

ISBN : 978-979-097-142-4



PROSIDING
SEMINAR NASIONAL STATISTIKA
UNIVERSITAS DIPONEGORO
2011



TEMA :

**Peran dan Implementasi Statistika Dalam Analisis Finansial
dan Pengambilan Keputusan Bisnis**

Semarang, 21 Mei 2011

Penyelenggara:
Program Studi Statistika FMIPA UNDIP

**Program Studi Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Diponegoro
2011**

PROSIDING

SEMINAR NASIONAL STATISTIKA

UNIVERSITAS DIPONEGORO

2011

Makalah dalam prosiding ini telah dipresentasikan pada
Seminar Nasional Statistika Universitas Diponegoro
tanggal 21 Mei 2011 di Gedung Prof. Soedharto, SH
Tembalang Semarang

Tim Penyunting Makalah:

Prof. Drs. Mustafid, M.Eng, Ph.D
Dra. Dwi Ispriyanti, M.Si
Erman Deni, SE, MM
Drs. Sudargo, M.Si

Tim Editor:

Dra. Tatik Widiharih, M.Si
Dra. Suparti, M.Si
Drs. Tarno, M.Si
Drs. Rukun Santoso, M.Si
Drs. Sudarno, M.Si

**Program Studi Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Diponegoro**

2011

ANALISIS *GEOGRAPHICALLY WEIGHTED REGRESSION (GWR)* DENGAN PEMBOBOT KERNEL GAUSSIAN UNTUK DATA KEMISKINAN

Rita Rahmawati¹, Anik Djuraidah²

¹⁾Program Studi Statistika, FMIPA Universitas Diponegoro

²⁾Jurusan Statistika, Institut Pertanian Bogor

Abstrak

Analisis yang digunakan dalam data kemiskinan kebanyakan masih bersifat global dan hasilnya diberlakukan untuk semua wilayah. 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 makalah ini pembobot yang digunakan adalah Kernel Gaussian, yang membutuhkan nilai *bandwidth* sebagai parameter jarak yang masih mempengaruhi suatu desa terhadap desa lainnya. *Bandwidth* optimum dapat diperoleh dengan meminimalkan nilai koefisien CV (*cross validation*). Data yang digunakan sebagai studi kasus adalah data 38 kota atau kabupaten di Propinsi Jawa Timur. Hasil menunjukkan bahwa untuk data kemiskinan yang digunakan, lebih baik dianalisis dengan GWR dibandingkan regresi biasa.

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

1. Pendahuluan

Salah satu masalah besar yang hingga saat ini masih dihadapi Indonesia adalah masalah kemiskinan. 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, diantaranya dengan mengidentifikasi wilayah-wilayah miskin, dari tingkat kabupaten atau kota bahkan hingga desa melalui program Instruksi Presiden tentang Desa Tertinggal (IDT).

Dalam menentukan strategi penanggulangan kemiskinan, pada dasarnya akan lebih efektif jika dilakukan dengan pendekatan geografis, yang berarti berhubungan

juga dengan sumber daya alam dan manusia di setiap wilayah. Hakim & Zuber (2008) menyatakan bahwa lokasi tempat tinggal, akses ke teknologi dan ketersediaan sumber alam berpengaruh terhadap kemiskinan. Untuk itu perlu ada upaya pendekatan analisis yang melibatkan unsur lokasi (faktor geografis) untuk mengolah data kemiskinan. 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 analisis yang 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 Kernel Gaussian.

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:

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

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. Namun berbeda dengan regresi global yang diberlakukan secara umum di setiap lokasi pengamatan, GWR 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 Kernel Gaussian yaitu:

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

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 koefisien validasi silang, dengan rumus:

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

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

Untuk mendeteksi secara global apakah GWR lebih baik daripada OLS, dapat diuji dengan *analysis of variance* (ANOVA) yang diusulkan Brunsdon *et al.* (1999) sebagai berikut:

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

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 , δ_1^2/δ_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 $F_{\text{hit}} > F_\alpha(v_1^2/v_2, \delta_1^2/\delta_2)$.

4. Data

Data yang digunakan dalam makalah ini adalah data sekunder dari BPS, yaitu data Potensi Desa (Podes) dan Susenas tahun 2008. Wilayah yang digunakan adalah 38 kota atau kabupaten di Propinsi Jawa Timur. Untuk peubah respon (Y) digunakan *Headcount Index* kemiskinan, yaitu persentase penduduk yang berada di bawah Garis Kemiskinan. Sedangkan untuk peubah bebas, dipilih sejumlah peubah berdasarkan studi BPS (2002) serta ketersediaan data pada Podes 2008, yang selanjutnya disaring kembali menggunakan regresi *stepwise*. Peubah-peubah bebas yang akhirnya digunakan dalam analisis adalah yang secara signifikan mempengaruhi peubah respon, yaitu :

- X1 = persentase penduduk yang mempunyai pendidikan rendah (di bawah SD)
- X2 = persentase rumah tangga yang menggunakan air minum yang berasal dari air mineral, air PAM, pompa air, sumur atau mata air terlindung
- X3 = persentase penduduk yang diperbolehkan membeli beras (raskin) dengan harga murah bersubsidi, dan
- X4 = persentase penduduk yang mendapat surat miskin yang merupakan kelompok rumah tangga di bawah 20% kelompok pengeluaran terbawah.

