

Efektifitas Model Regresi OLS (*Ordinary Least Square*) dan *Geographically Weighted Regression* (GWR) pada Indeks Pembangunan Manusia (IPM) di Provinsi Jawa Timur

Nur Azizah¹, Prof. Henny Pramoedyo²

^{1,2}Jurusan Statistika, Program Studi Magister Statistika, Universitas Brawijaya Malang
nur.azizah.1907@gmail.com, pramoedyohp@yahoo.com

Info Artikel

Riwayat Artikel:

Diterima: 21 Oktober 2019
Direvisi: 18 November 2019
Diterbitkan: 15 Januari 2020

Kata Kunci:

IPM (Indeks Pembangunan Manusia)
Regresi OLS (*Ordinary Least Square*)
GWR (*Geographically Weighted Regression*)
Adaptive Gaussian Kernel
AIC (*Akaike's Information Criterion*)

ABSTRAK

Indeks Pembangunan Manusia (IPM) merupakan indikator penting untuk menilai keberhasilan dalam upaya membangun kualitas hidup manusia (masyarakat/penduduk). Pembangunan manusia di Jawa Timur mengalami peningkatan selama periode 2011 hingga 2016. Oleh karena itu, peningkatan IPM tersebut perlu dijaga agar IPM di Provinsi Jawa Timur lebih meningkat. Peningkatan IPM berdampak kepada peningkatan kualitas hidup masyarakat di Provinsi Jawa Timur. Guna meningkatkan IPM, faktor-faktor yang berpengaruh terhadap peningkatan IPM perlu ditingkatkan. Penelitian ini bertujuan untuk memodelkan IPM di Provinsi Jawa Timur tahun 2016 menggunakan model regresi linier OLS (*Ordinary Least Square*) dan *Geographically Weighted Regression* (GWR), mengetahui variabel yang mempengaruhi IPM dan membandingkan efektifitas kedua model tersebut. Hasil penelitian disimpulkan bahwa metode regresi OLS lebih efektif dalam memodelkan IPM dibandingkan dengan model GWR berdasarkan nilai Akaike's Information Criterion (AIC). Nilai AIC pada regresi OLS (77,78) lebih rendah dibandingkan dengan model GWR (190,5837). Variabel independen yang mempengaruhi nilai IPM berdasarkan regresi OLS yaitu angka partisipasi kasar SMU (X_1), jumlah rata-rata pengeluaran per kapita sebulan untuk makanan (X_2) dan angka kematian bayi (X_3) dengan nilai R^2 sebesar 78,86%. Hal ini menunjukkan bahwa keragaman yang dapat dijelaskan oleh model regresi OLS pada IPM di Jawa Timur tahun 2016 sebesar 78,86%.

Copyright © 2019 SIMANIS.
All rights reserved.

Korespondensi:

Nur Azizah,
Jurusan Statistika, Program Studi Magister Statistika
Universitas Brawijaya,
Jl. Veteran Malang, Jawa Timur, Indonesia 65145
nur.azizah.1907@gmail.com

1. PENDAHULUAN

Indeks Pembangunan Manusia (IPM) merupakan indikator yang penting guna (1) menilai keberhasilan dalam upaya membangun kualitas hidup masyarakat, (2) menjelaskan bagaimana masyarakat dapat mengakses hasil pembangunan dalam memperoleh pendapatan, kesehatan, pendidikan dan lain-lain serta (3) mengamati perkembangan pembangunan di suatu daerah dalam jangka panjang. Berdasarkan data dari BPS, diketahui bahwa pembangunan manusia di Jawa Timur mengalami peningkatan selama periode 2011 hingga 2016. Data yang digunakan pada penelitian ini menggunakan data sekunder yang diperoleh dari BPS Jawa

Timur tahun 2016 berupa IPM dari semua Kabupaten/Kota di Provinsi Jawa Timur. Oleh karena itu, guna mengetahui faktor-faktor apa dan seberapa besar pengaruh faktor tersebut terhadap IPM di Provinsi Jawa Timur tahun 2016, perlu dilakukan analisis regresi *Ordinary Least Square* (OLS). Di sisi lain, data penelitian ini diperoleh dari berbagai Kabupaten/Kota yang secara geografis tersebar di Indonesia. Sehingga adanya keberagaman antar lokasi dalam suatu wilayah atau heterogenitas spasial pada data yang diperoleh bisa terjadi. Oleh karena itu, diperlukan adanya analisis regresi yang mempertimbangkan faktor heterogenitas spasial yaitu metode *Geographically Weighted Regression* (GWR). Penelitian ini dilakukan untuk memodelkan IPM dengan menggunakan pendekatan analisis regresi OLS dan GWR, mendapatkan faktor-faktor yang mempengaruhi IPM serta membandingkan efektivitas kedua model tersebut.

Penelitian tentang IPM di Provinsi Jawa Timur telah banyak dilakukan sebelumnya antara lain pengaruh faktor pendidikan (rasio guru-siswa SMP/MTs, rasio sekolah-murid SMP/MTs, angka partisipasi SMP/MTs), faktor kesehatan (jumlah sarana kesehatan dan rumah tangga dengan akses air bersih), faktor kependudukan (kepadatan penduduk, tingkat partisipasi kerja dan PDRB per kapita) terhadap IPM menggunakan regresi panel [1], analisis pengaruh IPM (menggunakan variabel harapan lamanya hidup dan saat lahir serta tingkat pendidikan) terhadap pertumbuhan ekonomi melalui analisis regresi linier berganda [2], pengaruh variabel Produk Domestik Regional Bruto, Tingkat Pengangguran, Jumlah Penduduk Miskin dan Gini Ratio terhadap IPM menggunakan analisis regresi data panel [3], pengaruh faktor kesehatan (angka kematian bayi, prosentase keluhan kesehatan, jumlah sarana kesehatan), faktor pendidikan (angka buta huruf, angka partisipasi sekolah (SMA), rasio guru-siswa (SMA) dan rasio sekolah-murid (SMA), faktor ekonomi (prosentase penduduk miskin, PDRB perkapita, pertumbuhan ekonomi, prosentase penduduk usia 15 tahun ke atas yang bekerja, tingkat pengangguran terbuka dan tingkat partisipasi angkatan kerja) menggunakan analisis regresi logistik *ridge* [4]. Sedangkan pada penelitian ini, variabel-variabel yang digunakan yaitu angka partisipasi kasar SMU (X_1), jumlah rata-rata pengeluaran per kapita sebulan untuk makanan (X_2), angka kematian bayi (X_3) dan jumlah Puskesmas (X_4) dengan menggunakan analisis regresi dan GWR.

