

PENGGUNAAN GEOGRAPHICALLY WEIGHTED REGRESSION (GWR) DENGAN PEMBOBOT GAUSS KERNEL UNTUK KLASIFIKASI DESA MISKIN (Studi kasus desa-desa di Kabupaten Jember, Jawa Timur)

Rita Rahmawati¹, Anik Djuraidah², M. Nur Aidi²

¹Mahasiswa Pascasarjana Statistika, Institut Pertanian Bogor

²Jurusan Statistika, Institut Pertanian Bogor

Abstrak. Penentuan suatu wilayah desa tergolong miskin atau tidak, umumnya menggunakan rata-rata pengeluaran per kapita sebagai indikator utamanya. Hanya saja, analisis yang biasanya digunakan kebanyakan masih bersifat global dan hasilnya diberlakukan untuk semua desa. Padahal masalah kemiskinan sangat mungkin dipengaruhi oleh lokasi (*space*) dan ketetanggaan (*neighboring*), sehingga data antar pengamatan sulit untuk diasumsikan saling bebas. Salah satu analisis yang mengakomodir masalah spasial ini adalah *Geographically Weighted Regression* (GWR), yaitu regresi yang terboboti secara geografis. Pengamatan di lokasi yang lebih jauh diboboti dengan pembobot yang lebih kecil, sesuai *Tobler's first law of geography* yang menyatakan bahwa semakin dekat suatu lokasi maka pengaruhnya akan semakin besar. Dalam banyak analisis GWR, juga dalam tulisan ini pembobot yang digunakan adalah Gauss Kernel, yang membutuhkan nilai *bandwidth* sebagai parameter jarak yang masih mempengaruhi suatu desa terhadap desa lainnya. *Bandwidth* optimum dapat diperoleh dengan meminimalkan nilai CV (*cross validation*). Model GWR tidak dapat digunakan untuk menduga parameter selain parameter di lokasi pengamatan (Walter, Carsten and Jeremy 2005). Untuk desa-desa yang tidak diobservasi peubah responnya, digunakan interpolasi *Inverse Distance Weighted* (IDW) berdasarkan nilai-nilai penduga GWR rata-rata pengeluaran per kapita tiap desa.

Kata kunci: *Geographically Weighted Regression, Gauss Kernel, bandwidth, cross validation*

1. PENDAHULUAN

Masalah kemiskinan masih menjadi salah satu masalah besar di Indonesia. Badan Pusat Statistik (BPS) menyatakan, jumlah penduduk miskin pada bulan Maret 2008 di Indonesia mencapai 15,42% atau 34,96 juta orang (BPS, 2008). Berbagai upaya dilakukan pemerintah untuk menanggulangi masalah ini, di antaranya dengan Instruksi Presiden tentang Desa Tertinggal (IDT). Dengan memprediksi wilayah-wilayah miskin hingga tingkat administrasi desa, diharapkan upaya pengentasan kemiskinan akan lebih tepat sasaran.

Dalam menentukan suatu wilayah desa tergolong miskin atau tidak, analisis yang biasanya digunakan oleh BPS masih bersifat global dan diberlakukan untuk semua lokasi desa. Pendekatan model global berarti menggunakan rata-rata dari wilayah-wilayah yang lebih kecil (wilayah lokal), dan akan memberikan informasi yang reliabel untuk wilayah lokal jika tidak ada atau hanya ada sedikit keragaman antar wilayah lokal tersebut (Fotheringham, Brunson and Charlton 2002). Sementara kondisi kemiskinan desa sangat mungkin dipengaruhi oleh lokasi desa, termasuk posisinya terhadap desa lain sehingga asumsi kebebasan data antar desa pun sulit dipenuhi. Hukum pertama tentang geografi dikemukakan oleh Tobler (*Tobler's first law of geography*) dalam Schabenberger and Gotway (2005), yang menyatakan "*everything is related to everything else, but near things are more related than distant things*". Segala sesuatu adalah saling berhubungan, tetapi sesuatu yang lebih dekat akan lebih berpengaruh daripada sesuatu yang jauh. *Geographically Weighted Regression* (GWR) adalah salah satu solusi yang dapat digunakan untuk membentuk analisis regresi namun bersifat lokal untuk setiap lokasi. Hasil analisis ini adalah model regresi yang nilai-nilai parameternya berlaku hanya pada tiap lokasi pengamatan, dan berbeda dengan lokasi lainnya. Dalam GWR digunakan unsur matriks

