

# PEMODELAN INDEKS PEMBANGUNAN KESEHATAN MASYARAKAT PROVINSI JAWA TIMUR DENGAN METODE GEOGRAPHICALLY WEIGHTED LOGISTIC REGRESSION

Quraini Septy Wardhani<sup>1</sup>, Sri Sulistijowati Handajani<sup>2</sup>, Irwan Susanto<sup>3</sup>

<sup>1,2,3</sup> Program Studi Statistika, FMIPA Universitas Sebelas Maret, Surakarta  
e-mail: <sup>2</sup>rr\_ssh@staff.uns.ac.id

## Abstrak

*Geographically weighted logistic regression (GWLR)* adalah metode yang dapat memodelkan data bersifat kategorik dengan mempertimbangkan efek spasial. Pada penelitian ini, metode GWLR digunakan untuk memodelkan indeks pembangunan kesehatan masyarakat (IPKM) di Provinsi Jawa Timur. IPKM merupakan indikator keberhasilan pemerintah dalam bidang kesehatan. Tingginya IPKM disuatu wilayah menunjukkan kualitas kesehatan yang baik di wilayah tersebut. Variabel yang digunakan adalah prevalensi balita *stunting*, hipertensi, pneumonia, persalinan ditangani nakes, pengguna KB dengan metode kontrasepsi jangka panjang (MKJP), dan penduduk dengan perilaku cuci tangan benar. Hasil penelitian ini menunjukkan bahwa pemodelan GWLR dengan pembobot *adaptive Gaussian Kernel* lebih baik daripada pembobot lain dengan faktor yang berpengaruh signifikan secara lokal adalah prevalensi balita *stunting* dan hipertensi. Nilai akurasi, *sensitivity*, dan *specificity* yang dihasilkan berturut-turut sebesar 97,4%, 100% dan 85,71%.

**Kata kunci:** IPKM, GWLR, *Adaptive Gaussian Kernel*.

## Abstract

*Geographically weighted logistic regression (GWLR)* is a method that can model categorical data by considering spatial effects. In this study, the GWLR method was used to model the public health development index (IPKM) in East Java Province. IPKM is an indicator of the government's success in the health sector. The high IPKM in an area shows the good quality of health in that region. The variables used are prevalence of stunting in children under five years, hypertension, pneumonia, childbirth handled by health workers, family planning users with long-term methods of contraception, and residents with proper hand washing behavior. The results of this study indicate that the GWLR modeling with the adaptive Gaussian Kernel weighting is better than the other weights with factors that have a significant local effect are the prevalence of stunting and hypertension under five. The resulting accuracy, sensitivity, and specificity values are 97.4%, 100% and 85.71%, respectively.

**Keywords:** PHDI, GWLR, *Adaptive Gaussian Kernel*.

## PENDAHULUAN

Indeks pembangunan kesehatan masyarakat (IPKM) merupakan alat ukur yang digunakan pemerintah untuk mengetahui kondisi kesehatan masyarakat di suatu wilayah. IPKM memuat 30 indikator yang dikelompokkan ke dalam 7 sub-indeks. Sub-indeks yang termuat dalam IPKM yaitu kesehatan balita, kesehatan reproduksi, pelayanan kesehatan, perilaku kesehatan, penyakit tidak menular, penyakit menular, dan kesehatan lingkungan (Kemenkes, 2019). Data yang tercantum dalam setiap indikator diperoleh dari tiga survei yaitu riset kesehatan dasar (Riskesdas), survei sosial ekonomi nasional (Susenas), dan potensi desa (Podes) (Kemenkes, 2011).

Pada tahun 2018, Jawa Timur merupakan salah satu provinsi di Indonesia yang lebih dari 30% kabupaten/kotanya mengalami penurunan peringkat IPKM dari IPKM tahun 2013. Sepanjang tahun 2018, jumlah kematian bayi yang terjadi adalah sebanyak 4.016 bayi, dengan 4.338 kematian balita serta 3.148 kematian neonatal (Kemenkes, 2019). Selain itu, beberapa kasus kesehatan juga terjadi di Jawa Timur pada 2018, mulai dari tingginya kasus demam berdarah dengue (DBD), hingga angka *stunting* yang lebih tinggi dari nasional (Kemenkes, 2019). Oleh karena itu, diperlukan analisis statistik untuk mengetahui faktor-faktor yang mempengaruhi IPKM di Provinsi Jawa Timur dengan mempertimbangkan faktor spasial. Salah satu metode yang dapat digunakan untuk memodelkan data IPKM adalah metode *Geographically Weighted Logistic Regression* (GWLR).

GWLR adalah metode non-parametrik untuk menganalisis data spasial yang merupakan bentuk lokal dari regresi logistik (Desriwendi, dkk, 2015). Regresi logistik merupakan teknik analisis multivariat yang bertujuan untuk menjelaskan hubungan antara variabel independen dengan variabel dependen yang bersifat kategorik (Hosmer dan Lemeshow, 2000). Data variabel dependen pada model GWLR diasumsikan berdistribusi Bernoulli

yang menghasilkan nilai koefisien regresi pada setiap lokasi berbeda-beda (Fotheringham, dkk, 2002).

Penelitian dengan menggunakan metode GWLR sebelumnya pernah dilakukan oleh Faturrahman (2016) untuk mengetahui faktor-faktor yang mempengaruhi IPKM di Provinsi Papua. Selain itu, Albuquerque (2016) melakukan penelitian terkait pemodelan penilaian kredit dengan menggunakan metode GWLR, serta Sabat (2017) yang menggunakan metode GWLR untuk memodelkan data angka kematian bayi. Penelitian terkait dengan IPKM sebelumnya pernah dilakukan oleh Octavany (2017) yang melakukan penelitian terkait faktor yang berpengaruh terhadap IPKM di Provinsi Jawa Timur dengan pendekatan regresi *spline*. Faturrahman (2019) juga pernah melakukan penelitian mengenai pemodelan IPKM kabupaten/kota di Pulau Kalimantan dengan menggunakan pendekatan regresi probit. Dibandingkan dengan metode regresi *spline* dan probit yang telah dilakukan pada penelitian sebelumnya, metode GWLR mampu menyajikan model regresi berdasarkan lokasi sehingga akan terbentuk model sejumlah  $n$  lokasi. Perbedaan model tiap lokasi memungkinkan peneliti untuk mengetahui lebih spesifik terkait dengan permasalahan yang lebih mendesak dan harus segera diatasi di masing-masing lokasi.