Penjelasan yang berkaitan dengan model Regresi OLS, Analisis GWR dan pemilihan model terbaik, berturut-turut akan dijelaskan pada Bab 1.1, Bab 1.2 dan Bab 1.3.

1.1. Model Regresi OLS

Model regresi OLS dapat digunakan untuk menemukan hubungan terbaik antara variabel independen X dan variabel dependen Y , mengukur kekuatan hubungan dan melakukan prediksi terhadap variabel Y melalui variabel X yang diamati [5]. Metode pendugaan parameter pada regresi OLS menggunakan Metode Kuadrat Terkecil (MKT). Model analisis regresi linier berganda digambarkan dengan

$$Y = \beta_0 + \sum_{i=1}^p \beta_i X_i \quad (1)$$

di mana:

- Y : variabel dependen (variabel respons)
- X : variabel independen (variabel prediktor)
- p : banyaknya variabel yang digunakan
- β_0 : konstanta (*intersept*)
- β_i : parameter pada peubah ke k , di mana $i : 1, 2, \dots, p$.

Sehingga, penduga parameter pada regresi linier berganda diperoleh dari matriks:

$$\hat{\beta} = (X^T X)^{-1} (X^T Y) \quad (2)$$

Sedangkan varians (keragaman) bagi penduga parameter $\hat{\beta}$ yaitu

$$\text{var}(\hat{\beta}) = \sigma^2 (X^T X)^{-1} \quad (3)$$

$$\text{di mana } \sigma^2 = \frac{(Y_i - X_i' \hat{\beta})' (Y_i - X_i' \hat{\beta})}{n - k} \quad (4)$$

1.1.1. Uji Asumsi Klasik

Terdapat beberapa asumsi klasik pada analisis regresi OLS sebagaimana penjelasan berikut [6]:

1. Asumsi normalitas

Hipotesis penelitian:

H_0 : Data sisaan menyebar normal

H_1 : Data sisaan tidak menyebar normal

Sisaan diperoleh dari

$$\hat{u}_i = Y_i - \hat{Y}_i$$

Pengujian asumsi dilakukan dengan menggunakan uji statistika *Jarque-Bera* dengan persamaan di bawah ini

$$JB = n \left(\frac{s^2}{6} + \frac{(k-3)^2}{24} \right) \quad (5)$$

Dimana

n : banyaknya sampel

S : *skewness*, merupakan ukuran kemenjuluran suatu kurva.

K : kurtosis, merupakan ukuran keruncingan dari suatu kurva.

Data sisaan dikatakan menyebar norma, ketika S bernilai 0 dan K bernilai 3.

Keputusan: Tolak H_0 jika $p\text{-value} \leq \alpha$ (taraf nyata) atau statistik uji $JB > \chi^2$.

2. Asumsi heterogenitas

Hipotesis penelitian:

H_0 : Tidak terdapat heteroskedastisitas ragam sisaan.

H_1 : Terdapat heteroskedastisitas ragam sisaan (asumsi kehomogenan ragam tidak terpenuhi).

Pengujian asumsi dilakukan dengan menggunakan uji *Glejser* pada penelitian ini menggunakan fungsi persamaan berikut:

$$|\hat{u}_i| = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_p X_p \quad (6)$$

Di mana $|\hat{u}_i|$ merupakan nilai absolut dari nilai sisaan pada regresi OLS.

Keputusan: Tolak H_0 jika $p\text{-value} \leq \alpha$ (taraf nyata).

3. Asumsi non multikolinieritas.

Asumsi ini dilakukan untuk melihat keeratan hubungan antar variabel X melalui nilai VIF (*Variance Inflation Factor*) yang diperoleh. Asumsi non multikolinieritas tidak terpenuhi jika nilai VIF yang diperoleh lebih besar dari 10 ($VIF > 10$).

Pengujian VIF diperoleh dari

$$VIF = \frac{1}{1-R_j^2} \quad (7)$$

nilai R_j^2 merupakan koefisien determinasi (R^2) pada regresi X_i sebanyak $p-2$ peubah, di mana penjelasan mengenai (R^2) dijelaskan pada Bab 1.1.3.

1.1.2. Pengujian Signifikansi

Pengujian signifikansi dilakukan untuk mengetahui apakah variabel-variabel independen yang digunakan berpengaruh atau tidak terhadap variabel dependen baik secara parsial maupun serentak. Berikut penjelasan masing-masing pengujian tersebut.

1.1.2.1. Pengujian secara parsial

Uji signifikansi secara parsial dilakukan dengan menggunakan statistik uji t.

Hipotesis: $H_0 : \hat{\beta}_i = \beta_i$ vs $H_1 : \hat{\beta}_i \neq \beta_i$, di mana $i = 0, 1, 2, 3, \dots, p$.

Statistik uji t:

$$t = \frac{\hat{\beta}_i - \beta_i}{s.e.[\hat{\beta}_i]} \quad (8)$$

Jika statistik uji $|t| > t$ (db = $n-i$) atau $p\text{-value} \leq \alpha$, maka:

Keputusan: tolak H_0 .

Kesimpulan: variabel X_i secara parsial berpengaruh terhadap Y .

1.1.2.2. Pengujian secara serentak

Uji signifikansi secara serentak dilakukan dengan menggunakan statistik uji F dengan hipotesis berikut [7]:

$H_0 : \beta_0 = \beta_1 = \beta_2 = \dots = \beta_p = 0$

H_1 : paling tidak terdapat sepasang β_i yang tidak sama dengan nol di mana $i = 0, 1, 2, 3, \dots, p$.

