

SELEKSI KOEFISIEN MODEL PRODUK DOMESTIK REGIONAL BRUTO KABUPATEN/KOTA PULAU JAWA MENGGUNAKAN *GEOGRAPHICALLY WEIGHTED LASSO (GWL)*

Embay Rohaeti^{1,2}, Khairil Anwar Notodiputro^{2*}, Bagus Sartono²

¹ Program Studi Matematika Universitas Pakuan, Bogor, Jawa Barat, Indonesia.

² Departemen Statistika, IPB University, Bogor, Jawa Barat, Indonesia.

*email: khairil@apps.ipb.ac.id

Abstrak. Keheterogenan spasial dan *local collinearity* (multikolinearitas lokal) masih menjadi masalah dalam analisis data spasial karena hal tersebut dapat meningkatkan ragam dari pendugaan koefisien regresi. Masalah ini dapat diatasi dengan model *Geographically Weighted Lasso (GWL)* yang merupakan pengembangan dari *Geographically Weighted Regression (GWR)* dengan menerapkan teknik LASSO dalam pendugaannya. Dalam penelitian ini dilakukan seleksi koefisien dari model produk domestik regional bruto (PDRB) Kabupaten/Kota di Pulau Jawa yang mengandung masalah keheterogenan spasial dan kekolinearitas lokal sebagai konsekuensi dari adanya perbedaan kondisi geografis dan kebijakan ekonomi di setiap lokasi pengamatan. Dengan menerapkan algoritma LARS (*Least Angle Regression*) yang dimodifikasi, diperoleh hasil bahwa *Geographically Weighted Lasso* dapat mengatasi kekolinearitas lokal pada model PDRB Kabupaten/Kota di Pulau Jawa, dengan menyeleksi peubah pada setiap lokasi. Peubah yang memiliki koefisien sama dengan nol akan terseleksi untuk wilayah tertentu, sehingga dugaan parameter model menjadi lebih stabil.

Kata Kunci: keheterogenan spasial, *local collinearity*, *geographically weighted regression*, *geographically weighted lasso*.

1. Pendahuluan

Keheterogenan spasial masih menjadi masalah dalam analisis data spasial dikarenakan ragam yang tidak selalu homogen pada setiap lokasi pengamatan. Mengatasi masalah tersebut dapat diatasi dengan pemodelan *Geographically Weighted Regression (GWR)* yang dapat mengeksplorasi keragaman spasial (Fotheringham *et al.* 2002). Selain itu menurut Wheeler (2009) bahwa terdapat masalah yang biasanya muncul pada GWR yaitu *local collinearity* (multikolinearitas lokal) pada koefisien yang diduga, yang dapat meningkatkan ragam dari pendugaan koefisien regresi, untuk mengatasinya Wheeler (2009) mengusulkan metode *Geographically Weighted Lasso (GWL)* yang merupakan pengembangan dari GWR dengan menerapkan teknik lasso dalam pendugaannya sehingga hasil dugaan yang diperoleh menjadi lebih stabil. Berdasarkan hal tersebut GWL akan diterapkan pada pemodelan produk domestik regional bruto Kabupaten/Kota di Pulau Jawa.

Menurut BPS (2018) produk domestik regional bruto (PDRB) merupakan salah satu indikator penting untuk mengetahui kondisi ekonomi di suatu daerah dalam suatu periode tertentu, baik atas dasar harga berlaku maupun atas dasar harga konstan. Pertumbuhan PDRB bergantung pada jumlah penduduk, persentasi penduduk miskin, tenaga kerja, tingkat pengangguran terbuka, dana alokasi umum, dana alokasi daerah dan pendapatan asli daerah. (Nasution, 2010). Pertumbuhan jumlah penduduk dan jumlah tenaga kerja merupakan faktor penting yang dapat mendorong meningkatnya PDRB suatu daerah dikarenakan semakin banyaknya penduduk yang bekerja maka akan semakin banyak yang menghasilkan output, tetapi sebaliknya jika semakin tinggi tingkat pengangguran terbuka dan persentase penduduk miskin akan meningkatkan ketidakmampuan dalam menghasilkan output, selain itu semakin besar pendapat asli suatu daerah, mendapatkan dana alokasi umum dan dana alokasi khusus dari pemerintah pusat maka pertumbuhan PDRB meningkat (Wahyoedi, 2009).

