

Pemodelan Indeks Pembangunan Manusia Provinsi Jawa Barat Tahun 2018 Menggunakan *Geographically Weighted Regression* dengan Pembobotan *Kernel Gaussian*

Retnowati Sulastri*, Nusar Hajarisman

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Islam Bandung, Indonesia.

*retnowatisulastri2@gmail.com

Abstract, Modeling on the Human Development Index in West Java province. Location differences, of course, also provide various location characteristics. One method to determine the factors that affect the Human Development Index is to use linear regression analysis. Spatial data regression analysis can be used because it has a dependency between data measurement and location. The local form of spatial regression analysis is Geographically Weighted Regression (GWR). GWR shows spatial heterogeneity (location). This GWR model is used to consider elements of geography or location as weighting in estimating model parameters. In the GWR model, the estimation of the model parameters is obtained using the Weighted Least Square (WLS), namely by giving different weights at each location where the data is collected. In 2018 the HDI of West Java Province has reached 71.30. This figure increased by 0.61 points compared to the HDI of West Java Province in 2017 which was 70.69. The parameter estimator model with the GWR method produces $R^2 = 99.68\%$ and $SSE = 1.899759$. The GWR model is proven to be able to increase the R^2 value and decrease the AIC value. So it can be said that the GWR model is the most appropriate model that can be used in estimating the District /City Level Human Development Index in West Java Province in 2018.

Keywords: Human Development Index Modeling. Spatial data. Regression. Geographically Weighted Regression. Gaussian Kernel.

Abstrak, Pemodelan pada Indeks Pembangunan Manusia di provinsi Jawa Barat. Perbedaan lokasi tentunya juga memberikan karakteristik lokasi yang beragam. Salah satu metode untuk mengetahui faktor yang berpengaruh terhadap Indeks Pembangunan Manusia adalah menggunakan analisis regresi linear. Analisis regresi data spasial dapat digunakan karena memiliki sifat ketergantungan antara pengukuran data dengan lokasi. Bentuk lokal dari analisis regresi spasial adalah *Geographically Weighted Regression* (GWR). GWR menunjukkan adanya *heterogenitas* secara spasial (lokasi) Model GWR ini digunakan untuk mempertimbangkan unsur geografi atau lokasi sebagai pembobot dalam menaksir parameter modelnya. Dalam model GWR penaksiran parameter modelnya diperoleh dengan menggunakan *Weighted Least Square* (WLS) yaitu dengan memberikan pembobot yang berbeda pada setiap lokasi di mana data tersebut dikumpulkan. pada tahun 2018 IPM Provinsi Jawa Barat telah mencapai 71.30. Angka ini meningkat sebesar 0.61 poin dibandingkan dengan IPM Provinsi Jawa Barat pada tahun 2017 yang sebesar

70.69. Model penduga parameter dengan metode GWR menghasilkan $R^2 = 99.68\%$ dan $SSE = 1.899759$. Model GWR terbukti mampu meningkatkan nilai R^2 dan menurunkan nilai AIC. Sehingga dapat dikatakan model GWR merupakan model yang paling sesuai yang dapat digunakan dalam pendugaan terhadap Indeks Pembangunan Manusia Tingkat Kabupaten/Kota di Provinsi Jawa Barat pada Tahun 2018.

Kata Kunci : Pemodelan Indeks Pembangunan Manusia. Data spasial. Regresi. Geographically Weighted Regression. Kernel Gaussian.

