

# Simulasi Radius Jarak Pengaruhnya terhadap Keباikan Model Regresi Logistik Spasial<sup>1</sup>

Utami Dyah Syafitri<sup>2</sup>, Agus M Sholeh<sup>2</sup>, Poppy Suprapti<sup>3</sup>

## Abstrak

Pemodelan regresi logistik dengan basis ruang spasial perlu mengakomodir pengaruh spasial tersebut. Salah satu pendekatan untuk mengakomodir pengaruh spasial tersebut dengan pendekatan variogram. Pembobot spasial yang diberikan tergantung jarak antar desanya. Pada penelitian ini akan dilakukan simulasi radius jarak yang digunakan dalam pembobotan untuk keباikan hasil pendugaan dengan regresi logistik spasial. Studi kasus yang diambil dalam penelitian ini adalah pendugaan status kemiskinan desa di Jawa Barat.

Hasil simulasi menunjukkan bahwa jarak yang digunakan dalam pembobot spasial mempengaruhi hasil pendugaan yang dihasilkan. Berdasarkan nilai *c-statistic* dan *correct clasification rate*-nya, bahwa jarak radius yang semakin dekat akan menghasilkan prediksi yang lebih baik.

*Kata kunci* : model regresi logistik spasial, variogram, simulasi

## PENDAHULUAN

### Latar Belakang

Pemodelan regresi logistik dengan basis ruang (spasial) perlu mengakomodir pengaruh spasial tersebut. Beberapa pendekatan yang dapat digunakan dalam mengakomodir pengaruh spasial antara lain dengan pendekatan *generalized linear mixed model*, variogram, serta matriks kontiguity.

Pratama (2008) melakukan pemodelan regresi logistik spasial terhadap status kemiskinan desa di Jawa Barat. Pembobot spasial yang digunakan menggunakan pendekatan fungsi-fungsi variogram. Pembobot spasial dilakukan pada jarak desa yang ber-radius 27,5 km. Hasil yang diperoleh menunjukkan bahwa regresi logistik spasial dengan pendekatan variogram menunjukkan tingkat akurasi prediksi sedikit lebih baik dibandingkan dengan regresi logistik klasik. Namun hasil yang didapatkan belum cukup memuaskan. Salah satu faktor yang ditengarai mampu memperbaiki model adalah definisi jarak desa yang digunakan. Sehingga dalam penelitian ini dilakukan simulasi berbagai kemungkinan jarak yang digunakan untuk memperbaiki pendugaan model regresi logistik spasial.

## Tujuan

Penelitian ini bertujuan untuk mengetahui jarak antar desa yang menghasilkan prediksi yang lebih baik dari Pratama (2008) dalam menduga status kemiskinan desa di Jawa Barat.

## Kerangka Pikir Penelitian

Andaikan sebuah bidang  $S$  tersekat-sekat  $n$  buah sub bidang (lokasi) yang saling lepas, yaitu  $s_1, s_2, \dots, s_n$  dengan  $s_1 \cup s_2 \cup \dots \cup s_n = S$  dan  $s_i \cap s_j = \emptyset$ . Setiap subbidang  $s_i$  memiliki sentroid pada titik  $c_i$ . Nilai pengukuran  $Y, X_1, X_2, \dots, X_p$  di masing-masing sub-bidang adalah  $y_i, x_{1i}, x_{2i}, \dots, x_{pi}$ .

Jika  $Y$  memiliki hubungan spasial, dan dipengaruhi oleh  $X_1, X_2, \dots, X_p$  maka model yang bisa dibentuk adalah :  $y_i = \mathbf{x}_i' \boldsymbol{\beta} + \mathbf{z}_i' \boldsymbol{\beta}^* + \varepsilon_i \dots (1)$  dengan  $\mathbf{x}_i = (1, x_{1i}, x_{2i}, \dots, x_{pi})'$  merupakan vektor  $p \times 1$  yang berisi nilai-nilai pengamatan peubah  $X_1, X_2, \dots, X_p$  pada sub-bidang  $s_i$  dan  $\mathbf{y}' = (y_1, y_2, \dots, y_n)$  adalah nilai-nilai pengamatan peubah  $Y$  dari seluruh lokasi.