pembobot $W(i)$ yang besarnya tergantung pada kedekatan antar lokasi. Semakin dekat suatu lokasi, bobot pengaruhnya akan semakin besar. Fungsi pembobot yang digunakan untuk GWR dalam tulisan ini adalah fungsi Gauss Kernel. Model GWR tidak dapat digunakan untuk menduga parameter selain parameter di lokasi pengamatan (Walter, Carsten and Jeremy 2005). Untuk desa-desa yang tidak diobservasi peubah responnya, digunakan interpolasi *Inverse Distance Weighted* (IDW) berdasarkan nilai-nilai penduga GWR rata-rata pengeluaran per kapita tiap desa.

Dalam penelitian ini, studi kasus dipilih Kabupaten Jember, di antaranya karena kabupaten ini merupakan salah satu wilayah termiskin di Pulau Jawa, khususnya di Propinsi Jawa Timur. Hasil pendataan program perlindungan sosial pada September 2008 dan data survei sosial ekonomi nasional (Susenas) Maret 2009, mencatat bahwa Kabupaten Jember memiliki penduduk miskin terbanyak di Jawa Timur yaitu mencapai 237.700 rumah tangga miskin (Djunaidy, 2010). Gubernur Jawa Timur, Soekarwo, menyebutkan jumlah rumah tangga miskin tahun 2010 di Kabupaten Jember mencapai angka 370.000 dan menjadikan Jember sebagai kabupaten dengan penduduk miskin terbesar di Propinsi Jawa Timur (Ant, 2010).

2. Regresi global (*global regression*)

Persamaan regresi global yang biasa didefinisikan dengan menggunakan metode pendugaan parameter *Ordinary Least Square* (OLS), secara umum dapat dituliskan dalam persamaan matematis sebagai berikut:

$$y_i = b_0 + \sum_{j=1}^p x_{ij}b_j + e_i, \quad i = 1, 2, 3, \dots, n$$

dimana b_0 adalah konstanta, b_j adalah nilai koefisien peubah penjelas x_j , p adalah banyaknya peubah penjelas yang digunakan dalam model, n adalah banyaknya pengamatan (contoh) dan e adalah galat acak yang diasumsikan menyebar $N(0, \sigma^2 I)$, dengan $e = (e_1, e_2, e_3, \dots, e_n)'$ dan I adalah matriks identitas. Dengan meminimumkan jumlah kuadrat galat, nilai penduga parameter dengan OLS dalam bentuk vektor adalah sebagai berikut:

$$\hat{\mathbf{b}} = (X'X)^{-1}X'Y$$

dimana $\hat{\mathbf{b}} = (\hat{b}_0, \hat{b}_1, \dots, \hat{b}_p)'$ adalah vektor $p+1$ sebagai koefisien regresi, X adalah matriks peubah penjelas berukuran $n \times (p+1)$ dengan kolom pertama bernilai 1 untuk konstanta, dan Y adalah vektor peubah respon.

3. Geographically Weighted Regression (GWR)

Model GWR merupakan pengembangan dari model regresi global. Persamaan regresi global diasumsikan berlaku secara umum di setiap lokasi pengamatan. GWR adalah model regresi linier lokal yang menghasilkan penduga parameter model yang bersifat lokal untuk setiap lokasi pengamatan dengan metode *Weighted Least Square* (WLS), yaitu :

$$\hat{\mathbf{b}}(i) = (X'W(i)X)^{-1}X'W(i)Y$$

dimana $W(i) = \text{diag}[w_1(i), w_2(i), \dots, w_n(i)]$, dengan $0 \leq w_j(i) \leq 1$ ($i, j = 1, 2, 3, \dots, n$).