Dalam penelitian ini akan dilakukan seleksi koefisien dari model produk domestik regional bruto (PDRB) Kabupaten/Kota di Pulau Jawa yang mengandung masalah keheterogenan spasial dan kekolinearan lokal sebagai konsekuensi dari adanya perbedaan kondisi geografis dan kebijakan ekonomi di setiap lokasi pengamatan. Proses seleksi model menerapkan algoritma LARS (*Least Angle Regression*) dari Efron *et al* (2004) yang dimodifikasi, kemudian untuk mengetahui kebaikan model yang diperoleh maka dibandingkan kinerja dari GWL dengan metode kuadrat terkecil dan GWR berdasarkan nilai R^2 .

1.1 Geographically Weighted Regression (GWR)

Fotheringham *et al.* (2002) dan Wheeler (2009), Model *Geographically Weighted Regression* (GWR) pada lokasi ke $i = 1, 2, \dots, n$ yaitu:

$$y(i) = \mathbf{X}(i)\boldsymbol{\beta}(i) + \varepsilon(i) \quad (1)$$

dengan $y(i)$ peubah respon di lokasi ke i , $\mathbf{X}(i)$ peubah penjelas di lokasi ke i , $\boldsymbol{\beta}(i)$ koefisien regresi di lokasi ke i , $\varepsilon(i)$ galat di lokasi ke i . Pendugaan koefisien regresi di lokasi ke i adalah

$$\hat{\boldsymbol{\beta}}(i) = [\mathbf{X}^T \mathbf{W}(i) \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{W}(i) \mathbf{y} \quad (2)$$

dengan $\mathbf{X} = [\mathbf{X}^T(1); \dots; \mathbf{X}^T(n)]^T$ matrik peubah penjelas $\mathbf{W}(i) = \text{diag}[w_1(i), \dots, w_n(i)]$ diagonal matrik pembobot yang dihitung untuk setiap lokasi ke i dan y adalah peubah respon, dan $\hat{\boldsymbol{\beta}}(i) = (\hat{\beta}_{i0}, \hat{\beta}_{i1}, \dots, \hat{\beta}_{ip})^T$ koefisien regresi lokal di lokasi ke i untuk peubah penjelas p .

1.2 Fungsi Pembobot Spasial

Dalam Wheeler (2009), matriks pembobot $\mathbf{W}(i)$ dengan bobot dari fungsi eksponensial kernel antara lokasi ke j dan lokasi ke i lokasi dihitung dengan:

$$w_j(i) = \exp\left(-\frac{d_{ij}}{\phi}\right) \quad (3)$$

Dengan d_{ij} adalah jarak antara titik di lokasi ke i dan lokasi ke j yang diperoleh dari jarak Euclidean, sedangkan ϕ merupakan *bandwidth* kernel pada lokasi ke i .

1.3 Cross Validation (CV)

Fotheringham *et al.* (2002), formula *Cross Validation* (CV) sebagai berikut:

$$CV(h) = \sum_{i=1}^n [y_i - \hat{y}_{\neq i}(h)]^2 \quad (4)$$

dengan $\hat{y}_{\neq i}(h)$ merupakan nilai dugaan untuk y_i dengan pengamatan ke i dihilangkan dari proses prediksi, dan *bandwidth* optimum (h) akan diperoleh dengan proses iterasi sampai diperoleh CV yang minimum. Sehingga dugaan peubah respon diekspresikan dengan $\hat{y}(i) = X(i)\hat{\beta}(i)$. (Wheeler, 2009).