Berikut statistik uji F untuk pengujian signifikansi secara serempak [7]:

$$F = \frac{JKR/(p-1)}{JKG/(n-p)} \quad (9)$$

Di mana:

JKR (Jumlah Kuadrat Regresi) = $\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$

JKG (Jumlah Kuadrat Galat) = $\sum_{i=1}^n (Y_i - \hat{Y}_i)^2$

Jika statistik uji $F > F$ (dbregresi, dbgalat) di mana db regresi = $p-1$ sedangkan db galat = $n-p$ atau $p\text{-value} < \alpha$

Keputusan: Tolak H_0 .

Kesimpulan: variabel X_1, X_2, \dots, X_p secara bersama-sama, berpengaruh terhadap Y .

1.1.3. Koefisien Determinasi

Koefisien determinasi (R^2) mengukur proporsi keragaman atau variasi total variabel respon yang dapat dijelaskan oleh variabel prediktor [8]. Besar nilai R^2 dapat diperoleh dari

$$R^2 = \frac{JK_{regresi}}{JK_{total}} = \frac{\sum_{i=1}^n (\bar{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (10)$$

Nilai R^2 berkisar antara $0 \leq R^2 \leq 1$. Semakin tinggi nilai R^2 maka semakin tinggi kekuatan variabel prediktor dalam menjelaskan variabel respon dalam analisis regresi. Sehingga dapat diperoleh prediksi variabel respon yang lebih baik [9].

1.2. Analisis GWR

Pada model GWR, jarak antar lokasi digunakan sebagai acuan dalam pendugaan parameter. Adapun penjelasan model GWR pada notasi matriks sebagai berikut [10]:

$$y_i = X_i \beta_i + \varepsilon_i \quad (11)$$

dimana:

β_i : vektor kolom dari koefisien regresi.

X_i : vektor baris dari variabel penjelas pada lokasi i .

Metode pendugaan parameter pada model GWR dilakukan dengan menggunakan metode *Weighted Least Square* (WLS) sehingga diperoleh vektor koefisien regresi pada lokasi i yaitu

$$\hat{\beta}_i = [X^T W_i X]^{-1} X^T W_i Y \quad (12)$$

dimana:

Y : vektor variabel respon berukuran $n \times 1$.

X : $[X_1^T, X_2^T, \dots, X_n^T]^T$ merupakan matriks variabel penjelas (kolom pertama untuk intersep)

W_i : diagonal $[W_{i1}, W_{i1}, \dots, W_{in}]$ merupakan matriks pembobot diagonal $n \times n$ untuk setiap lokasi i .

$\hat{\beta}_i$: $[\hat{\beta}_{i0}, \hat{\beta}_{i1}, \dots, \hat{\beta}_{ip-1}]^T$ merupakan vektor p koefisien regresi lokal pada lokasi i untuk $p-1$ variabel penjelas dan sebuah intersep.

Dasar penentuan bobot pada masing-masing pengamatan terhadap model regresi di masing-masing lokasi dapat dilakukan dengan menggunakan *bandwith* yang berupa lingkaran dengan radius h [11]. Semakin dekat suatu pengamatan dengan lokasi i , maka pengamatan tersebut akan mempunyai pengaruh yang lebih besar pada model regresi yang terbentuk pada lokasi i tersebut. Jika *bandwith* yang digunakan terlalu besar akan menghasilkan *oversmoothing* dan jika nilainya terlalu kecil akan menghasilkan *undersmoothing*. Dengan demikian, penentuan *bandwith* pada masing-masing lokasi merupakan hal penting. Semakin dekat lokasi tersebut, maka pembobot akan semakin besar begitu juga sebaliknya. Metode penentuan nilai *bandwith* yang paling banyak digunakan adalah *Cross Validation* (CV) [12]. *Bandwith* yang optimum diperoleh melalui proses iterasi *Golden Section Research*. Berikut persamaan *bandwith kernel* ($\hat{\gamma}$) yang diestimasi menggunakan fungsi CV yang dapat meminimalkan RMSPE (*Root Mean Squared Prediction Error*):

$$\hat{\gamma} = \operatorname{argmin} \sum_{i=1}^n [y_i - \hat{y}_i(\gamma)]^2 \quad (13)$$

Di mana

\hat{y}_i : nilai prediksi dari pengamatan t dengan lokasi kalibrasi l yang keluar dari kumpulan estimasi data

$\hat{\gamma}$: nilai *bandwith Kernel* yang meminimalkan RMSPE.

Metode CV cocok digunakan dalam menentukan *bandwith* secara optimal ketika menggunakan matriks pembobot berupa fungsi *Adaptive Gaussian Kernel* [11].

1.2.1. Matriks Pembobot dengan Fungsi Adaptive Gaussian Kernel

Fungsi *Adaptive Gaussian Kernel* cocok digunakan untuk pengamatan yang tersebar pada pola yang tidak beraturan serta mampu menyesuaikan kondisi titik-titik pengamatan [11]. Penelitian ini menggunakan fungsi pembobot *Adaptive Gaussian Kernel*. Adapun fungsi *Adaptive Gaussian Kernel* dapat dituliskan sebagai berikut:

$$w_{ij} = \exp \left[-\frac{1}{2} \left(\frac{d_{ij}}{\gamma_i} \right)^2 \right] \quad (14)$$

di mana

γ_i : parameter *bandwidth Kernel*

d_{ij} : jarak *Euclidian* antara pengamatan pada titik ke- i dan ke- j yang diperoleh dari persamaan berikut

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2} \quad (15)$$

di mana:

(u_i, v_i) : titik koordinat pada lokasi ke- i .

(u_j, v_j) : titik koordinat pada lokasi ke- j .

