIFF of Criterion	
Intercept	7.883499	0.833	18.903	-0.029317
PENDIDIKAN	3.972293	0.687	18.903	-0.004252
UMUR	2.521470	0.789	18.903	-0.017625
SANITASI	7.229851	0.643	18.903	-0.030661
KEPADTRT	1.019751	0.961	18.903	0.014095
SAMPAH	5.369233	0.847	18.903	-0.019995
AIRNONLAYAK5	1.94609	1.269	18.903	-0.009167

Dari nilai *DIFF of Criterion* hanya variabel KEPADATANRT yang



memiliki nilai positif sedangkan variabel prediktor lainnya bernilai negatif. Hal ini menunjukkan bahwa variabel PENDIDIKAN, UMUR, SANITASI, SAMPAH, dan AIRNONLAYAK memiliki variasi spasial di masing-masing provinsi. Variasi masing-masing parameter model GWR menunjukkan bahwa semua variabel bersifat *spatial nonstationarity*. Maka model yang terbentuk dalam penelitian ini merupakan *mixed GWR* yang dapat dituliskan sebagai berikut:

$$\hat{y}_i = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)X_{1i} + \beta_2(u_i, v_i)X_{2i} + \beta_3(u_i, v_i)X_{3i} + \beta_4X_{4i} + \beta_5(u_i, v_i)X_{5i} + \beta_6(u_i, v_i)X_{6i}$$

dimana  $i = 1, 2, \dots, 33$

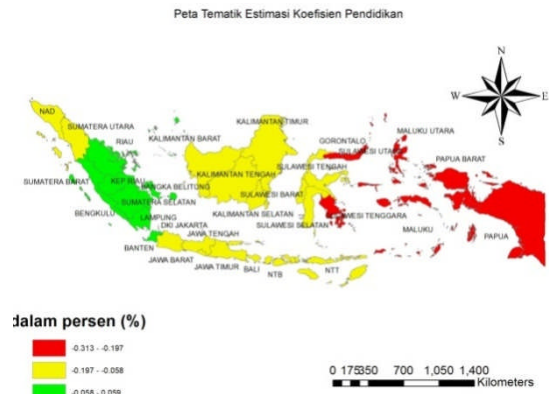
sehingga akan terdapat 33 persamaan yang berbeda untuk masing-masing provinsi, misal persamaan untuk variabel yang signifikan memengaruhi prevalensi hepatitis di Provinsi Aceh sebagai berikut:

$$\hat{y}_{ACEH} = 1,299 + 0,253X_{5ACEH} + 0,254X_{6ACEH}$$

Untuk koefisien masing-masing provinsi bisa dilihat pada lampiran.

Selain mempunyai persamaan berbeda-beda untuk masing-masing provinsi, berikut contoh penyajian lain untuk menyajikan variasi estimasi koefisien pendidikan dalam peta dan bagaimana cara menginterpretasikan. Estimasi koefisien pendidikan terhadap prevalensi Hepatitis di Indonesia adalah negatif di seluruh wilayah Indonesia. Hal ini menunjukkan bahwasemakin sedikit persentase penduduk yang berpendidikan SD kebawah di suatu daerah dapat menurunkan prevalensi Hepatitis di Indonesia. Sebagai contoh misalnya seluruh Provinsi di Sulawesi dan Maluku adalah daerah-daerah yang memiliki pengaruh negatif dan kuat (-0,31 sampai -0,19). Artinya, penurunan satu persen persentase penduduk berpendidikan SD kebawah di daerah tersebut akan menurunkan prevalensi

Hepatitis sebesar (0,19 sampai 0,31 persen)



Gambar 2. Peta tematik estimasi koefisien pendidikan

## KESIMPULAN

Model analisis terbaik untuk variabel yang memengaruhi prevalensi hepatitis di Indonesia adalah model GWR. Dengan terbentuk 12 kelompok provinsi berdasarkan variabel prediktor yang memengaruhi. Variabel prediktor yang tidak signifikan di model regresi global, ternyata berpengaruh signifikan di beberapa provinsi. Hal ini menandakan bahwa variabel tidak berpengaruh secara global belum tentu tidak berpengaruh secara lokal. Hampir seluruh variabel prediktor memiliki variasi spasial, hanya satu variabel yang tidak bervariasi secara spasial yaitu variabel KEPADATANRT.