5. Hasil Dan Pembahasan

Dengan analisis regresi biasa (global), model persamaan regresi yang dihasilkan adalah $\hat{Y} = 2.69 - 0.824X_1 + 0.279X_2 + 0.244X_3 + 0.392X_4$, dengan rincian penduga parameter dan ANOVA sebagai berikut:

Tabel 1. Penduga Parameter

Penduga	Koeffisien	Koeff. SE	t-hitung	nilai-p
Konstanta	2.6879	0.4688	5.73	0.000
X1	0.82420	0.07189	11.46	0.000
X2	0.27933	0.07147	3.91	0.000
X3	0.24439	0.09863	2.48	0.019
X4	0.39175	0.04270	9.17	0.000

Tabel 2. ANOVA

Sumber Keragaman	DF	SS	MS	F	P
Regression	4	1874.13	468.53	440.97	0.000
Residual Error	33	35.06	1.06		
Total	37	1909.19			

Hasil di atas, diasumsikan sama dan digunakan untuk semua wilayah kota atau kabupaten di Propinsi Jawa Timur.

Langkah pertama untuk analisis GWR adalah menentukan *bandwidth* yang akan digunakan dalam fungsi pembobot Kernel Gaussian. Dengan iterasi hingga didapatkan CV minimum, diperoleh nilai *bandwidth* 125.884 km yang menghasilkan nilai CV 49.79414, sehingga fungsi pembobot spasial GWR-nya menjadi :

$$w_j(i) = \exp\left[-1/2\left(\frac{d_{ij}}{125.884}\right)^2\right]$$

Nilai *bandwidth* dalam fungsi pembobot di atas menggambarkan batas jarak suatu wilayah yang masih memberikan pengaruh cukup besar terhadap wilayah lain di sekitarnya. Berarti jika jarak antar kota atau kabupaten lebih dari 125.884 km maka

pengaruhnya semakin kecil dan akan turun lebih cepat dengan semakin bertambahnya jarak antar kota atau kabupaten.

Dengan analisis GWR, maka nilai-nilai parameter yang dihasilkan adalah sebanyak lokasi data yang digunakan. Untuk memperoleh hasil analisis ini, penulis menggunakan software R. Rangkuman nilai-nilai penduga parameter model GWR dengan fungsi pembobot Kernel Gaussian adalah sebagai berikut:

Tabel 3. Penduga Parameter Model GWR

Penduga	Min.	Q1	Med.	Q3	Max.	Global
Konstanta	2.55400	2.70800	2.78600	2.84000	2.86800	2.6879
X1	0.82190	0.83030	0.84280	0.86600	0.90460	0.8242
X2	0.19090	0.20960	0.23160	0.27530	0.35860	0.2793
X3	0.06256	0.17970	0.23580	0.27130	0.30980	0.2444
X4	0.31600	0.38630	0.41050	0.42100	0.43030	0.3918

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

Tabel 4. ANOVA

Sumber Keragaman	Df	SS	MS	F-hitung	nilai-p
OLS Residuals	5.0000	35.062			
GWR Improvement	4.1332	10.140	2.45339		
GWR Residuals	28.8668	24.922	0.86335	2.8417	0.004156

Dari ANOVA di atas, dapat dilihat bahwa dengan menggunakan GWR maka nilai sisaan (*residuals*) akan jauh berkurang. Nilai-p yang kurang dari 5%, menunjukkan bahwa hipotesis nol yang menyebutkan bahwa dengan tingkat kepercayaan 95% regresi global sama baiknya dengan GWR, ditolak. Berarti terdapat perubahan yang signifikan dalam menjelaskan hubungan peubah-peubah yang digunakan, jika analisis yang digunakan adalah GWR. Sehingga dapat dikatakan bahwa *Headcount Index* kemiskinan kota atau kabupaten di Jawa Timur lebih baik jika dijelaskan oleh peubah-peubah

penjelas dengan koefisien bervariasi secara geografis, dibandingkan jika menggunakan regresi global dengan koefisien tetap di seluruh lokasi kota atau kabupaten.

6. Kesimpulan

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. Dengan demikian hasil analisis diharapkan akan lebih akurat.

Daftar Pustaka

- BPS. 2002. Identifikasi dan Penentuan Desa Tertinggal 2002, Buku II = Jawa. Badan Pusat Statistik, Jakarta.
- BPS. 2008. Profil Kemiskinan di Indonesia Maret 2008. Berita Resmi Statistik No. 37/07/Th. XI 1 Juli 2008, Jakarta.
- Brunsdon C, Fotheringham AS, Charlton M. 1999. *Some notes on parametric significance tests for geographically weighted regression*, Journal of Regional Science, Vol. 39, No 3, 497- 524.
- Fotheringham A.S., Brunsdon C., Charlton M. 2002. *Geographically Weighted Regression, the analysis of spatially varying relationships*. John Wiley and Sons, LTD.
- Hakim L, Zuber A. 2008. Dimensi Geografis dan Pengentasan Kemiskinan Pedesaan. Media Ekonomi. Jakarta: Fakultas Ekonomi Universitas Trisakti.
- Schabenberger O., Gotway C.A. 2005. *Statistical Methods for Spatial Data Analysis*. Chapman & Hall/CRC.
- 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.