## METODOLOGI

### Regresi Logistik

Model regresi logistik merupakan bentuk khusus dalam model linier umum (*Generalized Linear Models* atau GLM) di mana variabel acak dependen bernilai kategorik (0 dan 1) dan mengikuti distribusi Bernoulli. Model logistik ganda dapat dituliskan sebagai berikut (Agresti, 2007):

$$\pi(x_i) = \frac{\exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ji})}{1 + \exp(\beta_0 + \sum_{j=1}^p \beta_j x_{ji})}$$

di mana  $\pi(x_i)$  adalah observasi variabel dependen ke- $i$  dengan  $i = 1, 2, 3, \dots, n$ ;  $\beta_0$  adalah intersep model regresi logistik;

$\beta_j$  adalah nilai parameter regresi variabel independen ke- $k$ ; dan  $x_{ji}$  adalah nilai observasi variabel independen ke- $j$  pada lokasi ke- $i$ . Pendugaan parameter regresi logistik menggunakan metode *maximum likelihood estimation* (MLE) dengan persamaan logaritma *likelihood* yang terbentuk yaitu:

$$\ell(\beta) = \sum_{j=0}^p \left( \sum_{i=1}^n y_i x_{ji} \right) \beta_j - \sum_{i=1}^n \ln \left( 1 + \exp \left( \sum_{j=0}^p \beta_j x_{ji} \right) \right)$$

Penyelesaian pendugaan parameter regresi logistik dari fungsi logaritma *likelihood* dapat dilakukan dengan melalui prosedur *iteratively reweighted least square* (IRLS) dari metode *Newthon Raphson* (Lestari, 2020).

### Pembobot Spasial

Pada analisis spasial, diperlukan pembobot spasial pada masing-masing lokasi ke- $i$ . Sebelum melakukan perhitungan pembobot spasial, perlu dilakukan perhitungan jarak *Euclidean*. Apabila lokasi ke- $j$  terletak pada koordinat  $(u_j, v_j)$  maka akan diperoleh jarak *Euclidean* antara lokasi ke- $i$  dan lokasi ke- $j$  sebagai berikut (Yrigoyen, dkk, 2007).

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

Terdapat beberapa fungsi pembobot *Kernel* yang dapat digunakan untuk analisis spasial yaitu (Foringham, dkk, 2002):

#### 1. Fixed Gaussian Kernel

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

#### 2. Adaptive Gaussian Kernel

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

#### 3. Adaptive Bisquare Kernel

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

dengan  $h$  adalah *bandwith* yang bernilai sama untuk setiap lokasi;  $h_i$  adalah *bandwith* pada lokasi ke- $i$ ; dan  $d_{ij}$  adalah jarak *Euclidean* antara lokasi  $(u_i, v_i)$  ke  $(u_j, v_j)$ .

Dalam mencari pembobot pada masing-masing lokasi didasarkan pada jarak *Euclidean* ( $d_{ij}$ ) dan *bandwidth* ( $h$ ) yang dihasilkan pada masing-masing lokasi. Pemilihan ukuran *bandwidth* yang optimum menjadi salah satu hal yang penting karena akan mempengaruhi ketepatan hasil regresi (Fotheringham, *et al.*, 2002). Pada penelitian ini digunakan metode *Generalized Cross Validation* (GCV) untuk penentuan *bandwidth* optimal dan pemilihan model terbaik. GCV merupakan modifikasi dari *Cross Validation* (CV) yang didapat dengan meminimumkan fungsi CV (Eubank, dkk, 1988). Nilai *bandwidth* ( $h$ ) yang optimal diperoleh dari *bandwidth* yang menghasilkan nilai GCV minimum. GCV dapat dihitung dengan rumus (Suyitno, dkk.,2016):

$$GCV = \frac{n \sum_{i=1}^n [\hat{\pi}(u_i, v_i) - \bar{Y}]^2}{(n - v)^2}$$

dengan  $v = tr(\mathbf{S})$ , di mana matriks  $\mathbf{S}$  merupakan matriks berukuran  $(n \times n)$  yang dituliskan dengan rumus  $\mathbf{S} = \mathbf{X}(\mathbf{X}^T \mathbf{W} \mathbf{V} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{V}$ ;  $\mathbf{W}$  adalah matriks diagonal pembobot spasial; dan  $\mathbf{V}$  adalah matriks diagonal yang memiliki elemen diagonal  $\hat{\pi}_i [1 - \hat{\pi}_i]$ .

### Geographically Weighted Logistic Regression (GWLR)

*Geographically Weighted Logistic Regression* (GWLR) adalah metode non parametrik yang merupakan bentuk lokal dari regresi logistik di mana lokasi diperhatikan dan diasumsikan bahwa data dari variabel dependen berdistribusi Bernoulli yang digunakan untuk menganalisis data spasial (Desriwendi, dkk., 2015). Model GWLR dapat dituliskan sebagai berikut (Caraka, dkk, 2017).

$$\pi(x_i) = \frac{\exp(\sum_{j=0}^p \beta_j(u_i, v_i) x_{ji})}{1 + \exp(\sum_{j=0}^p \beta_j(u_i, v_i) x_{ji})}$$

$; i = 1, 2, \dots, n$

dengan  $\pi(x_i)$  adalah nilai observasi variabel dependen ke- $i$ ;  $\beta_{j(u_i, v_i)}$  adalah vektor nilai parameter variabel independen ke- $j$  pada lokasi ke- $i$ ;  $(u_i, v_i)$  merupakan koordinat letak geografis (*longitude*, *latitude*) dari lokasi ke- $i$ ; dan  $x_{ji}$  adalah nilai observasi variabel independen ke- $j$  pada lokasi ke- $i$ .

Pendugaan parameter model GWLR dilakukan dengan metode *maximum likelihood estimation* (MLE) dengan fungsi logaritma *likelihood* untuk model GWLR dengan pembobot spasial ditunjukkan pada rumus persamaan berikut (Aji, 2014).

$$\ell^*(\beta_{(u_i, v_i)}) = \sum_{j=0}^p \left( \sum_{i=1}^n y_i w_{i(u_i, v_i)} \right) \beta_{j(u_i, v_i)} x_{ji} - \sum_{i=1}^n w_{i(u_i, v_i)} \ln \left( 1 + \exp \left( \sum_{j=0}^p \beta_{j(u_i, v_i)} x_{ji} \right) \right)$$

Terdapat dua pengujian parameter yang harus dilakukan untuk memodelkan data dengan metode GWLR, yaitu pengujian parameter secara serentak dan parsial. Pengujian parameter GWLR secara serentak bertujuan untuk mengetahui signifikan tidaknya pengaruh antara variabel independen dan variabel dependen secara bersamaan atau serentak. Hipotesis uji yang dapat digunakan untuk pengujian parameter GWLR secara serentak sebagai berikut (Atkinson, 2003).

