

***Geographically Weighted Ridge Regression* dalam Kasus Multikolinearitas Pada Indeks Pembangunan Manusia di Kabupaten/Kota Provinsi Jawa Timur**

Yani Arthayanti

Program Studi Matematika, Fakultas MIPA – Universitas Udayana
Email: yaniarthayanti@gmail.com

I Gusti Ayu Made Srinadi

Program Studi Matematika, Fakultas MIPA – Universitas Udayana
Email: srinadi@unud.ac.id

G.K. Gandhiadi

Program Studi Matematika, Fakultas MIPA – Universitas Udayana
Email: gandhiadigk@yahoo.com

Abstract: *Linear Regression Analysis is a statistical method for modeling relation between two variable, response and explanatory variable. Geographically Weighted Regression (GWR) is the development of linier regression analysis if the case of spatial divers case. Local multicollinearity is a condition when explanatory variables had correlated with each observation location. Geographically Weighted Ridge Regression (GWRR) is a method used to model data containing local multicollinearity on spatial data. GWRR model was developed from ridge regression by adding weight as additional information. The study aims to model spatial data containing local multicollinearity to the Human Development Index (HDI) in the districts/municipalities of eastern Java Province in 2015. The result of this study was indicate that the indicator of the average length of school is a dominant indicator that affects HDI.*

Keywords: *GWR, GWRR, HDI, Local Multicollinearity, Spatial Data*

1. Pendahuluan

Analisis Regresi merupakan metode statistika yang digunakan untuk memodelkan hubungan variabel respon dengan variabel penjelas. Menurut Anselin [1], apabila menggunakan model regresi linear pada pengamatan yang dipengaruhi oleh faktor geografis atau data spasial, maka dapat menyebabkan kesimpulan yang kurang tepat karena unsur spasial tidak tercakup di dalamnya. Metode yang digunakan untuk data yang dipengaruhi faktor geografis adalah regresi spasial.

Menurut Fotheringham *et.al.*[2] metode statistika yang dapat digunakan untuk

memodelkan hubungan variabel respon dan penjelas pada data yang mengandung heteroskedastisitas adalah *Geographically Weighted Regression* (GWR). Dalam metode GWR digunakan matriks pembobot yang besarnya tergantung pada kedekatan antar lokasi pengamatan. Semakin dekat suatu lokasi pengamatan maka pengaruh bobotnya akan semakin besar.

Dalam aplikasi GWR terkadang ditemukan adanya variabel penjelas yang saling berkorelasi, permasalahan tersebut dinamakan multikolinearitas lokal. Multikolinearitas lokal adalah suatu kondisi dengan variabel penjelas saling berkorelasi di setiap lokasi pengamatan. Adanya multikolinearitas dapat menyebabkan pendugaan parameter dari model yang dihasilkan memiliki varians yang besar, sehingga dapat menyebabkan terjadinya kesalahan dalam menginterpretasikan parameter.

Menurut Hastie *et.al.* [3] Regresi Ridge merupakan metode yang dapat digunakan pada data yang mengandung multikolinearitas. Regresi ridge merupakan pengembangan dari metode kuadrat terkecil yaitu meminimumkan jumlah kuadrat galat dengan menambahkan tetapan bias. Pada data spasial yang mengandung multikolinearitas lokal dapat digunakan modifikasi regresi ridge dan GWR yang disebut dengan *Geographically Weighted Ridge Regression* (GWRR). GWRR merupakan pengembangan dari regresi ridge yaitu dengan menambahkan unsur pembobot sebagai informasi tambahan.