$W(i)$ adalah matriks pembobot spasial lokasi ke- i yang nilai elemen-elemen diagonalnya ditentukan oleh kedekatan lokasi ke- i dengan lokasi lainnya (lokasi ke- j). Semakin dekat lokasinya maka semakin besar nilai pembobot pada elemen yang bersesuaian. Salah satu fungsi pembobot spasial dalam GWR, diadopsi dari bentuk fungsi Gauss Kernel yaitu:

$$w_j(i) = \exp \left[-1/2 \left(\frac{d_{ij}}{b} \right)^2 \right]$$

dengan d_{ij} =jarak dari lokasi- i ke lokasi ke- j dan b =bandwidth, yaitu suatu nilai yang harus ditetapkan, sebagai gambaran jarak maksimal suatu lokasi masih mempengaruhi lokasi lainnya.

Salah satu cara yang dapat digunakan sebagai kriteria untuk mendapatkan nilai *bandwidth* optimum adalah dengan meminimumkan nilai CV, dengan rumus:

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

$\hat{y}_{\neq i}(b)$ adalah nilai dugaan y_i (*fitting value*) dengan pengamatan di lokasi- i dihilangkan dari proses prediksi (Fotheringham, Brunson and Charlton 2002). *Bandwidth* optimum dapat diperoleh dengan proses iterasi hingga didapatkan CV minimum.

Untuk mendeteksi secara global apakah GWR lebih baik daripada OLS untuk data kasus yang digunakan, dapat diuji dengan *analysis of variance* (ANOVA) yang diusulkan Brunson dalam Dimulyo (2009) sebagai berikut:

$$F_{hit} = \frac{(RSS_{OLS} - RSS_{GWR})/v_1}{RSS_{GWR}/\delta_1}$$

dimana RSS_{OLS} adalah jumlah kuadrat galat dari model OLS dan RSS_{GWR} adalah jumlah kuadrat galat dari model GWR. Nilai F_{hit} akan mendekati sebaran F dengan derajat bebas $v_1^2/v_2, \delta_1^2/\delta_2$, dimana $\delta_i = \text{tr}[(I - S)'(I - S)]^i, i = 1, 2$.

v_1 adalah nilai dari $n-p-1-\delta_1$, v_2 adalah nilai dari $n-p-1-2\delta_1+\delta_2$, dan S adalah hat matrix dari model GWR. Nilai F_{hit} yang kecil akan mendukung diterimanya hipotesis nol yang menyatakan bahwa model GWR dan OLS sama efektifnya dalam menjelaskan hubungan antar peubah. Dengan tingkat signifikansi α , hipotesis nol akan ditolak jika $F_{hit} > F_{\alpha}(v_1^2/v_2, \delta_1^2/\delta_2)$.

4. Data

Data yang digunakan dalam tulisan ini adalah data sekunder dari BPS, yaitu data Potensi Desa (Podes) dan Susenas tahun 2008. Wilayah yang digunakan adalah 35 desa/kelurahan di Kabupaten Jember yang terambil dalam Susenas 2008. Peubah-peubah penjelas diperoleh dari data Podes. Berdasarkan studi yang dilakukan oleh BPS, kemajuan atau ketertinggalan suatu desa dicerminkan oleh indikator utama, yaitu tinggi rendahnya rata-rata pengeluaran per kapita penduduk desa tersebut. Beberapa faktor yang diduga menjadi penyebab maju tidaknya suatu desa, diantaranya adalah faktor alam dan lingkungan, faktor sarana dan prasarana serta faktor sosial ekonomi penduduknya (BPS, 2002).