$H_0$  :  $\beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \dots = \beta_p(u_i, v_i) = 0$  (Tidak ada pengaruh yang signifikan antara variabel independen terhadap variabel dependen secara serentak).

$H_1$  : Paling sedikit satu  $\beta_j(u_i, v_i) \neq 0$ ;  $j = 1, 2, \dots, p$  (Ada pengaruh yang signifikan antara variabel independen terhadap variabel dependen secara serentak).

$H_0$  ditolak jika nilai  $p$ -value  $< \alpha$  atau jika nilai  $G_2 > \chi^2_{\alpha, p}$  dengan rumus pengujian statistik yang digunakan adalah

$$G_2 = 2(\ell(\hat{\Omega}_{GWLR}) - \ell(\hat{\omega}_{GWLR}))$$

$$P\text{-value}_{\text{serentak(GWLR)}} = P(G_v > G_2)$$

di mana

$$\ell(\hat{\Omega}_{GWLR}) = \sum_{i=1}^n (y_i \ln \hat{\pi}_i(u_i, v_i) + (1 - y_i) \ln (1 - \hat{\pi}_i(u_i, v_i)))$$

$$\ell(\hat{\omega}) = \sum_i^n (n_{1i} \ln(n_{1i}) + n_{0i} \ln(n_{0i}) - n \ln(n))$$

Pengujian parameter GWLR yang selanjutnya adalah pengujian parameter secara parsial. Pengujian ini bertujuan untuk mengetahui variabel independen mana yang berpengaruh signifikan terhadap variabel dependen. Hipotesis pengujian yang digunakan sebagai berikut (Foringham, 2002).

$H_0$  :  $\beta_j(u_i, v_i) = 0$  (Tidak ada pengaruh yang signifikan antara variabel independen terhadap variabel dependen).

$H_1$  :  $\beta_j(u_i, v_i) \neq 0$  (Ada pengaruh yang signifikan antara variabel independen terhadap variabel dependen).

$H_0$  ditolak jika nilai  $p$ -value  $< \alpha$  atau  $|W_j| > Z_{\alpha/2}$  dengan rumus pengujian statistik yang digunakan adalah

$$W_j = \frac{\hat{\beta}_j(u_i, v_i)}{SE(\hat{\beta}_j(u_i, v_i))}$$

$$P\text{-value}_{\text{parsial(GWLR)}} = P(Z > |W_j|)$$

dengan

$$SE(\hat{\beta}_j(u_i, v_i)) = \sqrt{\text{var} \hat{\beta}_j(u_i, v_i)}$$

Pengujian kesamaan model regresi logistik dan model GWLR menggunakan perbandingan nilai *deviance* model Regresi Logistik dan model GWLR dengan hipotesis sebagai berikut (Faturrahman, 2016).

$H_0$  :  $\beta_j(u_i, v_i) = \beta_j$ ;  $i = 1, 2, \dots, n, j = 1, 2, \dots, p$  (Tidak ada perbedaan yang signifikan antara model regresi logistik dan model GWLR)

$H_1$  : paling sedikit satu  $\beta_j(u_i, v_i) \neq \beta_j$ ;  $i = 1, 2, \dots, n, j = 1, 2, \dots, p$  (Ada perbedaan yang signifikan antara model Regresi Logistik dan model GWLR)

$H_0$  ditolak jika  $F_{hit} > F_{(\alpha, v_1, v_2)}$ . Misalkan model regresi logistik *deviance* dinyatakan dengan  $D(\hat{\beta})$  dengan derajat bebas  $v_1$  dan model GWLR dinyatakan dengan  $D(\hat{\beta}(u_i, v_i))$  dengan derajat bebas  $v_2$  maka statistik uji yang digunakan yaitu:

$$F_{hit} = \frac{D(\hat{\beta})/v_1}{D(\hat{\beta}(u_i, v_i))/v_2}$$

$$P\text{-value}_{\text{kesamaan(GWLR)}} = P(F_v > F_{hit})$$

## Pemilihan Model Terbaik

Pemilihan model terbaik bertujuan untuk mencari metode yang sesuai untuk memodelkan data. Pada penelitian ini, pemilihan model terbaik dilakukan dengan membandingkan nilai *Generalized Cross Validation* (GCV), *Aikake Information Criterion* (AIC), dan koefisien determinasi ( $R^2$ ). AIC merupakan pengukuran untuk kualitas relatif dari model statistik dari data yang diberikan. Semakin kecil nilai AIC maka model akan semakin baik. Rumus AIC didefinisikan sebagai berikut (Suyitno, dkk., 2016).

$$AIC = 2n \log(\hat{\sigma}) + n \log(2\pi) + n \left( \frac{n + v}{n - v - 2} \right)$$

Koefisien determinasi ( $R^2$ ) adalah nilai yang menjelaskan seberapa besar kemampuan variabel independen dalam menjelaskan variabel dependen. Apabila nilai  $R^2$  mendekati 1 maka model semakin baik. Rumus dari *R-squared* McFadden's didefinisikan sebagai (Lestari, 2020)

$$R_{MF}^2 = 1 - \frac{\ell(\hat{\Omega}_{GWLR})}{\ell(\hat{\omega}_{GWLR})}$$

## Variabel Penelitian

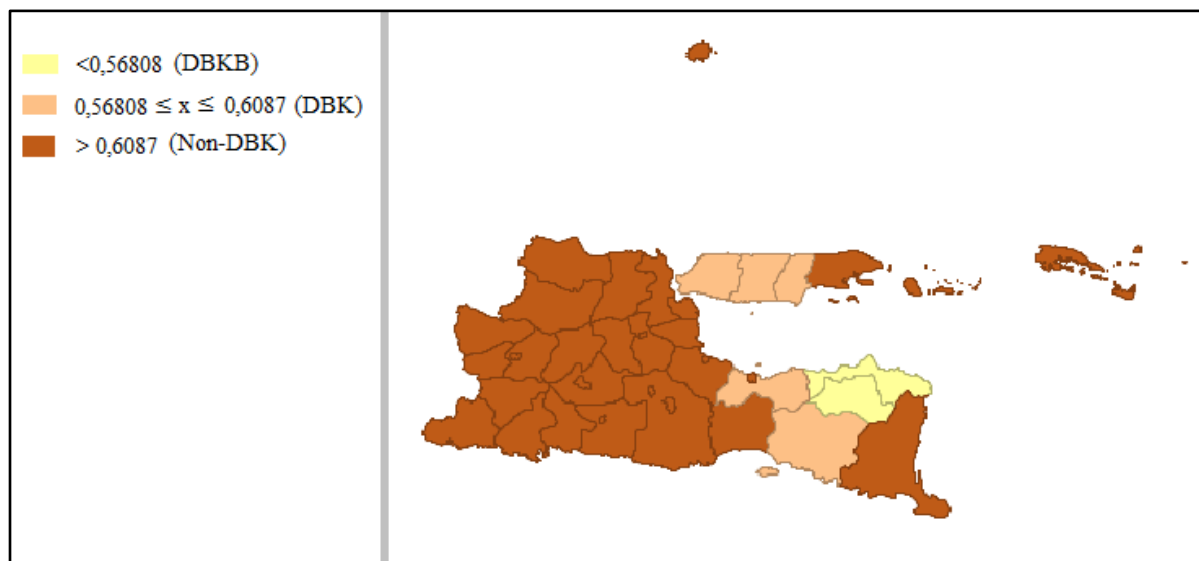
Data yang digunakan dalam penelitian ini merupakan data sekunder yang diambil dari buku IPKM 2018 oleh Kementerian kesehatan Republik Indonesia yang diterbitkan pada tahun 2019. Terdapat