Indeks Pembangunan Manusia merupakan indikator kinerja pembangunan yang dibentuk melalui pendekatan tiga dimensi dasar. Tiga dimensi tersebut terdiri dari kesehatan, pengetahuan, dan kehidupan yang layak (BPS [4]). Jawa Timur merupakan salah satu provinsi besar di Indonesia dengan jumlah penduduk mencapai 38 juta jiwa lebih dan juga merupakan provinsi yang memiliki kabupaten/kota terbanyak di Indonesia yaitu 38 kabupaten/kota (BPS [5]). IPM di Jawa Timur pada tahun 2015 menduduki peringkat 16 sebesar 68,95[4]. Pembangunan pada kabupaten/kota di Jawa Timur masih belum merata, karena terdapat kesenjangan yang cukup tinggi. Berdasarkan [4] nilai IPM pada kota Malang sebesar 80,05 merupakan nilai IPM terbesar di Jawa Timur, sedangkan nilai IPM terendah adalah di kabupaten Sampang hanya sebesar 58,18. Maka dari itu perlu dilakukan penelitian terhadap factor-faktor yang berpengaruh terhadap Indeks Pembangunan Manusia di Provinsi Jawa Timur

Berdasarkan uraian di atas, penulis akan mengaplikasikan metode GWRR dalam memodelkan Indeks Pembangunan Manusia di kabupaten/kota Provinsi Jawa Timur tahun 2015. Adapun variabel respon yang digunakan dalam penelitian ini adalah nilai IPM kabupaten/kota Provinsi Jawa Timur sedangkan variabel penjelas yang digunakan adalah Angka Harapan Hidup (AHH), Angka Melek Huruf (AMH), Rata-rata Lama Sekolah (RLS), Pengeluaran per Kapita (PPK), Angka Kematian Bayi (AKB) dan Penduduk usia di atas 15 tahun yang bekerja (PK).

2. Metode Penelitian

Metode Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah berupa data sekunder yang diperoleh dari Badan Pusat Statistika (BPS) Provinsi Jawa Timur berupa data Indeks Pembangunan Manusia (IPM) dan faktor-faktor yang memengaruhi IPM pada tahun 2015.

Metode Analisis Data

Berikut merupakan tahapan – tahapan analisis dalam penelitian ini, yaitu :

1. Mendeskripsikan variabel respon dan variabel - variabel penjelas Indeks Pembangunan Manusia.
2. Menguji heteroskedastisitas pada data dengan menggunakan uji *Glejser*.
3. Menentukan model GWR dan mendeteksi multikolinearitas lokal dengan menggunakan *Variance Inflation Factor* (VIF) dengan rumus

$$VIF_k = \frac{1}{1-R_k^2(u_i, v_i)} \quad (1)$$

$$R_{(u_i, v_i)}^2 = \frac{\sum_{j=1}^p w_{ij}(y_j - \hat{y}_j)^2}{\sum_{j=1}^p w_{ij}(y_j - \bar{y})^2}. \quad (2)$$

4. Melakukan pemodelan *Geographically Weighted Ridge Regression* (GWRR) dengan tahapan sebagai berikut:

- a. Menghitung jarak *euclidean* serta menduga nilai *bandwidth* dan tetapan bias c dengan metode iterasi *Cross Validation* (CV) dengan rumus $CV =$

$$\sum_{i=1}^n (y_i - \hat{y}_{\neq i}(b))^2. \quad (3)$$

- b. Menghitung jarak *Euclidean* dengan rumus $d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2}$ dan menentukan matriks pembobot untuk setiap lokasi pengamatan dengan fungsi *fixed exponential* yaitu dengan rumus:

$$w_{ij} = \exp \left[- \left(\frac{d_{ij}}{b} \right) \right]. \quad (5)$$

5. Menduga nilai parameter model regresi setiap lokasi pengamatan berdasarkan matriks pembobot dengan rumus:

$$\hat{\beta}_R(u_i, v_i) = [X^{*T}W(u_i, v_i)X^* + cI]^{-1}X^{*T}W(u_i, v_i)Y^*. \quad (6)$$

6. Melakukan pengujian serentak pada model *Geographically Weighted Ridge Regression* (GWRR) dengan rumus:

$$F_{hitung} = \frac{\frac{y^{*T}y^*}{n}}{\frac{y^{*T}(I-G)^T(I-G)y^*}{u_1}} \quad (7)$$

dan pengujian parsial pada model *Geographically Weighted Ridge Regression* (GWRR) dengan rumus: $t_{hitung} = \frac{\hat{\beta}_k(u_i, v_i)}{\hat{\sigma} \sqrt{C_{kk}}}$.