1. Pendahuluan

Pembangunan manusia yang telah ditetapkan oleh perserikatan Bangsa-Bangsa yaitu disebut Human Development Index atau Indeks Pembangunan Manusia (IPM). Indeks pembangunan manusia menjelaskan bagaimana penduduk dapat mengakses hasil pembangunan dalam memperoleh kesehatan, pendidikan, pendapatan, dan sebagainya. IPM dibangun melalui tiga dimensi. Dimensi tersebut mencakup umur panjang dan hidup sehat (*a long and healthy life*), pengetahuan (*knowledge*), dan standar hidup layak (*decent standar of living*). Ketiga dimensi tersebut memiliki pengertian yang sangat luas karena terkait banyak faktor. Analisis regresi merupakan analisis statistika yang digunakan untuk mengetahui hubungan antara variabel dependen dengan variabel *independen*. Dalam model regresi linear memiliki beberapa asumsi yaitu normalitas, homoskedastisitas, non-multikolinearitas dan autokorelas. Adapun data spasial adalah data yang memuat adanya informasi lokasi atau geografis suatu wilayah. Kenyataannya kondisi semua wilayah yang diamati tidak sama, karena adanya faktor geografis, keadaan sosial budaya, maupun hal lainnya yang melatar belakangi kondisi yang seharusnya juga diteliti. Perbedaan ini sangat memungkinkan munculnya heterogenitas spasial. Bila kasus ini terjadi, maka regresi linear biasa kurang mampu dalam menjelaskan fenomena data yang sebenarnya. Pada data IPM terdapat ketergantungan antar pengukuran data dengan lokasi. Menurut Fotheringham, dkk. (2002) dalam GWR adalah metode statistika yang digunakan untuk menganalisis heterogenitas untuk menganalisis heterogenitas spasial. Heterogenitas spasial adalah suatu peubah bebas yang sama memberikan respon yang tidak sama pada lokasi yang berbeda dalam satu wilayah penelitian. Setiap parameter regresi diestimasi di setiap titik lokasi geografis sehingga hubungan antara variabel respon (Y) dan variabel prediktor (X) bervariasi (tidak sama) di sepanjang lokasi ini pembobot yang digunakan adalah Kernel Gaussian, yang membutuhkan nilai bandwidth sebagai parameter jarak yang masih mempengaruhi suatu Kabupaten/Kota terhadap Kabupaten/Kota lainnya. Bandwidth optimum dapat diperoleh dengan meminimalkan nilai koefisien CV (*cross validation*).

2. Landasan Teori

Regresi Linear

Menurut Anneke dan Nusar (2011), analisis regresi merupakan metode statistik untuk menganalisis hubungan antar variabel dan menghasilkan sebuah model, dimana model tersebut dapat digunakan sebagai alat prediksi nilai pada data yang mengandung variabel tersebut dengan tujuan untuk menaksir parameter didalam model atau disebut juga sebagai pencocokan model terhadap data. Kejadian dimana sebuah variabel, katakanlah Y, dipengaruhi berdasarkan satu atau lebih variabel, katakanlah X, disebut model regresi linear sederhana, karena hanya melibatkan satu variabel bebas. Dapat diperhatikan bahwa ε merupakan galat statistik yaitu merupakan variabel yang memperhitungkan kemelencengan dari nilai yang diperoleh berdasarkan model dengan nilai yang sebenarnya. Pada umumnya, variabel tak bebas dapat dipengaruhi oleh k variabel bebas (x_1, x_2, \dots, x_k), sehingga

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad \dots(2.1)$$

Model Geographically Weighted Regression (GWR)

Model ini merupakan pengembangan regresi linear dengan parameter model yang berbeda di setiap lokasi. Model GWR dirumuskan sebagai berikut (Fotheringham, et al., 2002) :

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i)x_{ik} + \varepsilon_i, i = 1, 2, \dots, n \quad \dots(2.2)$$

Weighted Least Square digunakan untuk melakukan penaksiran parameter dengan memberikan matriks pembobot yang berbeda di setiap lokasi amatan.

Pembobotan Model GWR

Adapun peran pembobot sangatlah penting karena dalam hal ini nilai pembobot mewakili letak data observasi satu dengan observasi lainnya. Adapun fungsi pembobot Adaptive Kernel Gaussian dapat dituliskan sebagai berikut:

$$w_{ij} = \exp\left(-\frac{1}{2}\left(d_{ij}/b\right)^2\right) \quad \dots(2.3)$$

Dimana b adalah *bandwidth* sedangkan d_{ij} adalah jarak *Euclidian* antara pengamatan pada titik ke- i dan ke- j .

Adalah jarak eucliden antara lokasi (u_i, v_i) ke lokasi (u_j, v_j) dan h adalah parameter non negative yang diketahui dan biasanya disebut parameter peghalus (*bandwidth*). Ada beberapa metode yang digunakan untuk memilih *bandwidth* optimum, salah satu diantaranya adalah metode *Cross Validation* yang secara matematis didefinisikan sebagai berikut:

$$CV = \sum_{i=1}^n (y_i - \hat{y}_{\neq i}(h))^2 \quad \dots(2.4)$$

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

Pengujian Kesesuaian Model (*Goodness of Fit*)

Pengujian ini dilakukan dengan hipotesis sebagai berikut :

H_0 : $\beta_k(u_i, v_i) = \beta_k, k = 1, 2, \dots, p$ (tidak ada perbedaan yang signifikan antara model regresi global dengan GWR).

H_1 : paling sedikit ada satu $\beta_k(u_i, v_i)$ yang berhubungan dengan lokasi (u_i, v_i) (ada perbedaan yang signifikan antara model regresi global dan GWR).