satu variabel dependen dan lima variabel independen pada penelitian ini. Variabel dependen yang digunakan adalah IPKM Provinsi Jawa Timur tahun 2018. Pada penelitian ini, IPKM dikelompokkan menjadi dua kategori, yaitu daerah tidak bermasalah kesehatan (non-DBK) dengan kategori 0 jika nilai IPKM  $> 0,6087$  dan daerah bermasalah kesehatan (DBK) dengan kategori 1 jika nilai IPKM  $< 0,6087$ . Sedangkan untuk variabel independen yang digunakan, yaitu prevalensi balita *stunting* ( $X_1$ ), prevalensi hipertensi ( $X_2$ ), prevalensi pneumonia ( $X_3$ ), persentase persalinan ditangani nakes ( $X_4$ ), persentase pengguna KB MKJP ( $X_5$ ), persentase proporsi penduduk dengan perilaku cuci tangan benar ( $X_6$ ).

## HASIL DAN PEMBAHASAN

Berikut merupakan statistik deskriptif dari data yang digunakan:

Pada peta tematik Gambar 1, menunjukkan bahwa terdapat dua daerah yang masuk dalam kelompok bermasalah kesehatan berat yaitu Kabupaten Bondowoso dan Kabupaten Situbondo. Selain itu, terdapat lima daerah yang masuk dalam kelompok daerah bermasalah kesehatan yaitu Kabupaten Jember, Probolinggo, Bangkalan, Sampang dan Pamekasan, sedangkan 31 daerah lainnya masuk dalam kelompok daerah non-bermasalah kesehatan.



Gambar 1. Klasifikasi Daerah Bermasalah Kesehatan Provinsi Jawa Timur Menurut Kemenkes

Pada penelitian ini hanya digunakan dua kelompok daerah yaitu daerah bermasalah kesehatan (DBK) dan daerah non-bermasalah kesehatan (non-DBK). Wilayah yang sebelumnya pada Gambar 4.1 masuk dalam kelompok daerah bermasalah kesehatan berat (DBKB) kemudian dimasukkan dalam kelompok daerah bermasalah kesehatan (DBK). Hal tersebut dikarenakan penelitian ini menggunakan metode regresi logistik dan *geographically weighted logistic regression* (GWLR) yang variabel dependennya bersifat kategori (0 dan 1) sehingga *cut-off* yang digunakan adalah rata-rata IPKM nasional tahun 2018.

### Diagnosis Multikolinearitas

Berdasarkan Gambar 1. dapat diketahui bahwa terdapat 11 kabupaten/kota yang masuk dalam kategori daerah bermasalah kesehatan (DBK) dan 13 kabupaten/kota lainnya masuk dalam kategori non-DBK. Setelah dikelompokkan menjadi dua kategori, maka dapat dilakukan pemodelan dengan regresi logistik. Akan tetapi, sebelum melakukan pemodelan dengan regresi logistik terlebih dahulu dilakukan uji multikolinearitas untuk mendeteksi ada tidaknya korelasi antar variabel independen. Hasil yang diperoleh tertera pada Tabel 3.

Tabel 1 menunjukkan bahwa semua variabel independen yang digunakan memiliki nilai VIF < 10. yang artinya tidak terdapat multikolinearitas antar variabel independen.

### Regresi Logistik

Terdapat dua uji parameter dalam memodelkan regresi logistik, yaitu uji parameter secara serentak dan parsial.

Tabel 3. Statistik Uji Wald

Variabel	Estimasi	Wald	P-value	Odds Ratio	Keputusan
Intercept	19,52	0,49957	0,617380	2,18x10 <sup>8</sup>	Tidak Signifikan
X <sub>1</sub>	0,77233	1,93330	0,053196	2,164804	Signifikan
X <sub>2</sub>	-0,79265	-1,8069	0,070781	0,452644	Signifikan
X <sub>3</sub>	0,15279	0,10148	0,919170	1,165080	Tidak Signifikan
X <sub>4</sub>	-0,19706	-0,56783	0,570150	0,821141	Tidak Signifikan
X <sub>5</sub>	0,026207	0,11573	0,907870	1,026553	Tidak Signifikan
X <sub>6</sub>	-0,087736	-0,86285	0,388220	0,916003	Tidak Signifikan

Tabel 1. Nilai VIF Variabel Independen

Variabel	VIF
X <sub>1</sub>	3,697
X <sub>2</sub>	5,140
X <sub>3</sub>	1,183
X <sub>4</sub>	1,994
X <sub>5</sub>	1,369
X <sub>6</sub>	1,254

Tabel 2. Statistik Uji G

G <sup>2</sup>	Df	P-value
25,811	6	0,000

Pengujian parameter secara serentak dilakukan dengan statistik uji G. Pada pengujian ini, H<sub>0</sub> akan ditolak jika  $G^2 > \chi^2_{(6;0,1)} = 10,64$  atau ketika nilai  $p\text{-value} < \alpha$ . Statistik uji G untuk pengujian parameter secara serentak dapat dilihat pada Tabel 2.

Berdasarkan Tabel 2, dapat diketahui bahwa terdapat minimal satu variabel independen yang berpengaruh signifikan terhadap variabel respon secara serentak karena memiliki nilai statistik uji  $G^2 = 25,811 > \chi^2_{(6;0,1)} = 10,64$  dan  $p\text{-value} = 0,000 < \alpha = 0,1$  menunjukkan bahwa pada taraf signifikansi 0,1 terdapat minimal satu variabel independen yang berpengaruh signifikan, kemudian dilakukan pengujian parameter secara parsial dengan uji Wald.