1.2.2. Uji Heterogenitas Spasial

Uji ini dilakukan untuk melihat adanya keragaman antar lokasi yang disebabkan karena suatu lokasi memiliki struktur dan parameter hubungan yang berbeda dengan lokasi lainnya, melalui statistik uji *Breusch-Pagan* (BP). Adapun hipotesis yang digunakan yaitu [11] :

$H_0 : \sigma^2_{(u_1, v_1)} = \sigma^2_{(u_2, v_2)} = \dots = \sigma^2_{(u_n, v_n)} = \sigma^2$, yang berarti bahwa tidak terdapat heterogenitas spasial atau keragaman antar lokasi sama.

H_1 : minimal ada satu $\sigma^2_{u_i, v_i} \neq \sigma^2$, yang berarti bahwa terdapat heterogenitas spasial atau keragaman antar lokasi berbeda.

Berikut persamaan pada statistik uji BP :

$$BP = \left(\frac{1}{2}\right) f^T Z(Z^T Z)^{-1} Z^T f + \left(\frac{1}{T}\right) \left[\frac{e^T W e}{\sigma^2}\right]^2 \quad (16)$$

Di mana

Elemen vektor f adalah $f_i = \frac{e_i}{\sigma^2} - 1$

e_i : galat pada pengamatan ke- i

e : vektor galat e_i

σ^2 : ragam galat e_i

T : $Tr [W^T W + W^2]$

Jika $p\text{-value} \leq \alpha$ (taraf nyata) atau statistik uji BP $> \chi^2$

Keputusan : Tolak H_0 .

Kesimpulan: terdapat keragaman atau heterogenitas spasial. Jika terdapat heterogenitas spasial, maka analisis GWR dapat digunakan. Begitu juga sebaliknya.

1.2.3. Pengujian Signifikansi Pendugaan Parameter Model GWR

Hasil pendugaan parameter model GWR yang telah diperoleh perlu diuji signifikansinya untuk melihat apakah pendugaan parameter tersebut berpengaruh secara nyata atau tidak terhadap variabel dependen. Hipotesis yang digunakan untuk pengujian signifikansi secara parsial yaitu [11]:

$H_0: \beta_k(u_i, v_i) = 0$

$H_1: \beta_k(u_i, v_i) \neq 0$; untuk $k = 0, 1, 2, \dots, p$.

Pengujian ini dilakukan secara parsial dengan menggunakan statistik uji t sebagai berikut :

$$t = \frac{\hat{\beta}_k(u_i, v_i) - \beta_k(u_i, v_i)}{\hat{\sigma} \sqrt{c_{kk}}} \quad (17)$$

di mana:

Penduga parameter $\beta(u_i, v_i)$ mengikuti sebaran normal dan mengikui matriks kovarian CC^T , di mana matriks C yaitu

$$C = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i)$$

Jika nilai statistik uji $|t| > t_{\left(\frac{\alpha}{2}, n-k-1\right)}$ atau $p\text{-value} \leq \alpha$, maka keputusan H_0 ditolak atau pendugaan parameter berpengaruh secara nyata.

1.2.4. Pengujian Goodness of Fit Model GWR

Pengujian *Goodness Of Fit* pada model GWR dilakukan dengan menggunakan uji F dengan hipotesis di bawah ini [11]:

H_0 : parameter regresi konstan untuk semua titik di dalam area penelitian (regresi OLS lebih baik digunakan)

H_1 : parameter regresi bervariasi secara spasial (GWR lebih baik digunakan)

Pendugaan parameter pada GWR dilakukan dengan menggunakan persamaan matriks berikut:

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

Jika

$$S_1 = X_i(X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i)$$

Maka

$$\hat{Y} = S_1 Y$$

S_1 adalah *hat matrix* bagi model regresi GWR berdimensi $n \times n$, yang tidak tergantung pada nilai Y . *Hat matrix* tersebut digunakan untuk mendapatkan nilai JKG dari model regresi baik OLS maupun GWR. Sehingga, sisaan pada operasi GWR diperoleh

$$\hat{\epsilon}_z = (I - S_z) Y$$

Dengan:

$z = 0$ dan 1

I : matriks identitas.

Sehingga, perhitungan JKG diperoleh dengan persamaan

$$\hat{\epsilon}_z^T \hat{\epsilon}_z = Y^T (I - S_z)^T (I - S_z) Y$$

Jika dimisalkan

$$R_z = (I - S_z)^T (I - S_z)$$

maka

$$\varepsilon_z^T \varepsilon_z = Y^R R_z Y$$

Di mana R_z berdimensi $n \times n$. Notasi indeks z digunakan untuk menunjukkan 0 sebagai perhitungan dengan regresi OLS dan 1 perhitungan dengan GWR. Dengan demikian, uji *Goodness of Fit* model GWR diperoleh sebagai berikut

$$F_{hit} = \left[\frac{(Y^T R_0 Y) - (Y^T R_1 Y)}{v} \right] \left[\frac{(Y^T R_1 Y)}{\delta} \right]^{-1} \quad (18)$$

dimana :

$$v = Tr(R_0 - R_1) \text{ dan } \delta = Tr(R_1)$$

Statistik uji ini mengikuti sebaran F dengan db1 sebesar $\frac{v^2}{v}$ dan db2 sebesar $\frac{\delta^2}{\delta}$.

Notasi $v' = Tr[R_0 - R_1]$ dan $\delta' = Tr[R_1^2]$.

Jika statistik uji $F > F_{(\alpha; db1, db2)}$ atau jika $p\text{-value} \leq \alpha$, maka:

Keputusan: Tolak H_0

Kesimpulan: model GWR lebih baik digunakan pada data penelitian.

1.3. Pemilihan Model Terbaik

Pemilihan model terbaik antara model yang diperoleh melalui analisis regresi OLS dan analisis GWR, dilakukan berdasarkan nilai AIC. Nilai AIC diperoleh dari persamaan berikut [13]:

$$AIC = -2 \text{LogLikelihood} + 2p = -2 \sum_{i=1}^p \log(f(X_i | \hat{\beta})) + 2p \quad (19)$$

Semakin kecil nilai AIC, maka semakin baik model tersebut dalam memodelkan data penelitian. Dengan kata lain, semakin efektif model tersebut dalam menggambarkan data penelitian.