7. Interpretasikan Model.

(8)

3. Hasil dan Pembahasan

Analisis Deskriptif

Langkah awal dalam penelitian ini adalah melakukan analisis deskripsi untuk masing-masing variabel penjelas dan variabel respon. Analisis deskriptif disajikan pada Tabel 1.

Tabel 1. Analisis Desriptif

| Variabel | N | Min | Maks | Standar Deviasi |
|----------------|----|--------|---------|-----------------|
| Y | 38 | 58.18 | 80.05 | 5.405 |
| X ₁ | 38 | 65.73 | 73.85 | 2.079 |
| X ₂ | 38 | 78.03 | 98.86 | 5.272 |
| X ₃ | 38 | 4.84 | 11.22 | 1.587 |
| X ₄ | 38 | 532525 | 1722999 | 276347.03 |
| X ₅ | 38 | 16.05 | 56.24 | 11.369 |
| X ₆ | 38 | 63806 | 1365180 | 311294.05 |

Tabel 1 menunjukkan bahwa terdapat kesenjangan nilai IPM terbesar dan terendah di Jawa Timur yaitu nilai Indeks Pembangunan Manusia (Y) di Provinsi Jawa Timur tahun 2015 terbesar yaitu 80.05 berada di kota Malang sedangkan nilai IPM terendah berada pada kabupaten Sampang yaitu sebesar 58.18. Dapat diketahui bahwa nilai IPM di provinsi Jawa Timur sebesar 68.95 [4].

Heteroskedastisitas

Untuk mengetahui adanya ketaksamaan varians dari residual pengamatan satu ke pengamatan lainnya dilakukan uji heterokedastisitas. Uji yang digunakan adalah uji *glejser* yaitu dengan meregresikan nilai dari absolute residual pengamatan terhadap variabel penjelas. Langkah awal untuk memperoleh residual pengamatan adalah meregresikan variabel-variabel penjelasnya terhadap variabel respon. Dengan bantuan software SPSS 23, diperoleh *p-value* sebesar 0.036. Jika *p-value* kurang dari taraf kepercayaan $\alpha = 0.05$ maka data pengamatannya mengandung heterokedastisitas, karena *p-value* 0.036 sehingga *p value* lebih kecil dari 0.05 maka data tersebut mengandung heterokedastisitas.

Multikolinearitas

Mendeteksi adanya multikolinearitas lokal dengan menggunakan kriteria VIF. Jika variabel penjelas memiliki nilai VIF lebih besar dari lima maka variabel tersebut mengandung multikolinearitas (Montgomery & Peck, 1992). Dengan menggunakan software R 3.2.5, diperoleh bahwa pada variabel penjelas yang mengalami multikolinearitas lokal adalah Angka Harapan Hidup (X_1), Angka Melek Huruf (X_2), Rata-Rata Lama Sekolah (X_3), dan Pengeluaran Perkapita (X_4). Oleh sebab itu digunakan metode *Geographically Weighted Ridge Regression* (GWRR).

Pemodelan *Geographically Weighted Ridge Regression*

Pemodelan data spasial yang mengandung multikolinearitas dapat menggunakan *Geographically Weighted Ridge Regression* yaitu dengan menambahkan tetapan bias c tertentu pada matriks $X^T X$. Pada GWRR nilai tetapan bias c diperoleh secara bersamaan dengan metode iterasi untuk setiap *bandwidth* (b) yaitu tetapan bias c dan *bandwidth* (b) yang terpilih adalah yang meminimumkan nilai CV (*Cross Validation*)[6].