Pada hasil estimasi parameter, dengan taraf signifikansi 0,1 akan diperoleh dua variabel independen yang berpengaruh signifikan terhadap variabel dependen yaitu X<sub>1</sub> dan X<sub>2</sub> (prevalensi balita *stunting* dan prevalensi penderita hipertensi). Berdasarkan Tabel 3 juga dapat dibentuk model regresi logistik seperti berikut:

$$\hat{\pi} = \frac{\exp(19,52 + 0,77233X_1 - 0,79265X_2 + 0,15279X_3 - 0,19706X_4 + 0,026207X_5 - 0,087736X_6)}{1 + \exp(19,5200 + 0,77233X_1 - 0,79265X_2 + 0,15279X_3 - 0,19706X_4 + 0,026207X_5 - 0,087736X_6)}$$

dengan model transformasi logitnya:

$$g(x) = 19,52 + 0,77233X_1 - 0,79265X_2 + 0,15279X_3 - 0,19706X_4 + 0,026207X_5 - 0,087736X_6$$

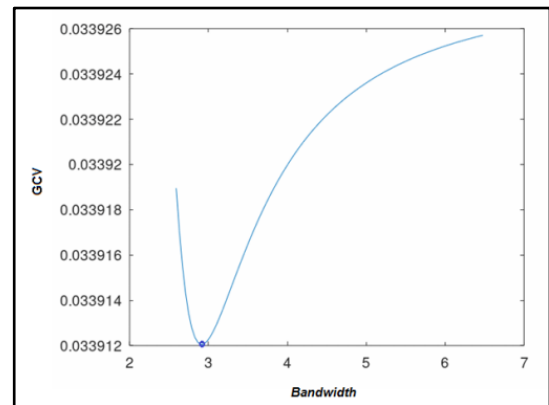
Interpretasi model logit di atas dapat dilakukan dengan cara mencari nilai *odds ratio* dari masing-masing parameter. Pada model di atas menjelaskan bahwa peningkatan prevalensi balita *stunting* sebesar 1% akan meningkatkan resiko Provinsi Jawa Timur masuk dalam kategori DBK sebesar 2,1648 kali. Apabila terjadi peningkatan prevalensi penderita hipertensi sebesar 1% akan meningkatkan resiko Provinsi Jawa Timur masuk dalam kategori DBK sebesar 0,4526 kali.

### Pembobot Spasial

Penentuan *bandwidth* optimal sangat diperlukan dalam analisis regresi yang terboboti lokasi. *Bandwidth* optimum dipilih berdasarkan nilai *Generalized Cross Validation (GCV)* paling minimum. Sebelum menentukan *bandwidth* dari masing-masing lokasi, terlebih dahulu diperlukan untuk mencari jarak *Euclidean*.

Setelah jarak *Euclidean* terbentuk, maka dapat dibentuk *bandwidth* optimum dari masing-masing pembobot. Pada penelitian ini digunakan rentang *bandwidth*

optimum pada kisaran [bwb; bwa] di mana bwb(i) atau batas *bandwidth* bawah adalah nilai maksimal dari jarak *Euclidean* pada lokasi ke-i. Pemilihan bwb ini dikarenakan pada pembobot *adaptive Kernel* terdapat ketentuan bahwa apabila nilai  $d_{ij} > h$  maka pembobot akan bernilai 0, sehingga untuk menghindari pembobot lokasi bernilai 0 maka batas bawah dari *bandwidth* adalah nilai maksimal dari jarak *Euclidean*. Sedangkan batas atas *bandwidth* (bwa)



Gambar 2. Grafik Bandwidth Optimal *Adaptive Bisquare* Lokasi Ke-19

yang digunakan dalam penelitian ini adalah 2,5 kali nilai maksimal jarak *Euclidean*. Dari interval *bandwidth* tersebut akan dipilih nilai *bandwidth* yang menghasilkan nilai *Generalized Cross Validation (GCV)* paling rendah. Grafik *bandwidth* yang terbentuk dengan pembobot *adaptive Bisquare Kernel* pada lokasi ke-19 (Kabupaten Madiun) tertera pada Gambar 2.

Lokasi ke-19 memiliki nilai *Euclidean* maksimal sebesar 2,5908 sehingga rentang *bandwidth* pada lokasi ke-19 atau Kabupaten Madiun adalah berkisar pada [2,5908; 6,477]. Pada Tabel 4

Tabel 4. Bandwidth optimal masing-masing lokasi

Lokasi	<i>Bandwidth</i> ( <i>Fixed Gaussian Kernel</i> )	<i>Bandwidth</i> ( <i>Adaptive Gaussian Kernel</i> )	<i>Bandwidth</i> ( <i>Adaptive Bisquare Kernel</i> )
1	2,2926	2,7742	3,05833
2	2,2926	3,0714	3,5056
3	2,2926	5,7806	5,7715
4	2,2926	2,6935	2,6897
⋮	⋮	⋮	⋮
38	2,2926	2,2876	2,2926

menunjukkan *bandwidth* optimal masing-masing lokasi.

Setelah terbentuk matriks *bandwidth* optimum dari masing-masing pembobot, maka dilanjutkan dengan penentuan pembobot untuk masing-masing lokasi.

Pembobot *fixed Gaussian Kernel*:

$$w_{ij} = \begin{bmatrix} 1 & 0,99173 & \dots & 0,97763 \\ 0,99173 & 1 & \dots & 0,94361 \\ \vdots & \vdots & \ddots & \vdots \\ 0,97763 & 0,94361 & \dots & 1 \end{bmatrix}$$

Pembobot *adaptive Gaussian Kernel*:

$$w_{ij} = \begin{bmatrix} 1 & 0,99538 & \dots & 0,93834 \\ 0,99434 & 1 & \dots & 0,97789 \\ \vdots & \vdots & \ddots & \vdots \\ 0,98467 & 0,96818 & \dots & 1 \end{bmatrix}$$

Pembobot *adaptive Bisquare Kernel*:

$$w_{ij} = \begin{bmatrix} 1 & 0,98584 & \dots & 0,91156 \\ 0,98142 & 1 & \dots & 0,78132 \\ \vdots & \vdots & \ddots & \vdots \\ 0,9498 & 0,90317 & \dots & 1 \end{bmatrix}$$

### Geographically Weighted Logistic Regression

Model GWLR akan terbentuk setelah pembobot untuk masing-masing lokasi ditemukan. Model akan terbentuk dengan nilai estimasi parameter yang berbeda untuk setiap lokasi. Sama halnya dengan regresi logistik global, terdapat tiga pengujian yang harus dilakukan sebelum membentuk model GWLR yaitu, pengujian parameter secara serentak, pengujian parameter parsial, dan uji kesamaan model. Pengujian parameter serentak dilakukan dengan tujuan untuk mengetahui apakah variabel independen secara bersama-sama berpengaruh signifikan terhadap variabel dependen dengan metode GWLR.