2. METODE PENELITIAN

Penerapan metode ini dilakukan dengan menggunakan data sekunder dari BPS berupa data IPM di Kabupaten/Kota Provinsi Jawa Timur tahun 2016. Model analisis yang digunakan yaitu metode analisis regresi berganda dan metode GWR.

1. Melakukan analisis deskriptif terhadap pemetaan IPM di Provinsi Jawa Timur tahun 2016.
2. Melakukan analisis regresi OLS

Langkah-langkah analisis regresi yang dilakukan sebagai berikut:

 - a) Menduga parameter pada regresi OLS melalui metode MKT dengan menggunakan persamaan (2).
 - b) Menguji signifikansi penduga parameter analisis regresi secara simultan dan parsial menggunakan persamaan (8) dan (9).
 - c) Melakukan pengujian asumsi klasik analisis regresi berupa:
 - Asumsi normalitas sisaan menggunakan statistik uji *Jarque-Bera* pada persamaan (5).
 - Asumsi kehomogenan ragam sisaan menggunakan uji *Glejser* pada persamaan (6).
 - Asumsi non multikolinieritas berdasarkan nilai VIF pada persamaan (7).
 - d) Interpretasi hasil analisis regresi berganda.
3. Melakukan analisis GWR

Langkah-langkah analisis GWR yang dilakukan antara lain:

 - a) Menguji asumsi heterogenitas spasial dengan menggunakan uji *Breush Pagan (BP)* pada persamaan (16).
 - b) Menentukan model GWR dengan cara berikut:
 - Menghitung jarak *Euclidean* antara lokasi pengamatan berdasarkan posisi geografis (nilai UTM) berdasarkan persamaan (15).
 - Menggunakan *Cross Validation (CV)* untuk menentukan *bandwith* paling optimal berdasarkan persamaan (13).
 - Menghitung matriks pembobot dengan menggunakan metode *Adaptive Gaussian Kernel* berdasarkan persamaan (14).
 - Melakukan pendugaan parameter GWR dengan menggunakan metode *WLS*.
 - Melakukan signifikansi pendugaan parameter model GWR dengan menggunakan persamaan (17).
 - Pengujian *goodness of fit* model GWR menggunakan persamaan (18).
 - Interpretasi model GWR.
 4. Melakukan pemilihan model terbaik berdasarkan nilai AIC pada persamaan (19).

Penelitian ini menggunakan *software R* untuk melakukan analisis regresi OLS dan GWR. Sedangkan pemetaan IPM, dilakukan melalui *software ArcGIS*.

3. HASIL DAN PEMBAHASAN

3.1. Pemetakan IPM di Provinsi Jawa Timur Tahun 2016

Klasifikasi IPM menurut BPS dibagi menjadi 4 tingkatan, yaitu klasifikasi IPM kategori sangat tinggi (jumlah IPM ≥ 80), kategori tinggi (70 ≤ IPM < 80), kategori sedang (60 ≤ IPM < 70) dan kategori rendah (IPM < 60).



Gambar 1. Pemetakan IPM di Provinsi Jawa Timur Tahun 2016

Berdasarkan hasil pemetakan IPM di Provinsi Jawa Timur tahun 2016 pada Gambar 1, terlihat bahwa terdapat 3 Kabupaten/Kota dengan kategori IPM sangat tinggi adalah Kota Malang, Kota Surabaya, dan Madiun. Sedangkan kategori rendah terletak di daerah Sampang. Kategori IPM tinggi adalah Jombang, Lamongan, Nganjuk, Tulungagung, Mojokerto, Kota Probolinggo, Magetan, Kota Batu, Kota Pasuruan, Gresik, Kota Kediri, Kota Mojokerto, Kota Blitar dan Sidoarjo. Sedangkan Kabupaten/Kota lainnya yang tidak termasuk ke dalam ketiga kategori tersebut termasuk dalam kategori IPM sedang.

3.2. Analisis Regresi OLS

3.2.1. Hasil Uji Pendugaan Parameter pada Regresi OLS

Persamaan pada model analisis regresi OLS pada penelitian ini yaitu

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_4 X_4 \tag{20}$$

Hasil pendugaan parameter menggunakan analisis regresi OLS berdasarkan persamaan (20) disajikan pada Tabel 1.

Tabel 1. Penduga Parameter dan Uji Signifikansi pada Regresi OLS

Penduga Parameter	Koefisien Penduga Parameter	p-value pada Statistik Uji t	p-value pada Statistik Uji F
$\hat{\beta}_0$	55,21	$5,35 \times 10^{-11}***$	
$\hat{\beta}_1$	0,1069	0,0831*	
$\hat{\beta}_2$	$2,872 \times 10^{-5}$	$1,08 \times 10^{-5}***$	$1,025 \times 10^{-10}***$
$\hat{\beta}_3$	-0,1435	0,0226**	
$\hat{\beta}_4$	-0,04648	0,1900 ^{ts}	

Keterangan: ***: sangat signifikan pada taraf nyata 0,01%; **: signifikan pada taraf nyata 2,5%; *: signifikan pada taraf nyata 8,5%; ts: tidak signifikan

Berdasarkan Tabel 1, model analisis regresi OLS yang diperoleh pada penelitian ini adalah

$$\hat{Y} = 55,21 + 0,1069 X_1 + 2,872 \times 10^{-5} X_2 - 0,1435 X_3 - 0,04648 X_4 \tag{21}$$

Hasil pengujian pendugaan parameter yang tertera pada Tabel 1 menunjukkan bahwa angka partisipasi kasar SMU (X_1), jumlah rata-rata pengeluaran per kapita sebulan untuk makanan (X_2), angka kematian bayi (X_3) dan jumlah Puskesmas (X_4) secara simultan berpengaruh sangat nyata terhadap IPM di Provinsi Jawa Timur tahun 2016 dengan tingkat kesalahan sebesar 0,01%.