Langkah awal pada metode ini adalah membentuk matriks pembobot yaitu dengan menggunakan fungsi kernel *exponential*. Dengan bantuan software R versi 3.2.5, diperoleh nilai *bandwidth* optimum sebesar 362203 dan tetapan bias c yang optimum adalah 0.01419 untuk setiap lokasi pengamatan. Jika suatu lokasi semakin jauh dari titik lokasi pengamatan maka nilai pembobotnya semakin menurun sehingga pengaruhnya semakin kecil dan juga sebaliknya. Untuk memperoleh pembobot masing-masing lokasi pengamatan dengan rumus Persamaan 5, yaitu:

$$w_{ij} = \exp \left[- \left(\frac{d_{ij}}{b} \right) \right]$$

$$= \exp \left[- \left(\frac{d_{ij}}{362203} \right) \right]$$

Dengan d_{ij} merupakan jarak *Euclid* masing-masing lokasi pengamatan. Sebagai contoh matriks pembobot untuk lokasi pengamatan di Kabupaten Ponorogo adalah:

$$W_{(u_i, v_i)} = \begin{bmatrix} 0.8623457 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0.7249052 \end{bmatrix}$$

Lebih lanjut lagi adalah pendugaan parameter model GWRR. Pemodelan GWRR dapat diselesaikan dengan bantuan software R versi 3.2.5, nilai *bandwidth* dan tetapan bias c digunakan dalam pendugaan parameter pada model GWRR. Sebagai contoh model GWRR untuk Kabupaten Ponorogo adalah sebagai berikut:

$$\hat{y}_{Ponorogo} = 49.8637 - 0.0417X_1 + 0.011X_2 + 2.4365X_3 + 3.76 \times 10^{-6}X_4 - 0.0389 X_5 + 1.23 \times 10^{-7}X_6$$

Uji Serentak GWRR

Uji serentak dilakukan untuk mengetahui signifikansi variabel-variabel penjelas pada model GWRR secara serentak. Dengan menggunakan software R versi 3.2.5 diperoleh F_{hitung} yaitu sebesar 18.57733. Dengan derajat bebas sebesar 38 dan 37.43246 maka diperoleh F_{tabel} sebesar 1.725073. Karena F_{hitung} lebih besar dari F_{tabel} maka kesimpulannya menolak H_0 yang artinya bahwa variabel – variabel penjelas signifikan berpengaruh terhadap model GWRR.

Uji Parsial GWRR

Setelah diketahui bahwa terdapat hubungan antara variabel-variabel penjelas dengan model GWRR selanjutnya dilakukan uji parsial. Uji parsial dilakukan untuk mengetahui parameter di setiap lokasi pengamatan yang berpengaruh signifikan terhadap Indeks Pembangunan Manusia di Jawa Timur tahun 2015.

Pengujian parsial menggunakan taraf kepercayaan $\alpha = 0.05$ dan derajat bebas sebesar 37 menghasilkan nilai t_{tabel} sebesar 2.02619, apabila nilai t_{hitung} di setiap lokasi pengamatan lebih besar dari t_{tabel} maka kesimpulannya menolak H_0 yang berarti variabel penjelas tersebut berpengaruh signifikan terhadap Indeks Pembangunan Manusia. Sebagai contoh untuk Kabupaten Ponorogo diperoleh hasil sebagai berikut:

Tabel 2. t_{hitung} untuk Kabupaten Ponorogo

| Variabel | t_{hitung} | Keterangan |
|----------|--------------|------------------|
| X_1 | -0.1880 | Tidak signifikan |
| X_2 | 0.1598 | Tidak signifikan |
| X_3 | 5.0345 | Signifikan |
| X_4 | 2.3241 | Signifikan |
| X_5 | -0.6428 | Tidak signifikan |
| X_6 | 0.0850 | Tidak signifikan |

Berdasarkan Tabel 2 variabel penjelas yang berpengaruh signifikan terhadap Indeks Pembangunan Manusia di Kabupaten Ponorogo adalah variabel Rata-rata lama sekolah (X_3) dan Pendapatan Perkapita (X_4).