Vektor  $\mathbf{z}$  berukuran  $n \times 1$  menyatakan bobot spasial antar sub-bidang terhadap nilai di sub-bidang lainnya. Menggunakan notasi matriks, model (1) dapat dituliskan sebagai  $\mathbf{y} = \mathbf{X} \boldsymbol{\beta} + \mathbf{Z} \boldsymbol{\beta}^* + \boldsymbol{\varepsilon} \dots (2)$  pada model (2) terdapat  $\mathbf{Z}$  yang merupakan sebuah matriks pembobot spasial. Matriks pembobot spasial diharapkan dapat menggambarkan pengaruh

<sup>1</sup> Diseminarkan pada Seminar Nasional Matematika dan Pendidikan Matematika di Universitas Negeri Yogyakarta, 45  
Jumat 28 November 2008,

<sup>2</sup> Staf Pengajar Departemen Statistika FMIPAIPB

<sup>3</sup> Mahasiswa Departemen Statistika FMIPA IPB

antar desa, dimana desa yang berdekatan memberikan pengaruh yang lebih besar dibandingkan dengan desa yang berjauhan.

Variogram yang akan digunakan untuk menghitung matriks pembobot spasial menggambarkan keragaman antar daerah berdasarkan jaraknya. Semakin jauh jarak antar daerah maka keragaman yang terbentuk akan semakin besar menuju kekonvergenan. Pada jarak tertentu maka keragaman yang terbentuk mulai menuju titik kekonvergenan. Jika titik telah konvergen maka pengaruh spasialnya juga semakin berkurang. Sehingga batasan jarak yang digunakan adalah di sekitar titik mulai kekonvergenan (sill).

Berkaitan dengan pembobotan, desa – desa yang mempunyai jarak yang dekat diberikan pembobot yang besar. Sebaliknya desa-desa yang berjarak jauh diberi pembobot yang kecil. Oleh karena itu, dilakukan proses pembalikan matriks variogram kemudian hasil pembalikan matriks variogram dijadikan matriks pembobot spasial. Matriks pembobot spasial ( $Z$ ) dikalikan dengan vektor  $y$  yang kemudian dianggap sebagai sebuah peubah penjelas baru ( $w$ ) dan akan digunakan dalam analisis regresi logistik.

## TINJAUAN PUSTAKA

### Regresi Logistik Spasial

Hosmer dan Lemeshow (1989) menjelaskan bahwa yang membedakan model regresi logistik dan model regresi linear adalah peubah hasil pada regresi logistik berskala biner atau dikhotom, sedangkan pada regresi linear peubah hasilnya berskala numerik. Model regresi logistik dengan  $E(Y=I|x)$  Sebagai  $\pi(x)$  adalah :

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}}$$

dalam regresi logistik diperlukan fungsi penghubung logit, transformasi logit sebagai fungsi dari  $\pi(x)$  adalah

$$g(x) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

Preisler *et al.* (1995) dalam Fernandez (2003) menyebutkan bahwa pendekatan dengan memasukkan hubungan spasial ke dalam model terdapat dua pendekatan yaitu

memasukkan lokasi ke dalam model dan memasukkan suku autologistik. Augustin *et al.* (1996) dalam Fernandez (2003) menggunakan model dalam bentuk :

$$y = \log\left[\frac{\pi(x)}{1-\pi(x)}\right] + \beta\psi$$

$$\psi_i = \frac{\sum_{j=1}^k w_{ij} \hat{y}_j}{w_{ij}}$$

model dari  $\psi$  merupakan bentuk dari autokovarian dan merupakan rata-rata terboboti dari jumlah kejadian dalam suatu lokasi ke-I yang terdiri dari  $k$  tetangganya. Pembobot dari lokasi ke-j adalah  $w_{ij} = 1/h_{ij}$  dimana  $h_{ij}$  adalah jarak *euclidean* antara lokasi ke-i dan ke-j. Serta  $\hat{y}$  adalah dugaan dari ada/tidaknya suatu kejadian.