Secara parsial, angka partisipasi kasar SMU (X_1) berpengaruh nyata terhadap IPM di Provinsi Jawa Timur tahun 2016 dengan tingkat kesalahan sebesar 8,5%. Rata-rata pengeluaran per kapita sebulan untuk makanan (X_2) secara parsial juga berpengaruh sangat nyata terhadap IPM di Provinsi Jawa Timur tahun 2016 dengan tingkat kesalahan sebesar 0,01%. Begitu pula dengan angka kematian bayi (X_3) secara parsial berpengaruh nyata terhadap IPM di Provinsi Jawa Timur tahun 2016 dengan tingkat kesalahan sebesar 2,5%. Sedangkan jumlah Puskesmas (X_4) secara parsial tidak berpengaruh nyata terhadap IPM dengan tingkat kesalahan 5%.

Informasi yang dapat disimpulkan pada persamaan (21) yaitu IPM akan naik sebesar 0,1069% apabila angka partisipasi kasar SMU (X_1) meningkat sebesar 1% serta variabel lain tidak mengalami perubahan, IPM akan naik sebesar 0,2872 % apabila rata-rata pengeluaran per kapita sebulan untuk makanan (X_2) naik sebesar Rp. 10.000 dengan asumsi kondisi variabel lainnya konstan, dan IPM akan turun sebesar 0,1435 % apabila angka kematian bayi (X_3) naik sebesar 1% serta variabel lain tidak mengalami perubahan.

3.2.2. Hasil Uji Asumsi pada regresi OLS

Hasil uji asumsi pada analisis regresi OLS ditampilkan pada Tabel 2.

Tabel 2. Hasil Uji Asumsi pada Regresi OLS

Asumsi	Metode	Hasil Uji	Kesimpulan
Normalitas sisaan	<i>Jarque-Bera</i>	<i>p-value</i> : 0,9919	Asumsi terpenuhi
Kehomogenan ragam sisaan	<i>Glejser</i>	<i>p-value</i> : 0,0005109	Asumsi tidak terpenuhi
Multikolinieritas antar variabel X	VIF	VIF pada semua variabel $X < 10$	Asumsi terpenuhi

Berdasarkan Tabel 2, terlihat bahwa asumsi sisaan menyebar normal dan tidak terjadi multikolinieritas (tidak terdapat hubungan antar variabel X) telah terpenuhi. Akan tetapi, berdasarkan hasil uji asumsi kehomogenan ragam sisaan pada regresi OLS tidak terpenuhi. Hal ini mengakibatkan penduga parameter dengan menggunakan metode OLS tidak lagi mempunyai ragam yang minimum dan terbaik.

3.2.3. Koefisien Determinasi pada Regresi OLS (R^2)

Nilai koefisien determinasi (R^2) yang diperoleh pada analisis regresi OLS sebesar 0,7886 atau 78,86%. Dengan demikian dapat dikatakan bahwa besarnya kontribusi variabel angka partisipasi kasar SMU (X_1), jumlah rata-rata pengeluaran per kapita sebulan untuk makanan (X_2), angka kematian bayi (X_3) dan jumlah puskesmas (X_4) dalam menjelaskan keragaman IPM di Provinsi Jawa Timur tahun 2016 sebesar 78,86%, sedangkan sisanya dijelaskan oleh faktor lain yang tidak diteliti pada penelitian ini.

3.3. Analisis GWR

Model GWR yang dalam notasi matriks sebagaimana persamaan (22):

$$y_i = X_i\beta_i + \varepsilon_i \quad (22)$$

di mana:

β_i : vektor kolom dari koefisien regresi.

X_i : vektor baris dari variabel penjelas pada lokasi i .

Melalui iterasi *Golden Section Research* dengan menggunakan bantuan *software R*, diperoleh nilai *bandwith* optimum sebesar 0,5789286 dan CV sebesar 299,407. *Bandwith* tersebut digunakan sebagai acuan penentuan *bandwith* pada masing-masing Kabupaten/Kota tersebut. Sehingga dapat diperoleh fungsi matriks pembobot *Adaptive Gaussian Kernel* untuk pendugaan parameter pada model GWR.

3.3.1. Hasil Uji Heterogenitas Spasial

Hasil pengujian heterogenitas spasial dengan menggunakan uji *Breusch-Pagan* (BP) terdapat pada Tabel 3.

Tabel 3. Nilai Duga pada Uji Heterogenitas Spasial

Nilai Duga	
Statistik uji BP	9,2025

p-value 0,05623

Berdasarkan Tabel 3, dapat disimpulkan bahwa terdapat keragaman atau heterogenitas spasial pada data Indeks Pembangunan Manusia di kabupaten/kota Provinsi Jawa Timur tahun 2016 dengan tingkat kesalahan sebesar 5,7%. Dengan demikian, analisis lanjutan yaitu GWR dapat dilakukan.

3.3.2. GWR Lokal

Pendugaan parameter model secara parsial pada setiap lokasi pengamatan terdapat pada hipotesis berikut:

$$H_0: \beta_k(u_i, v_i) = 0$$

$$H_1: \beta_k(u_i, v_i) \neq 0 ; \text{ untuk } k = 0, 1, 2, \dots, p.$$

Pengujian ini dilakukan secara parsial dengan menggunakan statistik uji t. Jika nilai statistik uji $|t| > t_{(0,025, 38-4-1)}$ sebesar 2.77, maka keputusan: tolak H_0 . Hal ini berarti bahwa variabel independen pada lokasi tersebut signifikan. Berdasarkan hasil pengujian signifikansi, diperoleh hasil pada Tabel 4.