## DAFTAR PUSTAKA

- [1] Fotheringham, A.S., Brunson, C., & Charlton, M. *Geographically Weighted Regression*, John Wiley & Sons, Chichester, UK, 2002.
- [2] Sugiyanto, 2008. *Model-model Pembelajaran Inovatif*. Surakarta: Panitia Sertifikasi Guru Rayon 13.
- [3] Kementerian Kesehatan RI. 2013. *Riset Kesehatan Dasar tahun 2013*. Jakarta : Kementerian Kesehatan RI.
- [4] Sulaiman, Ali. 2007. *Buku Ajar Ilmu Penyakit Hati*. Jakarta : Jaya abadi

- [5] Anselin, L., *Exploratory Spatial Data Analysis and Geographic Information Systems*, National Center for Geographic Information and Analysis of California Santa Barbara: CA93106,1993.
- [6] Chasco, C., Garcia, I., & Vicens, J. 2007, *Modeling Spastial Variations in Household Disposable Income with Geographically Weighted Regression*, Munich Personal RePEc Arkhive (MPRA) Working Papper No. 1682.

LAMPIRAN:

Tabel 9. Koefisien-koefisien berdasarkan estimasi model GWR

NO	PROVINSI	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	Local R <sup>2</sup>
1	Aceh	1.299	-0.078	0.108	0.160	0.123	0.253	0.294	0.838
2	Sumatera Utara	1.238	-0.059	0.085	0.135	0.099	0.180	0.282	0.801
3	Sumatera Barat	1.105	0.003	0.068	0.073	0.086	0.026	0.262	0.724
4	Riau	1.125	-0.009	0.068	0.081	0.085	0.056	0.259	0.729
5	Jambi	1.004	0.049	0.066	0.032	0.091	-0.037	0.185	0.560
6	Sumatera Selatan	0.959	0.060	0.066	0.040	0.089	-0.035	0.120	0.423
7	Bengkulu	0.976	0.056	0.071	0.046	0.094	-0.054	0.145	0.532
8	Lampung	1.042	0.001	0.079	0.143	0.081	0.082	0.104	0.633
9	Bangka Belitung	0.999	0.041	0.065	0.057	0.083	0.009	0.142	0.508
10	Kepulauan Riau	1.087	0.006	0.063	0.059	0.085	0.042	0.223	0.663
11	DKI Jakarta	1.092	-0.042	0.096	0.200	0.070	0.166	0.109	0.713
12	Jawa Barat	1.152	-0.076	0.116	0.233	0.070	0.246	0.141	0.787
13	Jawa Tengah	1.205	-0.121	0.154	0.249	0.064	0.350	0.213	0.832
14	DI Yogyakarta	1.215	-0.121	0.164	0.252	0.075	0.364	0.225	0.846
15	Jawa Timur	1.269	-0.120	0.207	0.203	0.123	0.433	0.318	0.872
16	Banten	1.039	-0.012	0.086	0.164	0.075	0.102	0.091	0.634
17	Bali	1.319	-0.110	0.251	0.160	0.192	0.475	0.401	0.898
18	Nusa Tenggara Barat	1.325	-0.112	0.263	0.145	0.206	0.490	0.422	0.899
19	Nusa Tenggara Timur	1.390	-0.165	0.273	0.211	0.210	0.473	0.320	0.909
20	Kalimantan Barat	1.213	-0.075	0.102	0.161	0.084	0.263	0.200	0.782
21	Kalimantan Tengah	1.305	-0.117	0.188	0.142	0.148	0.410	0.316	0.840
22	Kalimantan Selatan	1.296	-0.134	0.203	0.138	0.139	0.440	0.338	0.833
23	Kalimantan Timur	1.387	-0.122	0.242	0.081	0.251	0.451	0.387	0.866
24	Sulawesi Utara	1.468	-0.236	0.300	0.174	0.239	0.442	0.302	0.878
25	Sulawesi Tengah	1.428	-0.150	0.282	0.048	0.305	0.478	0.429	0.869
26	Sulawesi Selatan	1.392	-0.142	0.303	0.047	0.304	0.515	0.484	0.879
27	Sulawesi Tenggara	1.470	-0.203	0.334	0.039	0.356	0.499	0.492	0.877
28	Gorontalo	1.451	-0.197	0.287	0.131	0.259	0.454	0.345	0.880
29	Sulawesi Barat	1.405	-0.138	0.283	0.044	0.299	0.490	0.448	0.868
30	Maluku	1.482	-0.313	0.344	0.314	0.204	0.437	0.241	0.889
31	Maluku Utara	1.472	-0.271	0.311	0.251	0.202	0.432	0.238	0.883
32	Papua Barat	1.427	-0.244	0.283	0.313	0.154	0.437	0.147	0.892
33	Papua	1.401	-0.213	0.256	0.307	0.140	0.435	0.142	0.897