### Variogram

Analisis variogram melakukan penghitungan pada sejumlah lokasi dan melihat hubungan antar observasi pada berbagai lokasi. Variogram menghitung hubungan antara perbedaan pengukuran berpasangan dan jarak dari poin-poin yang bersesuaian satu sama lain. Variogram merupakan keragaman spasial antar lokasi dengan saling ketergantungan satu sama lain dalam ruang berdimensi  $m$ . Variogram merupakan fungsi spasial terbaik yang diketahui (Ashraf *et al.*, 1997). Persamaan umum untuk contoh variogram adalah (Matheron 1962, dalam Cressie 1993):

$$2\gamma(h) = \frac{1}{N(h)} \sum_{N(h)} (Z(x_i) - Z(x_j))^2; i, j \in N(h)$$

dengan  $N(h)$  adalah banyaknya pasangan lokasi (contoh) yang berjarak  $h$ . Variogram akan memenuhi beberapa asumsi. Misalkan terdapat gugusan nilai  $z(x_i)$  pada lokasi  $x_i$ ,  $i=1,2,3, \dots, n$  dalam ruang berdimensi  $m$ , maka asumsi yang harus terpenuhi adalah (Cressie, 1993):

1.  $E(Z(x+h) - Z(x)) = 0$
2.  $Var(Z(x+h) - Z(x)) = 2\gamma(h)$

dengan  $h$  adalah jarak antara dua lokasi yang terpisah. Sifat dari variogram, adalah:

1. Monoton tidak turun
2. Bernilai positif

### Model Power

Hubungan antara variogram dengan jarak untuk model power dibentuk dalam persamaan :

$$\gamma(h) = c_0 + ph^\alpha, \text{ dengan } 0 < \alpha < 2$$

dengan :

$c_0$  = intersep

$p$  = kemiringan kurva

$h$  = jarak antar pengamatan

### Correct Classification Rate (CCR)

Hosmer dan Lemeshow (1989) menjelaskan bahwa *correct classification rate* adalah persentasi kebenaran (kesesuaian) nilai pengamatan dengan dugaannya.

$$CCR = \frac{\text{banyaknya dugaan yang benar}}{\text{banyaknya pengamatan}} \times 100\%$$

## BAHAN DAN METODOLOGI

### Bahan

Data yang digunakan dalam penelitian ini sama dengan Pratama (2008) menggunakan data Potensi Desa (PODES) tahun 2006. Wilayah yang digunakan hanya sebagian dari Jawa Barat (Tabel 1).

Tabel 1. Daftar nama kota dan kabupaten yang digunakan

Kode Kabupaten	Nama Kabupaten	Jumlah Desa
3201000000	Kab Bogor	415
3202000000	Kab Sukabumi	340
3203000000	Kab Cianjur	344
3204000000	Kab Bandung	436
3213000000	Kab Subang	248
3214000000	Kab Purwakarta	190
3215000000	Kab Karawang	304
3216000000	Kab Bekasi	179
3271000000	Kota Bogor	63
3272000000	Kota Sukabumi	33
3273000000	Kota Bandung	139
3275000000	Kota Bekasi	43
3277000000	Kota Cimahi	15
Total		2749

Baik peubah penjas maupun peubah respon yang digunakan dalam penelitian ini mengacu kepada peubah yang digunakan oleh Pratama (2008). Peubah-peubah penjas yang digunakan selengkapnya dapat dilihat pada Tabel 2. Peubah respon yang digunakan adalah status kemiskinan desa. status kemiskinan desa ditentukan berdasarkan presentase keluarga miskin pada masing-masing desa. Apabila persentase keluarga miskin suatu desa lebih besar dari persentase keluarga miskin secara keseluruhan (36%) maka desa tersebut dikategorikan sebagai desa miskin. Sebaliknya dikategorikan sebagai desa tidak miskin. Sedangkan untuk memperoleh informasi hubungan spasial dibutuhkan data mengenai lokasi dari masing-masing desa (lintang dan bujur desa).