Hasil pengujian parameter secara serentak untuk masing-masing pembobot tertera pada Tabel 5.

Tabel 5 menunjukkan bahwa pada taraf signifikansi 10% terdapat pengaruh yang signifikan antara variabel independen terhadap variabel dependen secara serentak semua model GWLR baik dengan pembobot *fixed Gaussian*, *adaptive Gaussian*, maupun *adaptive Bisquare*.

Setelah dilakukan uji parameter GWLR secara serentak maka dilanjutkan dengan pengujian kesamaan model GWLR perlu dilakukan untuk melihat apakah terdapat perbedaan yang signifikan antara model GWLR dan regresi logistik global atau tidak. Pengujian dapat dilakukan dengan menggunakan uji *goodness of fit* di mana akan dilihat berdasarkan statistik uji F-nya.

Tabel 6 menunjukkan bahwa semua nilai  $F_{hitung}$  dari pembobot *fixed Gaussian*, *adaptive Gaussian*, maupun *adaptive Bisquare Kernel* memiliki nilai di bawah  $F_{(0,1;32;32)} = 1,582$  maka dapat diperoleh kesimpulan untuk tidak menolak  $H_0$  pada taraf signifikansi sebesar 10% yang artinya tidak terdapat perbedaan yang signifikan antara model regresi logistik global dengan model GWLR baik dengan pembobot *fixed Gaussian*, *adaptive Gaussian*, maupun *adaptive Bisquare*.

Sebelum melakukan pengujian parameter secara parsial, terlebih dahulu dilakukan pemilihan model GWLR terbaik. Pemilihan model ini dilakukan dengan membandingkan nilai GCV, AIC dan  $R^2$ .

Tabel 5. Pengujian Signifikansi Parameter GWLR Secara Serentak

Model	$G_2$	$\chi^2_{(6;0,1)}$	P-value	Keputusan
GWLR ( <i>fixed Gaussian</i> )	25,254		0,00141	$H_0$ ditolak
GWLR ( <i>adaptive Gaussian</i> )	25,516	10,64	0,00127	$H_0$ ditolak
GWLR ( <i>adaptive Bisquare</i> )	24,370		0,00668	$H_0$ ditolak

Tabel 6. Uji Kesamaan Model GWLR

Model	$F_{hitung}$	$F_{(0,1;32;32)}$	P-value	Keputusan
<i>Fixed Gaussian</i>	1,0122		0,4864	$H_0$ tidak ditolak
<i>Adaptive Gaussian</i>	1,0147	1,582	0,4837	$H_0$ tidak ditolak
<i>Adaptive Bisquare</i>	1,056		0,4393	$H_0$ tidak ditolak



Tabel 7. Perbandingan Model

Model GWLR	GCV	AIC	R <sup>2</sup>
<i>Fixed Gaussian Kernel</i>	0,055803	34,930	0,7127
<i>Adaptive Gaussian Kernel</i>	0,055013	34,164	0,7161
<i>Adaptive Bisquare Kernel</i>	0,05601	37,757	0,7292

Tabel 7 menunjukkan bahwa terdapat kontradiksi antara nilai AIC dan R<sup>2</sup> pada model GWLR pembobot *adaptive Gaussian Kernel* serta *adaptive Bisquare Kernel*. Nilai AIC minimum dimiliki oleh model GWLR pembobot *adaptive Gaussian Kernel*, sedangkan nilai R<sup>2</sup> maksimum dimiliki oleh model GWLR pembobot *adaptive Bisquare Kernel*. Meskipun terdapat kontradiksi, akan tetapi metode terbaik dapat dipilih dengan melihat nilai GCV masing-masing model. Nilai GCV minimum dimiliki oleh model GWLR pembobot *adaptive Gaussian Kernel* sehingga metode terbaik untuk memodelkan data IPKM Provinsi Jawa Timur tahun 2018 adalah metode GWLR dengan pembobot *adaptive Gaussian Kernel* karena memenuhi dua kriteria pemilihan model. Kemampuan variabel independen dalam menjelaskan variabel dependen dengan model GWLR pembobot *adaptive Gaussian Kernel* adalah sebesar 71,61%.

Pengujian yang selanjutnya adalah pengujian parameter dari masing-masing lokasi untuk melihat variabel mana yang berpengaruh signifikan. Model terbaik yang telah dipilih adalah model GWLR pembobot *adaptive Gaussian Kernel*, maka pengujian parameter parsial dilakukan pada model GWLR pembobot *adaptive Gaussian Kernel*.

Rangkuman dari kelompok kabupaten/kota berdasarkan variabel signifikan dari metode GWLR dengan pembobot *adaptive Gaussian Kernel* untuk memudahkan dalam membaca peta tematik pada Gambar 2 tertera pada Tabel 8 berikut.

Metode GWLR mampu menghasilkan model yang berbeda di masing-masing kabupaten dan kota di Jawa Timur sehingga pada Provinsi Jawa Timur, metode ini akan menghasilkan sebanyak 38 model yang memiliki nilai estimasi parameter yang berbeda pada setiap lokasi. Model GWLR dengan pembobot *adaptive Gaussian Kernel* untuk lokasi ke-11 (Kabupaten Bondowoso) sebagai berikut.

$$\hat{\pi}(u_{11}, v_{11}) =$$

$$\frac{\exp(19,19+0,76744X_1-0,78653X_2+0,15533X_3-0,19422X_4+0,026812X_5-0,08714X_6)}{1+\exp(19,19+0,76744X_1-0,78653X_2+0,15533X_3-0,19422X_4+0,026812X_5-0,08714X_6)}$$