Statistik uji yang digunakan adalah :

$$F^* = \frac{SSE(H_0)/df_1}{SSE(H_1)/df_2} \quad (2.5)$$

Dengan :

$$SSE(H_0) = Y^T(\mathbf{I}-\mathbf{H})\mathbf{Y} \text{ dimana } \mathbf{H}=\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \quad (2.6)$$

$$df_1 = n - p - 1$$

$$SSE(H_1) = Y^T(\mathbf{I} - \mathbf{S})^2(\mathbf{I} - \mathbf{S})\mathbf{Y}$$

$$df_2 = (n - 2tr(\mathbf{S}) +$$

$$tr(\mathbf{S}^T\mathbf{S}))$$

Jika F^* lebih besar dari F_{tabel} maka dapat diambil keputusan tolak (H_0). dengan kata lain model GWR mempunyai *goodness of fit* yang lebih baik daripada model regresi global. F^* akan mengikuti distribusi *F* dengan derajat bebas df_1 dan df_2 . Jika diberikan tingkat signifikan sebesar α . Maka diambil keputusan dengan menolak (H_0) jika nilai $F^* > F_{\alpha; df_1, df_2}$. (Caraka dan Yasin, 2017)

Indeks Pembangunan Manusia

Indeks Pembangunan Manusia (IPM) atau *Human Development Indeks* (HDI) adalah pengukuran perbandingan dari tiga bidang pembangunan manusia yang dianggap sebagai

mendasar yaitu usia hidup (*longevity*), pengetahuan (*knowledge*), dan standar hidup layak (*decent living*). Pengertian IPM oleh *United Nations Development Programme* (UNDP) yang menyatakan bahwa IPM atau HDI merupakan salah satu pendekatan untuk mengukur tingkat keberhasilan pembangunan manusia. Manfaat IPM yaitu IPM merupakan indikator penting untuk mengukur keberhasilan dalam upaya membangun kualitas hidup manusia serta dapat menentukan peringkat atau level pembangunan suatu wilayah atau negara. Bagi Indonesia sendiri IPM merupakan data strategis karena selain sebagai ukuran kinerja Pemerintah, IPM juga digunakan sebagai salah satu alokator penentuan Dana Alokasi Umum (DAU) (BPS, 2017). IPM dibentuk oleh tiga dimensi dasar yaitu umur panjang dan hidup sehat (*a long and healthy life*), pengetahuan (*knowledge*), dan standard hidup layak (*decent standard of living*).

3. Data

Data yang digunakan adalah data sekunder hasil pencatatan yang diperoleh dari instansi pemerintah yaitu Badan Perencanaan Pembangunan Daerah Provinsi Jawa Barat. Pada penelitian ini yang akan dijadikan unit penelitian adalah kabupaten/kota di provinsi Jawa Barat tahun 2018 yang terdiri dari 27 kabupaten dan kota, dimana terdiri 18 kabupaten dan 9 kota.

Tabel 1. Data Indeks Pembangunan Provinsi Jawa Barat Tahun 2018

Kabupaten/Kota	IPM	AHH	HLS	RLS	Pengeluaran per Kapita
Bogor	69.69	70.86	12.44	7.88	10,323
Sukabumi	66.05	70.49	12.2	6.8	8,618
Cianjur	64.62	69.7	11.9	6.93	7,874
Bandung	71.75	73.26	12.64	8.58	10,203
Garut	65.42	71.03	11.8	7.5	7,597
Tasikmalaya	65	68.96	12.48	7.13	7,761
⋮	⋮	⋮	⋮	⋮	⋮
Kota Banjar	71.25	70.59	13.2	8.6	10,329

4. Hasil Penelitian dan Pembahasan

Model Regresi Linear

Berikut merupakan nilai penduga/estimasi parameter model regresi dengan metode OLS. Dengan demikian model pendugaan regresi linear yang dihasilkan adalah :

$$\hat{y} = 0.504 AHH + 1.219 RLS + 1.109 HLS + 0.001 PpK$$

Model tersebut dapat diinterpretasikan jika angka harapan hidup bertambah satu persen maka angka Indeks Pembangunan Manusia akan bertambah sebesar 0.504% dengan variabel lain tetap. Jika angka rata-rata lama sekolah bertambah satu persen maka presentase Indeks Pembangunan Manusia akan bertambah sebesar 1.219% dengan variabel lain bersifat tetap. Sedangkan untuk angka harapan lama sekolah bertambah satu persen maka presentase angka Indeks Pembangunan Manusia sebesar 1.109% dengan variabel lainnya tetap. Untuk pengeluaran per kapita bertambah satu persen maka angka Indeks Pembangunan Manusia akan bertambah sebanyak 1.001% dengan variabel lain bersifat tetap.