1.4 Least Absolute Shrinkage and Selection Operator (LASSO)

Tibshirani (1996) mendefinisikan LASSO sebagai berikut :

$$\hat{\beta}^R = \arg \min_{\beta} \left\{ \sum_{i=1}^n (y_i - \beta_0 - \sum_{k=1}^p x_{ij}\beta_k)^2 + \lambda \sum_{k=1}^p |\beta_k| \right\} \quad (5)$$

Dengan syarat $\sum_{k=1}^p |\beta_k| \leq s$, dimana s adalah parameter yang mengontrol besarnya penyeleksian koefisien regresi.

1.5 Geographically Weighted Lasso (GWL)

Wheeler (2009), menerapkan teknik LASSO pada Model *Geographically Weighted Regression* (GWR) yang kemudian disebut dengan model *Geographically Weighted Lasso* (GWL) untuk mengatasi masalah heterogenitas spasial dan multikolinieritas lokal. Pada penyelesaian GWL digunakan algoritma LARS (Efron, *et al.* 2004) yang dimodifikasi dengan menambahkan matriks pembobot pada peubahnya, dengan algoritma sebagai berikut:

- a. Menduga *bandwidth* kernel yang optimum dengan *cross validation* (CV) :
 Menghitung matrik pembobot W berukuran $n \times n$ dari matriks jarak pada persamaan (3), dengan $[w_1(i), \dots, w_n(i)]$ adalah matriks diagonal dari $W(i)$ yang didefinisikan pada persamaan (2) .
- b. Untuk setiap lokasi ke i , $i = 1, \dots, n$:
 - 1) Bentuk $W^{\frac{1}{2}}(i) = \text{sqr}t(\text{diag}(W(i)))$
 - 2) Bentuk $X_w = W^{\frac{1}{2}}(i)X$ dan $y_w = W^{\frac{1}{2}}(i)y$ menggunakan akar kuadrat dari pembobot kernel $W(i)$ di lokasi ke i .
 - 3) Panggil algoritma lars (X_w, y_w), simpan rangkain penyelesaian Lasso, yang meminimalkan galat y_i dan simpan penyelesaian ini.
- c. Berhenti ketika jika hanya ada perubahan kecil dalam dugaan ϕ , lalu simpan dugaan ϕ tersebut.

2. Metode Penelitian

Data yang digunakan dalam pemodelan yaitu data sekunder dari BPS berupa Produk Domestik Regional Bruto (PDRB) atas dasar harga konstan Tahun 2019 pada 118 Kabupaten/Kota di Pulau Jawa, dengan peubah yang digunakan :

Peubah	Keterangan
Y	PDRB
X_1	Jumlah penduduk
X_2	Presentasi penduduk miskin
X_3	Tingkat pengangguran terbuka
X_4	Jumlah tenaga kerja
X_5	Dana alokasi umum
X_6	Dana alokasi khusus
X_7	Pendapatan asli daerah

Data PDRB tersebut mengandung efek spasial yang memiliki keragaman tidak homogen pada setiap lokasi pengamatannya, maka untuk mengeksplorasi keragaman spasial tersebut, akan dilakukan pemodelan *Geographically Weighted Regression* (GWR), dengan model sebagai berikut:

$$y(i) = \beta_0(i) + \sum_{k=1}^7 \beta_k(i) X_k(i) + \varepsilon(i), \text{ dengan } k = 1, \dots, 7 \text{ dan } i = 1, \dots, 118 .$$

dimana $y(i)$ adalah peubah respon pada lokasi ke i untuk data PDRB, $X_k(i)$ adalah peubah penjelas ke k pada lokasi ke i , $\beta_k(i)$ adalah koefisien parameter lokal untuk lokasi ke i dan sisaan diasumsikan $\varepsilon(i) \sim N(0, I\sigma^2)$.