Tabel 2. Peubah penjas yang digunakan

Peubah	Jenis	Keterangan
Ada/Tidak penduduk desa yang Bekerja Sebagai TKI ( $X_1$ )	Kategorik	1 = ada 0 = tidak ada
Persentase Keluarga Yang menerima kartu sehat ( $X_2$ )	Numerik	Jumlah penerima Jumlah keluarga
Persentase luas sawah ( $X_3$ )	Numerik	Luas lahan sawah Luas desa/kelurahan
Persentase Rumah tangga Yang menggunakan listrik ( $X_4$ )	Numerik	Jumlah pemakai Jumlah keluarga
Ada/Tidak puskesmas di Daerah tersebut ( $X_5$ )	Kategorik	1 = ada 0 = tidak ada

## Metodologi

Alur penelitian yang dilakukan adalah:

1. Membuat matriks jarak euklid antar desa berdasarkan lintang dan bujur desa.
2. Dari fungsi variogram model power yang dilakukan oleh Pratama (2008) dan matriks jarak dibuat matriks peragam spasial. Pada penelitian jarak yang digunakan pada radius 7.5 km, 10 km, 15 km, 20km, 25 km, 27.5 km, dan 30 km.
3. Membalikan matriks peragam spasial yang akan dijadikan matriks pembobot spasial ( $Z$ )
4. Membuat peubah penjelas baru ( $w$ ) yang telah diberi pengaruh spasial dengan mengalikan  $Z$  dan  $y$
5. Melakukan pendugaan dengan menggunakan regresi logistik yang telah ditambahkan peubah penjelasnya.
6. Membandingkan hasil pendugaan regresi logistik menggunakan matriks pembobot berdasarkan simulasi radius jarak variogram yang ditentukan.

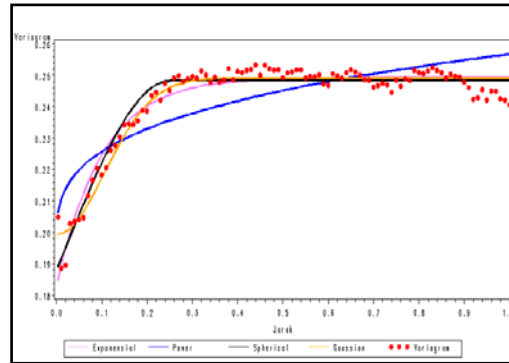
Penentuan radius jarak variogram terbaik dapat dilihat dari nilai *correct classification rate* dan *c-statistic*. Perangkat lunak yang digunakan dalam penelitian ini adalah SAS ver 9.1 dan Microsoft Office Excel 2003.

## HASIL DAN PEMBAHASAN

Dugaan fungsi variogram model power yang dilakukan Pratama (2008) adalah  $\hat{\gamma}(h) = 0.1753h^{0.3} + 0.0802$ . Model tersebut mempunyai koefisien determinasi sebesar 73.21%. Dibandingkan dengan model variogram lainnya model power menghasilkan prediksi yang kurang bagus dibandingkan dengan model variogram lainnya (Gambar 1).

Gambar 1. Plot dugaan dari masing-masing model variogram

Namun dari segi pemodelan regresi logistik spasial model ini lebih bagus. Maka dalam penelitian ini selanjutnya akan



menggunakan model power ini.

Faktor-faktor yang mempengaruhi status kemiskinan desa dengan radius jarak lebih besar atau sama dengan 20 km pada taraf alfa 10% adalah ada/tidak penduduk yang bekerja sebagai TKI, jarak dari desa ke ibukota kabupaten/kota, presentase keluarga yang menerima kartu sehat, persentase luas sawah, persentase keluarga yang memakai listrik, ada/tidak puskesmas di desa tersebut, dan hubungan spasial. Sedangkan untuk radius jarak kurang dari 20 km dengan taraf alfa 10%, faktor-faktor yang mempengaruhi status kemiskinan desa adalah jarak dari desa ke ibukota kabupaten/kota, presentase keluarga yang menerima kartu sehat, persentase luas sawah, ada/tidak puskesmas di desa tersebut, dan hubungan spasial.