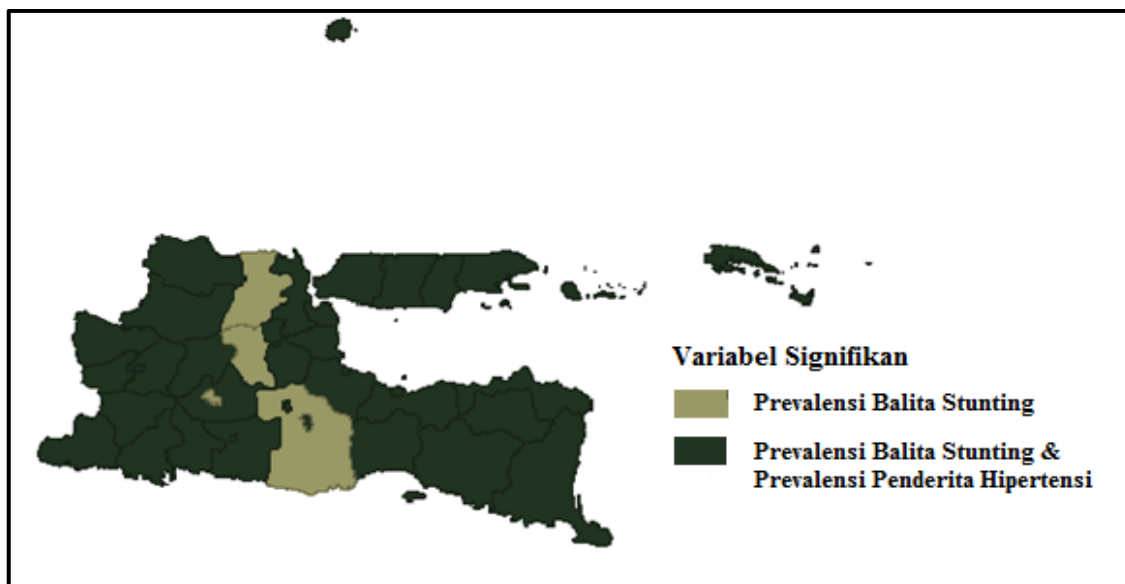
Transformasi logitnya:

$$g(x)_{(u_{11}, v_{11})} = 19,19 + 0,76744X_1 - 0,78653X_2 + 0,15533X_3 - 0,19422X_4 + 0,026812X_5 - 0,08714X_6$$

Interpretasi model untuk masing-masing lokasi dapat dilakukan dengan menghitung nilai *odds ratio* dari setiap parameter. Nilai *odds ratio* inilah yang akan memudahkan dalam memahami model yang terbentuk. Perhitungan *odds ratio*

Tabel 8. Variabel Yang Signifikan pada Model GWLR dengan Pembobot Adaptive Gaussian Kernel

Signifikan	Kabupaten/Kota
X <sub>1</sub> (prevalensi balita <i>stunting</i> )	Kediri, Malang, Jombang, dan Lamongan.
X <sub>1</sub> (prevalensi balita <i>stunting</i> ), X <sub>2</sub> (prevalensi penderita hipertensi)	Pacitan, Banyuwangi, Bondowoso, Probolinggo, Pasuruan, Gresik, Kota Blitar, Jember, Ponorogo, Trenggalek, Tulungagung, Blitar, Kediri, Malang, Lumajang, Madiun, Magetan, Bojonegoro, Tuban, Bangkalan, Sampang, Pamekasan, Situbondo, Sidoarjo, Mojokerto, Nganjuk, Ngawi, Sumenep, Kota Madiun, Kota Kediri, Kota Probolinggo, Kota Pasuruan, Kota Mojokerto, Kota Malang, Kota Surabaya, dan Kota Batu.



Gambar 3. Peta Tematik Berdasarkan Variabel Signifikan GWLR Adaptive Bisquare Kernel

dilakukan dengan mencari nilai exponential dari setiap parameter ( $\beta$ ) sehingga interpretasi model GWLR dengan pembobot *adaptive Gaussian Kernel* untuk lokasi ke-11 (Kabupaten Bondowoso) yaitu apabila terjadi peningkatan prevalensi balita *stunting* sebesar 1% akan meningkatkan resiko Kabupaten Bondowoso masuk dalam kategori DBK sebesar 2,15489 kali serta adanya peningkatan prevalensi penderita hipertensi sebesar 1% akan meningkatkan resiko Kabupaten Bondowoso masuk dalam kategori DBK sebesar 0,45542 kali.

Hasil estimasi dari GWLR tidak hanya mampu menampilkan resiko suatu wilayah masuk dalam kategori daerah bermasalah kesehatan (DBK) berdasarkan masing-masing variabel independennya,

akan tetapi metode ini juga mampu menunjukkan probabilitas suatu wilayah masuk dalam kategori DBK secara keseluruhan. Nilai probabilitas ini dilihat berdasarkan besaran nilai  $\hat{\pi}$  dari masing-masing kabupaten/kota. Nilai  $\hat{\pi}$  berada dalam rentang [0,1]. Semakin besar nilai  $\hat{\pi}$  maka probabilitas wilayah tersebut akan masuk dalam kategori daerah bermasalah kesehatan semakin tinggi. Matriks probabilitas untuk masing-masing wilayah di Provinsi Jawa Timur seperti table 9.

Nilai  $\hat{\pi}$  untuk lokasi ke-11 menunjukkan bahwa probabilitas Kabupaten Bondowoso masuk dalam kategori daerah bermasalah kesehatan adalah sebesar 0,87798 yang artinya Kabupaten Bondowoso beresiko tinggi

Tabel 9. Variabel Yang Signifikan pada Model GWLR dengan Pembobot Adaptive Gaussian Kernel

Lokasi	$\hat{\pi}$	Lokasi	$\hat{\pi}$	Lokasi	$\hat{\pi}$
1	$3,71 \times 10^{-6}$	14	0,20725	27	0,99164
2	0,001068	15	0,000438	28	0,98199
3	0,19862	16	$7,85 \times 10^{-5}$	29	0,22875
4	$2,18 \times 10^{-5}$	17	$6,00 \times 10^{-5}$	30	0,013221
5	$8,78 \times 10^{-5}$	18	$2,53 \times 10^6$	31	$2,62 \times 10^{-6}$
6	0,001414	19	$2,28 \times 10^{-5}$	32	$9,24 \times 10^{-5}$
7	$5,31 \times 10^{-5}$	20	$9,29 \times 10^{-6}$	33	0,038221
8	0,043799	21	0,41812	34	0,34006
9	0,061822	22	0,0028	35	$2,65 \times 10^{-7}$
10	0,000295	23	0,024739	36	$7,85 \times 10^{-9}$
11	0,87798	24	0,039521	37	0,001573
12	0,7421	25	$5,29 \times 10^{-5}$	38	$7,52 \times 10^{-8}$
13	0,98018	26	0,85876		

untuk masuk dalam kategori daerah bermasalah kesehatan. Nilai  $\hat{\pi}$  lokasi ke-38 menunjukkan bahwa probabilitas Kota Batu masuk dalam kategori daerah bermasalah kesehatan sangat rendah sehingga Kabupaten Bondowoso jauh lebih beresiko daripada Kota Batu.

Perhitungan akurasi, *sensitivity*, dan *specificity* pada model GWLR pembobot *adaptive Gaussian Kernel* sebagai berikut.

$$\text{Akurasi} = \frac{(31 + 6)}{31} = 0,9736$$

$$\text{Sensitivity} = \frac{31}{31} = 1$$

$$\text{Specificity} = \frac{6}{7} = 0,8571$$

Nilai akurasi model GWLR pembobot *adaptive Gaussian Kernel* sebesar 0,9736 menunjukkan bahwa kemampuan model dalam memprediksi secara benar adalah 97,36%. Nilai *sensitivity* sebesar 1 menunjukkan bahwa kemampuan model dalam memprediksi daerah non-bermasalah kesehatan dengan benar sebesar 100%. Sedangkan nilai *specificity* sebesar 0,8571 menunjukkan kemampuan model memprediksi daerah bermasalah kesehatan dengan benar sebesar 85,71%.