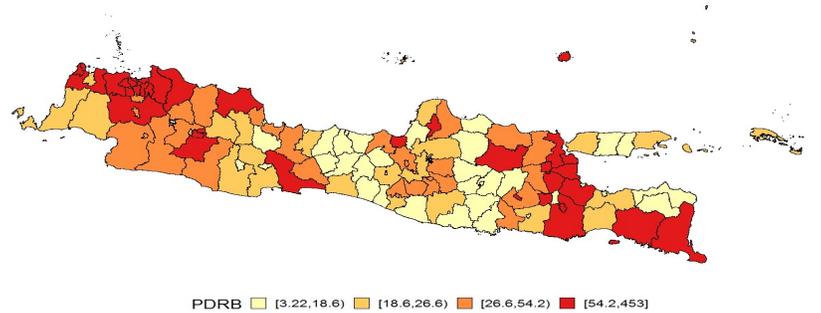
Selain itu dikarenakan menggunakan beberapa peubah penjelas yang dapat menyebabkan multikolinearitas lokal, maka setiap lokasi memiliki koefisien parameter yang berbeda-beda, sehingga metode pendugaan koefisien model lokal yang terbentuk dilakukan dengan *Geographically Weighted Lasso* (GWL) yaitu dengan menyeleksi peubah pada setiap lokasi. Peubah yang memiliki koefisien sama dengan nol akan terseleksi untuk wilayah tertentu, sehingga hasil dugaan yang diperoleh menjadi lebih stabil. Adapun tahapan analisisnya sebagai berikut :

- a. Membuat peta sebaran PDRB di Pulau Jawa.
- b. Melakukan uji heterogenitas spasial dengan uji Breusch Pagan.
- c. Menentukan matriks jarak dengan jarak Euclid.
- d. Menentukan *bandwidth* optimum berdasarkan CV minimum dengan kernel eksponensial.
- e. Memodelkan PDRB Kabupaten/Kota Pulau Jawa dengan GWR.
- f. Menentukan *local collinearity* dari model lokal yang terbentuk.
- g. Menyeleksi koefisien model PDRB dengan GWL.
- h. Menentukan kebaikan model GWL berdasarkan perbandingan dengan MKT dan GWR dilihat dari nilai R^2 .

Tahapan analisis dalam pemodelan dan seleksi menggunakan algoritma LARS (Efron *et al*, 2004) yang dimodifikasi.

3. Hasil dan pembahasan

Sebelum dilakukan pemodelan dengan *Geographically Weighted Regression* (GWR), pertama dilakukan pemetaan data PDRB (dalam trilyun rupiah) di 118 Kabupaten/Kota di Pulau Jawa, dengan hasil pemetaan memperlihatkan daerah dengan PDRB sangat rendah, rendah, sedang dan tinggi yang ditandai dengan perbedaan warna untuk daerahnya. Adapun hasil pemetaan data PDRB terdapat pada gambar 1.



Gambar 1. Peta PDRB Kabupaten/Kota di Pulau Jawa Tahun 2019.

3.1 Uji Heterogenitas Spasial

Nilai PDRB berbeda-beda pada setiap lokasi dan sangat beragam seperti diperlihatkan pada gambar 1, maka dilakukan uji Breusch-Pagan untuk mengetahui adanya keragaman spasial antar lokasi pengamatan, dengan menggunakan hipotesis sebagai berikut :

H_0 : Tidak terdapat heterogenitas spasial

H_1 : Terdapat heterogenitas spasial

Pengujian dengan uji Breusch-Pagan menghasilkan nilai-p kurang dari taraf nyata $\alpha = 5\%$, hal ini menunjukkan terdapat heterogenitas spasial pada model PDRB. Selanjutnya akan dilakukan pemodelan dengan GWR untuk menangani adanya heterogenitas spasial tersebut.

3.2 Geographically Weighted Regression (GWR)

Tahapan awal dalam pemodelan GWR adalah menentukan *bandwidth* (ϕ) atau lebar jendela, kemudian akan digunakan untuk membentuk matriks pembobot spasial. Besarnya *bandwidth* optimum merupakan nilai *bandwidth* dengan CV yang minimum. *Bandwidth* optimum pada penelitian ini adalah sebesar 1,17 atau sebesar 130,344 km. *Bandwidth* diilustrasikan sebagai jari-jari lingkaran di sekitar titik pengamatan sehingga wilayah yang berada di sekitar jarak tersebut masih memberikan pengaruh terhadap PDRB di wilayah tersebut.