Berdasarkan studi BPS (2002) dan ketersediaan data dari Podes 2008, serta analisis korelasi antar peubah penjelas maupun dengan peubah respon, peubah-peubah penjelas yang akhirnya digunakan adalah:

X_1 = persentase jumlah keluarga pertanian

X_2 = jumlah surat miskin/SKTM yang dikeluarkan desa 1 tahun terakhir

Peubah respon Y yang digunakan dalam penelitian ini adalah **rata-rata pengeluaran per kapita tiap desa** dari data Susenas 2008. Untuk pendugaan parameter dengan GWR, digunakan 35 desa/kelurahan yang teramati dalam Susenas tersebut hingga diperoleh penduga-penduga parameter pada lokasi-lokasi yang diamati. Sedangkan untuk jarak antar desa, diperoleh dengan terlebih dahulu menentukan titik pusat wilayah tiap desa dengan metode *polygon thiesen*, kemudian dihitung jarak *Euclid* antar titik-titik pusat tersebut. Untuk keperluan ini, digunakan data yang diperoleh dari BMKG (Badan Meteorologi Klimatologi dan Geofisika).

5. Hasil dan Pembahasan

Dengan analisis regresi biasa (global), hasil model persamaan regresinya adalah $\hat{Y} = 384195 - 2645X_1 + 615X_2$, dengan rincian dan ANOVA sebagai berikut:

Penduga	Koeffisien	SE Koeffisien	T	P
Konstan	384195	57050	6.73	0.000
X1	-2645.0	807.7	-3.27	0.003
X2	614.6	246.6	2.49	0.018

Sumber Keragaman	DF	SS	MS	F	P
Regression	2	1.87415E+11	93707700211	10.53	0.000
Residual Error	32	2.84767E+11	8898977607		
Total	34	4.72183E+11			

Hasil di atas, diasumsikan sama dan digunakan untuk semua wilayah desa di seluruh Kabupaten Jember.

Langkah pertama untuk analisis GWR adalah menentukan *bandwidth* yang akan digunakan dalam fungsi pembobot Gauss Kernel. Dengan iterasi hingga didapatkan minimum CV, diperoleh nilai 70,30617 sehingga fungsi pembobot spasial GWR-nya menjadi:

$$w_j(i) = \exp \left[-1/2 \left(d_{ij}/70.30617 \right)^2 \right]$$

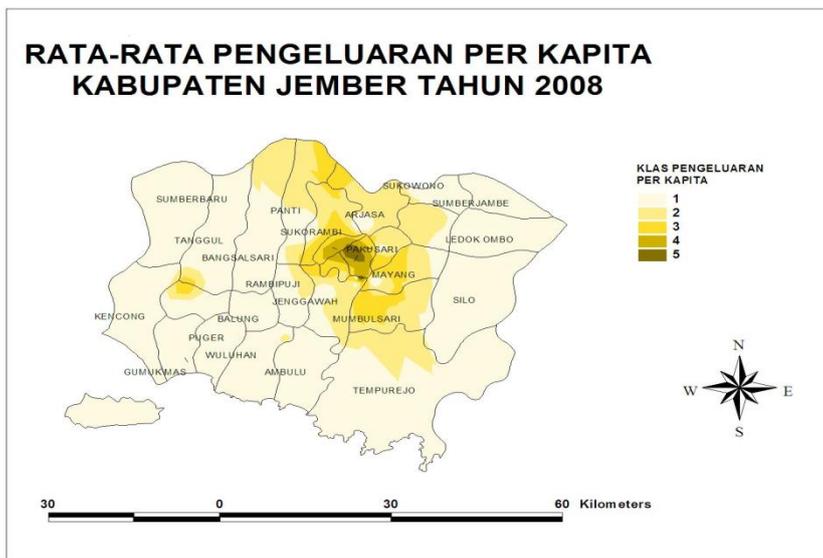
Tabel 1. Hasil nilai-nilai penduga parameter GWR dengan pembobot Gauss Kernel