## KESIMPULAN DAN SARAN

Meskipun tidak terdapat perbedaan yang signifikan antara model GWLR dengan regresi logistik global, metode GWLR merupakan salah satu metode yang dapat digunakan untuk memodelkan IPKM karena dengan menggunakan metode ini dapat diketahui faktor yang mempengaruhi suatu daerah masuk dalam kategori daerah bermasalah kesehatan (DBK) secara signifikan. Metode GWLR terbaik yang dapat digunakan untuk melakukan pemodelan data IPKM Provinsi Jawa Timur tahun 2018 adalah metode GWLR dengan pembobot *adaptive Gaussian Kernel* karena memiliki nilai GCV dan AIC paling kecil dibandingkan dua pembobot lainnya di mana nilai GCV-nya sebesar 0,055013 dan AIC sebesar 34,164. Pemodelan dengan GWLR dengan pembobot *adaptive Gaussian Kernel* membentuk satu variabel yang berpengaruh signifikan secara global

dan satu variabel yang berpengaruh secara lokal. Variabel yang berpengaruh secara global adalah  $X_1$  (prevalensi balita *stunting*) sedangkan variabel yang berpengaruh signifikan secara lokal adalah  $X_2$  (prevalensi penderita hipertensi).

Pada penelitian ini digunakan 6 variabel independen yang mewakili 7 indikator dari IPKM. Pada penelitian selanjutnya, peneliti menyarankan untuk menggunakan 7 variabel independen yang mana mewakili masing-masing satu variabel dari setiap indikator dan wilayah penelitian dapat dipersempit menjadi kecamatan atau bahkan desa karena semakin kecil satuan wilayah maka penelitian akan semakin tepat sasaran. Selain itu, penelitian menggunakan metode GCV untuk mencari *bandwith* optimal, pada penelitian selanjutnya dapat menggunakan metode pencarian *bandwidth* optimum lain seperti *golden section search*, CV, maupun AICc yang diharapkan dapat membentuk model GWLR yang berbeda signifikan dengan regresi logistik global.

## DAFTAR PUSTAKA

- Aji, Caraka. A. W. 2014. Analisis Faktor-Faktor yang Mempengaruhi Laju Pertumbuhan Penduduk Kota Semarang Tahun 2011 Menggunakan Geographically Weighted Logistic Regression. *Skripsi*. Semarang.
- Albuquerque, P., Median, F. dan Silva, A. 2016. Geographically Weighted Logistic Regression Applied to Credit Scoring Models. *Revista Contabilidade and Financas*. Vol. 28 (73), 93-112.
- Atkinson, P. M., German, S. E., Sear, D. A., & Clark, M. J. (2003). Exploring the relations between riverbank erosion and geomorphological controls using geographically weighted logistic regression. *Geographical Analysis*, Vol. 35(1), 58-82.
- Caraka, Rezzy. E. dan Yasin, Hasbi. 2017. *Geographically Weighted Regression (GWR)*. Yogyakarta: Mobius.
- Desriwendi, Hoyyi, A. dan Wuryandari, T. 2015. Pemodelan *Geographically Weighted Logistic Regression (GWLR)* dengan Fungsi Pembobot Fixed

- Gaussian Kernel dan Adaptive Gaussian Kernel. *Jurnal Gaussian*. Vol. 4 (2), 193-204.
- Eubank, Randall. 1988. Spline Smoothing and Nonparametric Regression. *Journal of the American Statistical Association*. Vol. 96 (456), 1523-1524.
- Faturrahman, M. 2019. Pemodelan Indeks Pembangunan Kesehatan Masyarakat Kabupaten/Kota Di Pulau Kalimantan Menggunakan Pendekatan Regresi Probit. *Jurnal Varian*. Vol. 2 (2), 47-54.
- Faturrahman, M., Purhadi, Sutikno, dan Ratnasari, V. 2016. Pemodelan Geographically Weighted Logistic Regression pada Indeks Pembangunan Kesehatan Masyarakat di Provinsi Papua. *Prosiding Seminar Nasional MIPA*. 34-42.
- Fotheringham, A. S., Lu, B., Charlton, M. dan Harris, P. 2002. Geographically Weighted Regression with a non-Euclidean Distance Metric: a Case Study Using Hedonic House Price Data. *International Journal of Geographical Information Science*. <http://dx.doi.org/10.1080/13658816.2013.865739> (Diakses 25 September, 2020).
- Hosmer, David. W. dan Lemeshow, Stanley. 1989. *Applied Logistic Regression*. Edisi Kedua. New York: A Willey-Interscience Publication.
- Kemenkes. 2011. Buku Saku Penanggulangan Daerah Bermasalah Kesehatan. Jakarta: Kemenkes.
- Kemenkes. 2019. *IPKM 2018*. Jakarta: Lembaga Penerbit Badan Penelitian dan Pengembangan Kesehatan (LPB).
- Kemenkes. 2019. Profil Kesehatan Provinsi Jawa Timur Tahun 2018. Surabaya: Kemenkes.
- Lestari, Vivi. D. 2020. Model Geographically Weighted Logistic Regression dengan Fungsi Pembobot Adaptive Tricube. *Skripsi*. Samarinda.
- Octavany, M. A., Budiantara dan Ratnasari. 2017. Pemodelan Faktor-Faktor yang Mempengaruhi Indeks Pembangunan Kesehatan Masyarakat Provinsi Jawa Timur Menggunakan Pendekatan Regresi Semiparametrik Spline. *Jurnal Sains dan Seni ITS*. Vol. 6 (1), 122-128.
- Sabat, T., Meilanani, S. dan Purnomo, W. (2017). Spatial Modeling of Infant Mortality Rate In South Central Timor Regency Using GWLR Method With Adaptive Bisquare Kernel and Gaussian Kernel. *Health Nations*. Vol. 1 (2), 90-98.
- Suyitno, Purhadi, Sutikno, dan Irhamah. (2016). Parameter Estimation of Geographically Weighted Trivariate Weibull Regression Model. *Applied Mathematical Science*. Vol. 10(18), 861-878.
- Yrigoyen, C. C., Rodriguez, I. G. dan Otero, J. V. 2007. Modeling Spatial Variations In Household Disposable Income With Geographically Weighted Regression. *MPRA Paper*. No. 1682.