Model GWR yang terbentuk yaitu model lokal dari 118 Kabupaten/Kota di Pulau Jawa, sehingga terbentuk 118 model, maka dugaan koefisien dari model lokal

tersebut sangat banyak, sehingga berikut ini diberikan ringkasan hasil dugaan koefisien tujuh peubah dari model lokal sebagai berikut :

Tabel 1. Ringkasan penduga koefisien GWR

Koefisien	Minimum	Maksimum	Rata-rata
b ₀	-18764.66	37181.35	-8526.80
b ₁	-221298.08	-3455.78	-33573.13
b ₂	-14423.29	180.37	-2310.09
b ₃	246.18	14242.09	3371.48
b ₄	0.03	0.60	0.11
b ₅	-179.68	41.57	3.73
b ₆	-156.48	397.82	-20.53
b ₇	-4.82	65.32	47.34

3.3 Local Collinearity (Multikolinearitas lokal)

Selanjutnya untuk mengetahui adanya *local collinearity* pada model lokal dari GWR, dapat dilihat berdasarkan pada nilai VIF lokal yang terbentuk, dimana jika nilai $VIF \geq 5$ atau $VIF \geq 10$ menunjukkan adanya multikolinearitas lokal (Friday dan Enenonye, 2012). Hasil dari perhitungan nilai VIF pada peubah penjelas di 118 lokasi menunjukkan tiap lokasi terdapat multikolinearitas lokal. Untuk mengatasi hal tersebut maka diterapkan *Geographically Weighted Lasso*.

3.4 Geographically Weighted Lasso (GWL)

Model GWL merupakan modifikasi model GWR dengan menambahkan algoritma LARS sehingga hasil pendugaan koefisien akan berbeda-beda untuk setiap wilayah amatan. Perbedaan nilai koefisien ini merupakan langkah untuk mengatasi adanya keragaman spasial pada model. Perbedaan koefisien ini juga dapat mengakibatkan perbedaan pengaruh peubah pada setiap wilayah. Tabel 2 menyajikan ringkasan koefisien model GWL dari 118 wilayah. Tabel 2 juga menyajikan jumlah koefisien yang disusutkan sampai 0 yang kemudian peubah dengan koefisien tersebut akan terseleksi.

Selanjutnya hasil pendugaan parameter model GWL dengan algoritma LARS yang dimodifikasi dalam proses seleksi koefisien model PDRB pada Kabupaten/Kota di Pulau Jawa dikelompokkan menjadi tiga nilai yaitu yang memberikan pengaruh positif, pengaruh negatif, dan memiliki nilai koefisien 0.

Peubah dengan nilai koefisien 0 ini merupakan koefisien dari peubah yang terseleksi. Adapun hasil seleksi ditunjukkan dalam bentuk pemetaan model seperti yang disajikan pada Gambar 2.

Tabel 2 Ringkasan koefisien GWL.

Koefisien	Minimum	Maksimum	Rata-rata	Jumlah koefisien 0
b ₁	-603543.11	599346.78	27606.70	29
b ₂	-64358.78	11852.07	-4844.12	20
b ₃	-18793.11	77041.34	4154.71	19
b ₄	-1.213	0.885	-0.065	24
b ₅	-446.42	687.38	25.71	27
b ₆	-1420.39	1411.53	2.90	19
b ₇	-190.38	231.36	50.62	1

