



NON-HYBRID ENSEMBLE SPATIAL REGRESSION ON HUMAN DEVELOPMENT INDEX (IPM) in CENTRAL JAVA

Evi Ardiati Sazaen^{1*}, Rochdi Wasono², Indab Manfaati Nur³

^{1,2,3} Department of Statistics, Universitas Muhammadiyah, Semarang, Indonesia

* E-mail: evisazaen@gmail.com

ABSTRACTS

The human development index (HDI) is a measure to see an increase in regional development that has a very broad dimension, because it increases the quality of the population of an area in terms of life expectancy, education, and decent standard of living. In 2010 the Central Java HDI increased by 66.08% and increased by 4.44%, with the total HDI in 2017 of 70.52 percent. Spatial regression is the development of classical linear regression involving the region model. Spatial regression ensemble is a technique to be sent spasi spatial regression models by adding noise (additive noise). The type of spatial weighting used is Queen Contiguity. The selection of the best model using AIC and RMSE values. The purpose of this study is to provide an assessment of the distribution of HDI data in the Province of Central Java in 2017 and to do modeling using non-hybrid spatial ensemble regression regression. The results of this study are the SAR spatial method with ensemble giving results with AIC value of 143 and RMSE value of 1.3899 with a value of R^2 90.09%. Significant variables on HDI are population density (X_1), poverty (X_2), school participation rates (X_5), and average per capita per month for food and non-food (X_7).

ABSTRAK

Indeks pembangunan manusia (IPM) merupakan ukuran untuk melihat dampak kinerja pembangunan wilayah yang mempunyai dimensi yang sangat luas, karena memperlihatkan kualitas penduduk suatu wilayah dalam hal harapan hidup, pendidikan, dan standar hidup layak. Pada tahun 2010 IPM Jawa Tengah berkisar 66,08% dan meningkat sebesar 4,44%, dengan angka IPM pada tahun 2017 sebesar 70,52 persen. Regresi spasial merupakan pengembangan dari regresi linier klasik yang melibatkan unsur wilayah (spasial) kedalam model. Regresi spasial *ensemble* merupakan suatu teknik untuk menggabungkan k model regresi spasial dengan menambahkan *noise* (additive noise). Jenis pembobot spasial yang digunakan yaitu *Queen Contiguity*. Pemilihan model terbaik menggunakan AIC dan nilai RMSE. Tujuan dari penelitian ini yaitu memberikan gambaran penyebaran data IPM di Provinsi Jawa Tengah Tahun 2017 dan melakukan pemodelan dengan menggunakan pendekatan regresi spasial *ensemble non-hybrid*. Hasil dari penelitian ini adalah metode spasial SAR dengan *ensemble* memberikan hasil dengan nilai AIC 143 dan nilai RMSE 1,3899 dengan nilai R^2 90,09%. Variabel yang berpengaruh secara signifikan terhadap IPM yaitu kepadatan penduduk (X_1), kemiskinan (X_2), angka partisipasi sekolah (X_5), dan rata-rata pengeluaran per kapita per bulan makanan dan non makanan (X_7).

ARTICLE INFO

Article History:

Received 07 Oct 2020

Revised 16 Oct 2020

Accepted 16 Oct 2020

Available online 18 Oct 2020

Keyword:

Human development index, Spatial ensemble non-hybrid regression, Queen contiguity, additive noise, AIC,

Keyword:

Indeks pembangunan manusia, Regresi spasial *ensemble non-hybrid*, *Queen contiguity*, *Additive noise*, AIC,