No	Kode BPS	Desa	Kecamatan	b0	b1	b2	\hat{Y} GWR
1	3509010001	PASEBAN	KENCONG	381224	-2596	628	221272
2	3509020005	GUMUKMAS	GUMUK MAS	381723	-2599	627	244765
3	3509020007	TEMBOKREJO	GUMUK MAS	382883	-2618	625	245331
4	3509030012	WRINGIN TELU	PUGER	384301	-2638	622	263341
5	3509040002	AMPEL	WULUHAN	385912	-2656	619	182717
6	3509040004	KESILIR	WULUHAN	386994	-2673	617	210174
7	3509050003	SABRANG	AMBULU	387054	-2671	617	235870
8	3509060005	SIDODADI	TEMPUREJO	389147	-2699	612	209947
9	3509070002	PACE	SILO	393513	-2767	602	215011
10	3509070006	SEMPOLAN	SILO	393910	-2778	601	231450
11	3509070008	GARAHAN	SILO	394692	-2787	599	223745
12	3509080007	MRAWAN	MAYANG	390996	-2735	608	176452
13	3509100001	KEMUNING SARI KIDUL	JENGGAWAH	388279	-2693	614	326609
14	3509110002	SUKAMAKMUR	AJUNG	389105	-2707	612	278093
15	3509110006	WIROWONGSO	AJUNG	390272	-2724	609	219080
16	3509130002	KARANG SEMANDING	BALUNG	386001	-2663	619	250033
17	3509130005	BALUNG KIDUL	BALUNG	386441	-2667	618	273224
18	3509140007	GADINGREJO	UMBULSARI	383097	-2625	624	221661
19	3509160005	WRINGIN AGUNG	JOMBANG	382208	-2613	626	239524
20	3509170004	PRINGGOWIRAWAN	SUMBER BARU	383775	-2641	623	321122
21	3509170007	JATIROTO	SUMBER BARU	382344	-2623	626	304342
22	3509190002	SUKOREJO	BANGSALSARI	386568	-2673	617	255185
23	3509190007	GAMBIRONO	BANGSALSARI	385878	-2666	619	238917
24	3509200003	SERUT	PANTI	388741	-2709	613	180603
25	3509220001	KEMUNINGLLOR	ARJASA	391024	-2742	608	245662
26	3509230005	SUMBER PINANG	PAKUSARI	391577	-2747	606	209792
27	3509240008	KALISAT	KALISAT	393570	-2778	601	441008
28	3509250001	SUREN	LEDOKOMBO	394673	-2790	599	179670
29	3509260001	RANDU AGUNG	SUMBERJAMBE	395059	-2801	598	292367

30	3509260005	SUMBERJAMBE	SUMBERJAMBE	395714	-2811	596	260332
31	3509270010	ARJASA	SUKOWONO	394395	-2793	599	211088
32	3509710004	TEGAL BESAR	KALIWATES	389918	-2720	610	406348
33	3509720003	KARANGREJO	SUMBERSARI	390695	-2733	608	479613
34	3509720005	SUMBERSARI	SUMBERSARI	390405	-2730	609	350673
35	3509730002	JEMBER LOR	PATRANG	389589	-2718	611	414492

Dengan pembobot Gauss Kernel, hasil nilai-nilai penduga parameter dari 35 desa yang digunakan dapat dilihat pada Tabel 1.

Dengan Interpolasi *Inverse Distance Weighted* (IDW) terhadap penduga rata-rata pengeluaran per kapita per desa dengan GWR, hasilnya dapat dilihat pada Gambar 1.