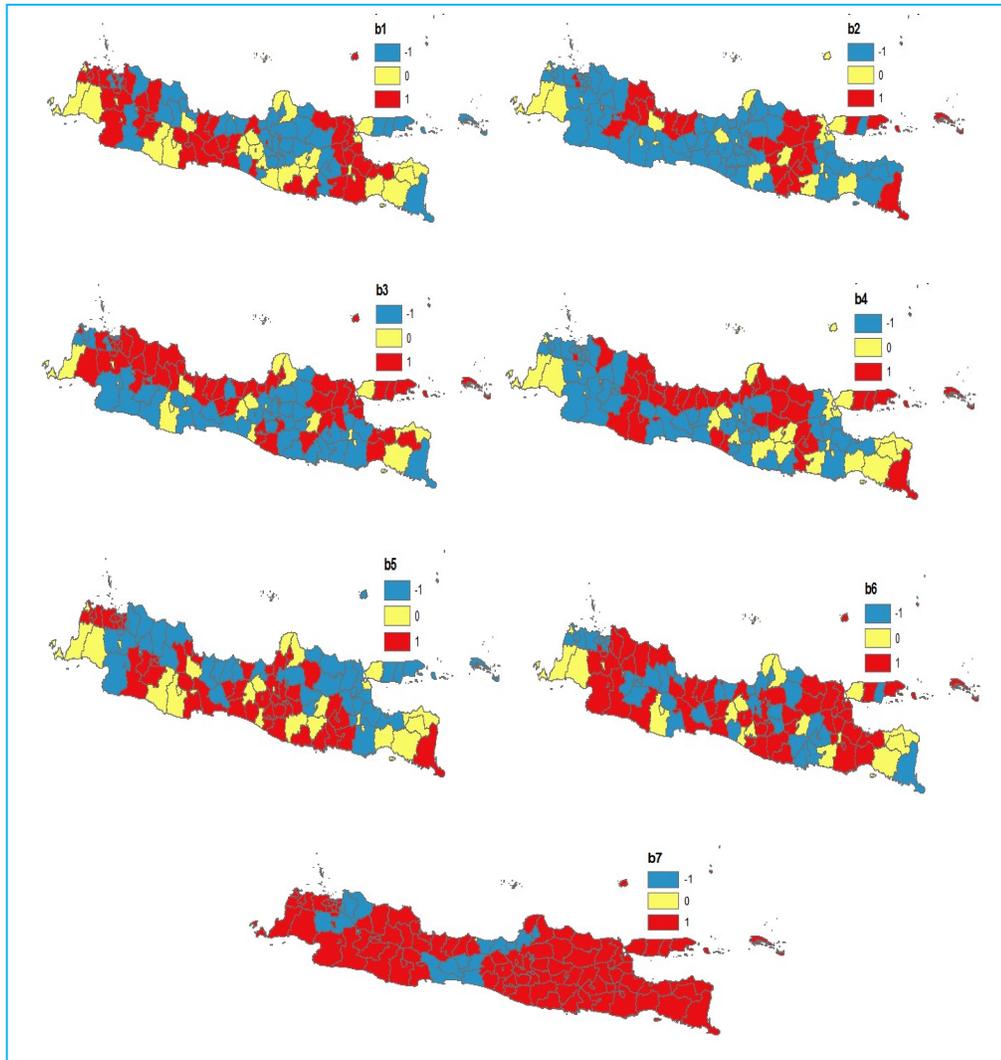
Pada Gambar 2 (dengan urutan ke samping) merupakan hasil pemetaan dari seleksi koefisien model GWL untuk peubah X₁ sampai X₇ di setiap lokasi. Wilayah dengan warna kuning menunjukkan terseleksi, sedangkan wilayah dengan warna biru menunjukkan bahwa peubah memberikan pengaruh negatif di wilayah tersebut dan wilayah dengan warna merah menunjukkan bahwa peubah memberikan pengaruh positif di wilayah tersebut.

Untuk peubah X₁ yang memiliki koefisien sama dengan nol berarti terseleksi untuk beberapa wilayah diantaranya Kabupaten Pandeglang, Kabupaten/Kota Tasikmalaya, Kabupaten Jepara, Kabupaten Sleman, Kabupaten Jember. Sedangkan yang memberikan pengaruh negatif pada dugaan nilai PDRB diantaranya Kabupaten Kuningan, Kabupaten Demak, Kabupaten Bantul, Kabupaten Sumenep. Kemudian yang memberikan pengaruh positif diantaranya Kota Serang, Kota Bogor, Kota Semarang, Kabupaten Malang. Dengan proses seleksi yang sama pada peubah X₂ sampai X₆ diperoleh peubah-peubah yang terseleksi di wilayah tertentu. Sedangkan pada peubah X₇ hanya ada satu wilayah yang memiliki koefisien sama dengan nol yaitu Kota Pekalongan. Sehingga diperoleh model lokal dengan nilai dugaannya yang lebih stabil.