Pengujian Kesesuaian Model GWR

Pada pengujian ini dilakukan dengan menggunakan uji F atau Goodness of Fit untuk mengetahui pengaruh pembobotan pada estimasi parameter. Berikut ini merupakan hipotesis untuk menguji kesesuaian model GWR :

$$H_0: \beta_k(u_i, v_i) = \beta_k, k = 1, 2, \dots, p \text{ (tidak ada perbedaan yang signifikan antara model regresi global dengan GWR).}$$

H_1 : paling sedikit ada satu $\beta_k(u_i.v_i) \neq \beta_k$ (ada perbedaan yang signifikan antara model regresi global dan GWR).

Berikut ini merupakan *output* yang diperoleh dari analisis GWR :

Tabel 2. Pengujian Model GWR

Source	DF	SS	MS	F
Global Residuals	22.000	1.907		
GWR Improvement	0.173	0.007	0.042	
GWR Residuals	21.827	1.900	0.087	0.587406

Dari tabel diatas didapatkan hasil F_{hitung} model GWR yaitu sebesar 0.587406. dengan kriteria penolakan ($F_{tabel} = 6.22667$).

karena $F_{hitung}(0.587406) < F_{tabel}(6.22667)$ maka dapat diambil keputusan bahwa H_0 diterima yang artinya tidak ada perbedaan yang signifikan antara model regresi global dengan GWR.

Pemilihan Model Terbaik

Tabel 3. Kriteria Pemilihan Model Terbaik

Kriteria	Regresi Linear	GWR
R^2	99.65%	99.68 %
AIC	7. 065672	3. 146592

Berdasarkan hasil dari tabel 3.3 dapat disimpulkan bahwa dengan menggunakan kriteria nilai R^2 dan AIC yang dihasilkan dari kedua model tersebut. Terlihat bahwa nilai R^2 , dan AIC untuk model GWR merupakan model yang lebih baik dibandingkan dengan model regresi linear dengan metode OLS. Model GWR terbukti mampu meningkatkan nilai R^2 dan menurunkan nilai dan AIC .

5. Kesimpulan

Berdasarkan pembahasan dalam penelitian ini, peneliti menyimpulkan beberapa hasil penelitian sebagai berikut :

1. Berdasarkan data yang digunakan pada penelitian ini. pembentukan model regresi linear dengan metode Ordinary Least Square (OLS) didapatkan dengan pengaruh yang signifikan dari setiap variabel AHH, RLS, HLS, PpK terhadap Indeks Pembangunan Manusia
2. Pemodelan terhadap Indeks Pembangunan Manusia tingkat Kabupaten/Kota di Provinsi Jawa Barat dengan metode GWR disimpulkan bahwa dengan menggunakan kriteria nilai R^2 dan AIC yang dihasilkan dari kedua model tersebut. Terlihat bahwa nilai R^2 , dan AIC untuk model GWR merupakan model yang lebih baik dibandingkan dengan model regresi linear dengan metode OLS. Model GWR terbukti mampu meningkatkan nilai R^2 dan menurunkan nilai AIC. Sehingga dapat dikatakan model GWR merupakan model yang paling sesuai yang dapat digunakan dalam pendugaan terhadap Indeks Pembangunan Manusia Tingkat Kabupaten/Kota di Provinsi Jawa Barat pada Tahun 2018

6. Saran

Pada penelitian ini menggunakan metode regresi spasial dengan pendekatan titik yaitu *Geographically Weighted Regression* dengan adanya penambahan faktor lain mungkin dapat

membuat model sangat baik dalam penelitian selanjutnya, dapat pula dilakukan penelitian model regresi spasial dengan pendekatan titik lainnya seperti *Geographically Weighted Logistic Regression (GWLR)*, atau *Geographically Weighted Poisson Regression (GWPR)*.

Daftar Pustaka

- [1] Caraka.R.E.,Yasin.H.2017.*Geographically Weighted Regression(GWR) Sebuah Pendekatan Regresi Geografis*.Yogyakarta:Mobius
- [2] Fortheringham. A. S., Brunson. C., & Charlton. M. E. 2002. *Geographically Weighted Regression : The Analysis of Spatially Varying Relationships*. England: John Wiley and Sons Ltd.
- [3] Hajarisman, Nusar, & Anneke Iswani Achmad. (2011). *Analisis Regresi Lanjut*. Bandung: Universitas Islam Bandung.