Keterangan Klas Pengeluaran Per Kapita:

- | | | |
|-------------------------------------|-------------------------------------|---------------------|
| 1 : < Rp. 282.500,- | 3 : Rp. 340.290,- s/d Rp. 455.869,- | 5 : > Rp. 571.449,- |
| 2 : Rp. 282.500,- s/d Rp. 340.290,- | 4 : Rp. 455.869,- s/d Rp. 571.449,- | |

Gambar 1. Hasil interpolasi *Inverse Distance Weighted* (IDW) terhadap rata-rata pengeluaran per kapita

ANOVA yang dapat menunjukkan bahwa model GWR dan model OLS menjelaskan hubungan antar peubah sama baiknya, ditolak, sebagai berikut:

	Df	Sum Sq	Mean Sq	F value
OLS Residuals	4.0000	2.6096e+11		
GWR Improvement	0.1937	5.1547e+09	2.6612e+10	
GWR Residuals	30.8063	2.5580e+11	8.3036e+09	3.2048

Dari ANOVA di atas, berarti terdapat perubahan yang signifikan dalam menjelaskan hubungan peubah-peubah yang digunakan, jika analisis yang digunakan adalah GWR. Sehingga dapat dikatakan bahwa rata-rata pengeluaran per kapita tiap desa di Kabupaten Jember lebih baik jika dijelaskan oleh peubah penjelas dengan koefisien bervariasi secara geografis, dibandingkan jika menggunakan regresi global dengan koefisien tetap di seluruh lokasi desa.

6. Kesimpulan dan Saran

Dalam melakukan analisis data perlu kiranya mempertimbangkan adanya faktor spasial yang mungkin mempengaruhi hasil pengukuran data, khususnya jika secara teori data yang dianalisa

sangat mungkin dipengaruhi posisi atau faktor geografis lokasi pengambilan data. Sehingga hasil analisis akan lebih akurat.

Selain GWR terdapat banyak analisis spasial lainnya dengan masing-masing karakter yang kesemuanya melibatkan faktor spasial dalam analisisnya. Untuk saran, kiranya dapat dicoba juga analisis spasial yang lain untuk analisis data kemiskinan ataupun analisis data-data spasial lainnya, sehingga dapat terlihat peran spasial dan perbandingan hasilnya jika dibandingkan analisis yang bersifat global.

7. Daftar Pustaka

- [1] Ant/BEY. 2010. Jember Berpenduduk Miskin Terbesar di Jatim. Nusantara/Rabu 10 Maret 2010. <http://metrotvnews.com/index.php/metromain/news/2010/03/10/12528/Jember-Berpenduduk-Miskin-Terbesar-di-Jatim> [1 Juli 2010].
- [2] BPS. 2002. Identifikasi dan Penentuan Desa Tertinggal 2002, Buku II = Jawa. Badan Pusat Statistik, Jakarta.
- [3] BPS. 2008. Profil Kemiskinan di Indonesia Maret 2008. Berita Resmi Statistik No. 37/07/Th. XI 1 Juli 2008, Jakarta.
- [4] Dimulyo S. 2009. Penggunaan *Geographically Weighted Regression-Kriging* untuk Klasifikasi Desa Tertinggal, dalam Seminar Nasional Aplikasi Teknologi Informasi (SNATI) 2009. Yogyakarta.
- [5] Djunaidy M. 2010. Jumlah Penduduk Miskin Kabupaten Jember Tertinggi, Jum'at 05 Februari 2010. <http://www.tempointeraktif.com/hg/nusa/2010/02/05/brk,20100205-223826,id.html> [1 Juli 2010].
- [6] Fotheringham A.S., Brunson C., Charlton M. 2002. *Geographically Weighted Regression, the analysis of spatially varying relationships*. John Wiley and Sons, LTD.
- [7] Schabenberger O., Gotway C.A. 2005. *Statistical Methods for Spatial Data Analysis*. Chapman & Hall/CRC.
- [8] Setyawan A. 2007. Potensi dan SDA di Kabupaten Jember, September 20, 2007. <http://indahjemberku.wordpress.com/2007/09/20/potensi-sda-di-kabupaten-jember/> [1 Juli 2010].
- [9] Walter J., Carsten R. and Jeremy W. Lichstein. 2005. *Local and Global Approaches to Spatial Data Analysis in Ecology*. *Global Ecology and Biogeography* 14, 97-98.