Dari berbagai simulasi jarak yang dilakukan terlihat bahwa nilai *correct classification rate* dan *c-statistic* semakin menurun seiring dengan bertambahnya radius jarak yang digunakan (Tabel 3). Hal tersebut menunjukkan bahwa daerah yang berdekatan akan memberikan pengaruh spasial yang cukup signifikan terhadap kebaikan pendugaan model.

Tabel 3. Nilai dugaan koefisien regresi logistik dengan berbagai simulasi jarak

Peubah	Dugaan koefisien						
	7.5 km	10 km	15 km	20 km	25 km	27.5 km	30 km
<i>Intercept</i>	2.9321	2.846	2.7393	2.6583	2.6988	2.7299	2.6913
Ada/tidak penduduk yang bekerja sebagai TKI	-0.1368*	-0.1436*	-0.1391*	-0.1835	-0.2186	-0.2305	-0.2428
Jarak dari desa ke ibukota kabupaten/kota	-0.018	-0.0179	-0.0172	-0.0161	-0.0154	-0.0157	-0.0159
Persentase keluarga yang menerima kartu sehat	-2.4536	-2.4424	-2.5127	-2.5456	-2.6021	-2.6364	-2.7053
Persentase luas sawah	-0.8902	-0.9264	-1.0475	-1.1434	-1.2346	-1.271	-1.3047
Persentase keluarga yang memakai listrik	0.0382*	0.1306	0.2593*	0.3806	0.4423	0.4702	0.4819
Ada/tidak puskesmas di desa tersebut	0.2283	0.2438	0.2567	0.2636	0.2563	0.252	0.246
X spasial	-4.218	-4.1598	-4.0465	-3.9481	-3.9719	-4.0039	-3.871
<i>C-statistic</i>	80.10%	78.40%	76.50%	75%	74.30%	74.00%	73.60%
<i>Correct classification rate</i>	71.73%	70.46%	69.33%	67.44%	67.00%	66.89%	66.42%

\* menunjukkan peubah tidak bersifat nyata pada alfa 10%

### KESIMPULAN

Pembobot spasial yang digunakan dalam model regresi logistik spasial tergantung terhadap radius jarak yang digunakan. Apabila pembobotan dilakukan dalam radius jarak yang lebih dekat maka akan dihasilkan hasil prediksi yang lebih baik.

### DAFTAR PUSTAKA

Ashraf, M., Jim C. L., K. G. Hubbard. 1997. *Application of Geostatistics to Evaluate Partial Weather Station Network*. J. Agricultural and Forest Meteorology. 84:255-271.

Bohling, Geoff. 2005. *Introduction to Geostatistics and Variogram Analysis*. Kansas Geological Survey 864-2093.

Cressie, NAC. 1993. *Statistics for Spatial Data*. Canada: John Wiley and Sons, Inc. 11

Frei, Allan. 2005. *Notes On Spatial Autocorrelation and Regression*. Hunter College.

Hosmer DW, Lemeshow S. 1989. *Applied Logistic Regression*. New York: J. Wiley.

Pratama, Vinda. 2008. *Perbandingan Hasil Akurasi Prediksi Model Regresi Spasial untuk Berbagai Model Variogram*. Skripsi. Fakultas Matematika dan Ilmu Pengetahuan Alam Institut Pertanian Bogor.

Searle, S. R. 1971. *Linear Models*. John Wiley & Sons, Inc: New York, hal 1-7.

Xie, Chenglin, Bo Huang, Christophe Claramunt and Magesh Chandramouli. 2005. *Spatial Logistic Regression and GIS to Model Rural-Urban Land Conversion*.