Interpretasi Model

Setelah diperoleh model GWRR untuk setiap lokasi pengamatan dan dilakukan pengujian signifikansi. Maka selanjutnya adalah interpretasi model GWRR yang telah diperoleh. Contoh interpretasi model GWRR pada Kabupaten Ponorogo yaitu:

$$\hat{y}_{Ponorogo} = 49.8637 - 0.0417X_1 + 0.011X_2 + 2.4365X_3 + 3.76 \times 10^{-6}X_4 - 0.0389 X_5 + 1.23 \times 10^{-7}X_6$$

Model GWRR Indeks Pembangunan Manusia di Kabupaten Ponorogo, menunjukkan bahwa variabel yang dominan dan signifikan memengaruhi Indeks Pembangunan Manusia adalah variabel rata-rata lama sekolah (X_3).

Berdasarkan pemodelan GWRR pada Indeks Pembangunan Manusia di Provinsi Jawa Timur, secara umum diperoleh indikator yang dominan memengaruhi IPM di masing-masing Kabupaten/Kota Provinsi Jawa Timur tahun 2015 adalah rata-rata lama sekolah.

4. Simpulan dan Saran

Simpulan Penelitian

Berdasarkan hasil dan pembahasan diperoleh beberapa kesimpulan yaitu:

1. Berdasarkan pemodelan GWRR pada Indeks Pembangunan Manusia di Provinsi Jawa Timur diperoleh 38 model GWRR berbeda di setiap lokasinya. Selanjutnya diperoleh indikator yang secara umum dominan memengaruhi IPM adalah rata-rata lama sekolah. Berdasarkan variabel penjelas yang berpengaruh signifikan di masing-masing Kabupaten/Kota dibagi menjadi dua kelompok.
2. Faktor – faktor yang berpengaruh terhadap Indeks Pembangunan Manusia tahun 2015 di setiap Kabupaten/Kota Provinsi Jawa Timur adalah variabel Rata-rata Lama Sekolah (X_3), sedangkan untuk variabel Pengeluaran Perkapita (X_4) berpengaruh di beberapa kabupaten/Kota di Jawa Timur kecuali di Kabupaten Pacitan, Blitar dan Kota Blitar.

Saran

Berdasarkan hasil analisis pada pembahasan, peneliti menggunakan modifikasi Regresi Ridge pada GWR untuk memodelkan data yang mengandung multikolinearitas sehingga diharapkan untuk penelitian selanjutnya yang menggunakan pemodelan GWR dengan kasus multikolinearitas dapat menggunakan modifikasi dengan metode lain seperti *Principal Component Analysis* (PCA) atau *Lasso*. Selanjutnya juga dapat dilakukan penelitian lanjutan dengan menggunakan variabel penelitian lain atau menambahkan variabel penjelas lain.

Daftar Pustaka

- [1] Anselin, L., 1988. *Spatial Econometrics : Methods and Models*. The Netherlands: Kluwer Academic Publishers.
- [2] Fotheringham, A. C. M. d. B. C., 2002. *Geographically Weighted Regression : The Analyse os Spatially Varying Relationships*. England: John Wiley and Sons Ltd.
- [3] Hastie, T., Tibshirani, R. & Friedman, J., 2009. *The Elements of Statistical Learning Data Mining, Inference, and Prediction*. New York: Springer.
- [4] BPS, 2016. *Indeks Pembangunan Manusia 2015*. In: Jakarta: Badan Pusat Statistika.
- [5] BPS, Jawa Timur, 2016. *Jawa Timur Dalam Angka 2016*. In: Surabaya: BPS Provinsi Jawa Timur.
- [6] Wheeler, D. C., 2007. Diagnostic Tools and A Remedial Method for Collinearity in Geographically Weighted Regression. *Environment and Planning A*, Volume 39, pp. 2464-2481.