Tabel 4. Variabel yang Signifikan pada GWR Lokal

Variabel	Kabupaten/Kota
X_2 dan X_3	Bondowoso, Situbondo, Probolinggo, Pasuruan, Sidoarjo, Mojokerto, Tuban, Lamongan, Gresik, Bangkalan, Sampang, Pamekasan, Sumenep, Probolinggo, Pasuruan, Mojokerto dan Surabaya
X_2	Semua Kabupaten/Kota di Provinsi Jawa Timur

Berdasarkan Tabel 4 terlihat bahwa pada model GWR, variabel pengeluaran per kapita sebulan untuk makanan (X_2) dan angka kematian bayi (X_3) yang berpengaruh nyata terhadap IPM. Variabel independen yang mempengaruhi IPM berbeda-beda pada masing-masing Kabupaten/Kota di Provinsi Jawa Timur. Sebagai contoh model GWR di Kota Malang. Berdasarkan hasil analisis, diperoleh bahwa IPM di Kota Malang hanya dipengaruhi oleh rata-rata pengeluaran per kapita sebulan untuk makanan (X_2) saja. Model GWR lokal yang terbentuk di Kota Malang yaitu

$$\hat{Y}_{KotaMalang} = 57,7490 + 0,0838 X_1 + 2,848 \times 10^{-5} X_2 - 0,1615 X_3 - 0,0471 X_4$$

Dengan demikian, model tersebut dapat diinterpretasikan bahwa IPM di Kota Malang akan meningkat sebesar 0,2848% jika rata-rata pengeluaran per kapita sebulan untuk makanan (X_2) naik sebesar Rp.10.000 dengan syarat kondisi variabel lain konstan. Nilai koefisien determinasi R^2 yang diperoleh untuk model IPM di Kota Malang sebesar 0,8199 atau 81,99% artinya besarnya kontribusi variabel angka partisipasi kasar SMU (X_1), jumlah rata-rata pengeluaran per kapita sebulan untuk makanan (X_2), angka kematian bayi (X_3) dan jumlah puskesmas (X_4) dalam menjelaskan keragaman indeks pembangunan manusia (IPM) di Kota Malang tahun 2016 sebesar 81,99%, sedangkan sisanya dijelaskan oleh faktor lain yang tidak diteliti.

Hasil nilai R^2 untuk Kabupaten/Kota di Provinsi Jawa Timur berkisar antara 80% hingga 83%. Nilai koefisien determinasi tinggi, namun hanya sedikit variabel yang signifikan dalam perhitungan uji t (beberapa nilai statistik uji t tidak nyata). Hal ini bisa saja disebabkan pemodelan GWR kurang cocok untuk mengestimasi IPM di Provinsi Jawa Timur.

3.3.3. GWR Global

Hasil pendugaan parameter pada GWR Global berdasarkan persamaan (7) dengan menggunakan metode pendugaan parameter WLS disajikan pada Tabel 5.

Tabel 5. Nilai Duga pada Regresi GWR Global

Parameter	Nilai Duga	<i>p-value</i>
β_0	55,2077	$5,35 \times 10^{-11}***$
β_1	0,1069	0,0831*
β_2	$2,872 \times 10^{-5}$	$1,08 \times 10^{-5}***$
β_3	-0,1435	0,0226**
β_4	-0,04645	0,1900 ^{ts}

Keterangan: ***: sangat signifikan pada taraf nyata 0,01%; **: signifikan pada taraf nyata 2,5%; *: signifikan pada taraf nyata 8,5%; ts: tidak signifikan

Berdasarkan Tabel 5, model analisis GWR Global yang diperoleh adalah

$$\hat{Y} = 55,2077 + 0,1069 X_1 + 2,872 \times 10^{-5} X_2 - 0,1435 X_3 - 0,04645 X_4 \quad (23)$$

Berdasarkan hasil pendugaan parameter pada persamaan (23) terlihat bahwa hasil pendugaan parameter pada GWR Global dan regresi OLS diperoleh hasil yang hampir sama. Hasil pengujian parameter

yang tertera pada Tabel 5 menunjukkan bahwa angka partisipasi kasar SMU (X_1) berpengaruh nyata terhadap IPM di Provinsi Jawa Timur tahun 2016 dengan tingkat kesalahan sebesar 8,5%. Rata-rata pengeluaran per kapita sebulan untuk makanan (X_2) juga berpengaruh sangat nyata terhadap IPM di Provinsi Jawa Timur tahun 2016 dengan tingkat kesalahan sebesar 0,01%. Begitu pula dengan angka kematian bayi (X_3) berpengaruh nyata terhadap terhadap IPM di Provinsi Jawa Timur tahun 2016 dengan tingkat kesalahan sebesar 2,5%. Sedangkan jumlah Puskesmas (X_4) tidak berpengaruh nyata terhadap IPM dengan tingkat kesalahan sebesar 5%.

Informasi yang dapat disimpulkan pada model GWR berdasarkan persamaan (23) yaitu IPM akan naik sebesar 0,1069% apabila angka partisipasi kasar SMU (X_1) meningkat sebesar 1% serta variabel lain tidak mengalami perubahan, IPM akan naik sebesar 0,2872 % apabila rata-rata pengeluaran per kapita sebulan untuk makanan (X_2) naik sebesar Rp. 10.000 dengan asumsi kondisi variabel lainnya konstan, dan IPM akan turun sebesar 0,1435 % apabila angka kematian bayi (X_3) naik sebesar 1% serta variabel lain tidak mengalami perubahan.

3.3.4. Pengujian Goodness of Fit Model GWR

Pemilihan model terbaik dapat dilakukan dengan menggunakan hasil analisis keragaman atau *analysis of variance (ANOVA)* dengan hipotesis berikut:

H_0 : parameter regresi konstan untuk semua titik di dalam area penelitian (regresi OLS lebih baik digunakan)

H_1 : parameter regresi bervariasi secara spasial (GWR lebih baik digunakan)

Hasil uji *goodness of fit* pada model GWR terdapat pada Tabel 6.

Tabel 6. Analisis Keragaman (Anova)

Sumber Keragaman	db	JK	KT	F
<i>OLS Residuals</i>	5,0000	226,143		
<i>GWR Improvement</i>	3,3323	32,426	9,7310	
<i>GWR Residuals</i>	29,6677	193,716	6,5295	1,4903

Ket : db : derajat bebas, JK: Jumlah Kuadrat, KT: Kuadrat Tengah dan F: statistik uji F.

Diperoleh statistik uji F sebesar $1,4903 < F_{(0,05;3,3323;29,6677)}$ sebesar 2,93403.