Salah satu model lokal yang terbentuk setelah adanya proses seleksi diantaranya di Kabupaten Garut dugaan nilai PDRB dipengaruhi oleh persentase penduduk miskin, tingkat pengangguran terbuka, tenaga kerja, alokasi dana khusus

dan pendapatan asli daerah, untuk peubah lainnya di Kabupaten Garut terseleksi dikarenakan adanya penyusutan koefisien sampai dengan nol. Sedangkan salah satu wilayah yang tidak mengalami proses seleksi di semua peubah penjelas diantaranya yaitu Kota Jogjakarta.

Gambar 2. Hasil seleksi koefisien model PDRB Kabupaten/Kota di Pulau Jawa



3.5 Kebaikan Model

Untuk melihat kebaikan model secara keseluruhan dihitung menggunakan nilai R-square. Sebagai bahan perbandingan tambahan maka dengan bantuan software R juga dihitung nilai R-square untuk metode kuadrat terkecil pada model

PDRB Kabupaten/Kota di Pulau Jawa. Adapun hasil perbandingan kinerja dari ketiga metode tersebut menunjukkan bahwa *Geographically Weighted Lasso* memiliki nilai R-square yang lebih tinggi, hal ini berarti model yang dihasilkan baik. Hasil perbandingan ditunjukkan pada tabel 2 berikut :

Tabel 3. Perbandingan kinerja model

Model	R-square
Metode Kuadrat Terkecil (MKT)	76.66
<i>Geographically Weighted Regression</i> (GWR)	84.64
<i>Geographically Weighted Lasso</i> (GWL)	90.32

4. Kesimpulan

Model *Geographically Weighted Lasso* dapat mengatasi kekolinearitas lokal pada model PDRB Kabupaten/Kota di Pulau Jawa, dengan menyeleksi peubah pada setiap lokasi. Peubah yang memiliki koefisien sama dengan nol akan terseleksi untuk wilayah tertentu, sehingga dugaan parameter model menjadi lebih stabil.

Daftar Pustaka

- Badan Pusat Statistik. (2018). *Tinjauan Regional Berdasarkan PDRB Kabupaten/Kota 2014-2018 Buku 2 Pulau Jawa dan Bali*. Jakarta. www.bps.go.id.
- Efron B, Hastie T, Johnstone I, Tibshirani R. (2004). Least Angle Regression. *The Annals of Statistics* 32(2): 407-451.
- Friday OR, Emenonye C. 2012. The Detection and Correction of Multicollinearity Effects in a Multiple Regression Diagnostics. *Elixir Statistics* 49:10108-10112. doi.10.1063/1.4979419.
- Fotheringham AS, Brunson C, Charlton M. (2002). *Geographically Weighted Regression the Analysis of Spatially Varying Relationships*. England (GB): John Wiley and Sons.
- Tibshirani R. 1996. Regression Shrinkage and Selection Via The Lasso. *Journal of the Royal Statistical Society B* 58(1): 267-288.
- Wahyoedi S. 2009. Kemampuan Dana Alokasi Umum dalam Meningkatkan Pendapatan Asli Daerah dan Produk Domestik Regional Bruto Di Indonesia Tahun 2001-2003. *Jurnal Ilmiah Manajemen Bisnis*, Vol. 9, No. 3: 139 –154.
- Wheeler DC. 2009. Simultaneous Coefficient Penalization and Model Selection in Geographically Weighted Regression: The Geographically Weighted Lasso. *Journal of Environment and Planning A* 41 (3): 722-742.