Keputusannya adalah Terima H_0 .

Kesimpulan: model OLS pada data IPM di Provinsi Jawa Timur tahun 2016 lebih baik digunakan.

3.4. Pemilihan Model Terbaik

Pemilihan model terbaik dilakukan berdasarkan nilai AIC yang terdapat pada Tabel 7.

Tabel 7. Nilai AIC Setiap Model

Fungsi	AIC
<i>Regresi OLS</i>	77,78
<i>Regresi GWR Global</i>	190,5837

Berdasarkan Tabel 7, nilai AIC pada regresi OLS lebih rendah dibandingkan dengan analisis GWR. Dengan demikian, dapat dikatakan bahwa regresi OLS merupakan regresi terbaik atau model yang paling efektif dalam memodelkan data IPM Provinsi Jawa Timur tahun 2016.

4. KESIMPULAN

Berdasarkan hasil analisis, diperoleh model regresi OLS yang diperoleh pada studi kasus ini yaitu

$$\hat{Y} = 55,21 + 0,1069 X_1 + 2,872 \times 10^{-5} X_2 - 0,1435 X_3 - 0,04648 X_4$$

Sedangkan model GWR yang diperoleh yaitu

$$\hat{Y} = 55,2077 + 0,1069 X_1 + 2,872 \times 10^{-5} X_2 - 0,1435 X_3 - 0,04645 X_4$$

Variabel yang berpengaruh pada Indeks Pembangunan Manusia di Provinsi Jawa Timur tahun 2016 baik pada regresi OLS maupun GWR Global sama yaitu rata-rata pengeluaran per kapita sebulan untuk makanan (X_2) dan angka kematian bayi (X_3) berpengaruh sangat signifikan terhadap IPM dengan tingkat kesalahan sebesar 2,5%. Sedangkan angka partisipasi kasar SMU (X_1) berpengaruh signifikan dengan tingkat kesalahan 8,5%. Akan tetapi pada GWR lokal diperoleh hasil signifikansi yang berbeda-beda pada masing-masing lokasi. Secara keseluruhan, variabel yang berpengaruh signifikan pada GWR lokal adalah variabel rata-rata pengeluaran per kapita sebulan untuk makanan (X_2) dan angka kematian bayi (X_3).

Nilai AIC pada model regresi OLS (77,78) lebih rendah dibandingkan dengan analisis GWR (190,5837). Sehingga dapat disimpulkan bahwa model regresi OLS lebih baik dari pada model GWR pada data

IPM Provinsi Jawa Timur tahun 2016. Dengan kata lain, model regresi OLS merupakan model paling efektif dalam menggambarkan IPM Provinsi Jawa Timur tahun 2016.

Banyaknya variabel independen yang digunakan untuk melihat pengaruhnya terhadap variabel independen pada penelitian ini hanya menggunakan empat variabel independen dan menggunakan fungsi pembobot *Adaptive Gaussian Kernel*. Oleh karena itu, penelitian berikutnya dapat ditambahkan beberapa variabel independen lain yang diduga dapat mempengaruhi IPM dan membandingkan pemodelan analisis GWR dengan menggunakan fungsi pembobot yang berbeda-beda.

5. UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak yang telah membantu proses penyelesaian penelitian ini.

DAFTAR PUSTAKA

- [1] Melliana, A., Zain, I. (2013). Analisis Statistika Faktor yang Mempengaruhi Indeks Pembangunan Manusia di Kabupaten/Provinsi Jawa Timur dengan Menggunakan Regresi Panel. *Jurnal Sains dan Seni Pomits*, 23337-3520.
- [2] Muqorrob, M., Soejoto, A. (2017). Pengaruh Indeks Pembangunan Manusia (IPM) Terhadap Pertumbuhan Ekonomi Provinsi di Jawa Timur. *Jurnal Pendidikan Ekonomi*, 5(3).
- [3] Irjaya, R. (2017). Analisis Faktor-Faktor yang Mempengaruhi Indeks Pembangunan Manusia (IPM) di Provinsi Jawa Timur Tahun 2012-2014. *Skripsi. Tidak dipublikasikan*. Universitas Muhammadiyah Surakarta.
- [4] Putra, D. M., Ratnasari, V. (2015). Pemodelan Indeks Pembangunan Manusia (IPM) Provinsi Jawa Timur dengan Menggunakan Metode Regresi Logistik Ridge. *Journal Sains dan Seni ITS*, 2337 – 3520.
- [5] Walpole, R. E., Myers, R. H., Myers, S. L., Ye, K. (2012). *Probability & Statistics for Engineers & Scientist*. 9th Edition. Boston: Pearson Education, Inc.
- [6] Gujarati, D. (2003). *Basic Econometrics*. 4th Edition. USA: TheMcGraw-Hill Companies, United States Military Academy, West Point.
- [7] Judd, C. M., McClelland, G.H, Ryan, C. S. (2017). *Data Analysis: A Model Comparison Approach to Regression, ANOVA, and Beyond*. 3rd Edition. New York: Routledge.
- [8] Draper N. R., Smith H. (1992). *Analisis Regresi Terapan*. Edisi Kedua. Sumantri B : Penerjemah. Terjemahan dari *Applied Regression Analysis*. Jakarta: Gramedia.
- [9] Hair Jr., J. F., Black, W. C., Babin, B. J., Anderson, R. E. (2014). *Multivariate Data Analysis*. 7th Edition. Harlow: Pearson Education Limited.
- [10] Bivand, R. S. (2010). *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*. Diedit oleh Fischer, M. M. & Getis, A. Germany: Springer.
- [11] Pramoedyo, H. (2017). *Analisis Spasial Dasar*. Malang: Universitas Negeri Malang (UM Press).
- [12] Bivand, R. S.; Pebesma, E., Gomez-Rubio, V. (2013). *Applied Spatial Data Analysis in R*. New York: Springer-Verlag.
- [13] Konishi, S., Kitagawa, G. (2018). *Information Criteria and Statistical Modeling*. USA: Springer Science + Business Media, LLC.