1. PENDAHULUAN

Pembangunan menjadi salah satu upaya dalam meningkatkan kemajuan dan kesejahteraan masyarakat. Kemajuan suatu negara tidak hanya diukur dari hasil pembangunan infrastruktur dan teknologi akan tetapi faktor utama yang berpengaruh adalah Sumber Daya Manusia (SDM). Menurut *United Nations Development Program* (UNDP) (1990) pembangunan manusia sebagai upaya perluasan pilihan bagi masyarakat dan sekaligus sebagai taraf yang dicapai dari upaya tersebut. Salah satu indikator dalam mengukur kemajuan pembangunan manusia adalah *Human Development Index* (HDI) atau lebih dikenal dengan indeks pembangunan manusia (IPM). IPM menjelaskan bagaimana penduduk dapat mengakses hasil pembangunan dalam memperoleh pendapatan, kesehatan, pendidikan, dan sebagainya (BPS, 2010). IPM merupakan ukuran untuk melihat dampak kinerja pembangunan wilayah yang mempunyai dimensi yang sangat luas, karena memperlihatkan kualitas penduduk suatu wilayah dalam hal harapan hidup, pendidikan, dan standar hidup layak (Melliana & Zain, 2013). Sebagai ukuran kualitas hidup, IPM dibangun melalui pendekatan tiga dimensi dasar yaitu umur panjang dan hidup sehat, pengetahuan, dan standar hidup layak (BPS, 2006-2007).

Berdasarkan data hasil publikasi Badan Pusat Statistik (BPS) tahun 2019 menunjukkan bahwa IPM Indonesia pada tahun 2017 sebesar 70,81 persen. Pada tahun 2010 angka IPM Indonesia berkisar 66,53 persen, hal ini menunjukkan bahwa IPM Indonesia meningkat sebesar 4,28 persen. Provinsi Jawa Tengah menjadi salah satu penyumbang IPM terbesar ke 22 dari 34 provinsi di Indonesia, artinya IPM di Jawa Tengah masih tergolong dalam kategori menengah. Hasil publikasi dari BPS Jawa Tengah (2018) menunjukkan bahwa pada tahun 2010 IPM Jawa Tengah berkisar 66,08 persen dan meningkat sebesar 4,44 persen pada tahun 2017, dengan angka IPM pada tahun 2017 sebesar 70,52 persen.

Pembangunan pada suatu wilayah tidak lepas dari pengaruh pembangunan di wilayah sekelilingnya. Hal ini mengindikasikan bahwa

faktor wilayah diduga dapat memberikan pengaruh spasial. Untuk mengatasi hal tersebut dengan cara memasukkan efek spasial antar wilayah ke dalam model, sehingga model yang digunakan adalah model regresi spasial Savita, et.al (2017). Regresi spasial merupakan pengembangan dari regresi linier klasik yang melibatkan unsur wilayah (spasial) kedalam model (Oktaviani, 2018). Ciri dari pemodelan spasial adalah adanya matriks pembobot yang merupakan penanda adanya hubungan antara suatu wilayah dengan wilayah lain (Pratiwi, et.al, 2013). Salah satu pembobot spasial yaitu *Queen Contiguity* (persinggungan sisi dan sudut) yang akan digunakan pada penelitian ini.

Model yang baik tentunya memiliki tingkat akurasi yang lebih baik dan dapat mengurangi kesalahan dalam memprediksi. Teknik *ensemble* dapat digunakan untuk mengurangi keragaman yang terdapat pada model prediksi, serta terbukti dapat memberikan peningkatan akurasi prediksi (Milton, et.al, 2015). Regresi spasial *ensemble* merupakan suatu teknik untuk menggabungkan k model regresi spasial (Rohmawati, 2015). Terdapat dua teknik *ensemble*, yaitu teknik *ensemble hybrid* dan teknik *ensemble non-hybrid* (Canuto, et.al, 2007). Teknik *ensemble hybrid* yaitu teknik yang menggunakan banyak metode dan teknik *ensemble non-hybrid* yaitu teknik yang menggunakan satu jenis metode.

Arifli (2017) telah melakukan penelitian tentang Analisis Faktor-Faktor yang Mempengaruhi Indeks Pembangunan Manusia di Provinsi Sumatera Barat Periode Tahun 2010-2015 menggunakan metode analisis regresi linier berganda dengan metode pendugaan *Ordinary Least Square* (OLS). Ali dan Nugraha (2019) melakukan penelitian Penerapan Metode Regresi Ridge dalam Mengatasi Masalah Multikolinearitas pada Kasus Indeks Pembangunan Manusia di Indonesia Tahun 2017. Kusumaningrum (2018) melakukan penelitian tentang Analisis Faktor-Faktor yang Mempengaruhi Indeks Pembangunan Manusia Provinsi Jawa Tengah Periode Tahun 2006-2016.

Savita (2017) melakukan penelitian tentang Penerapan Model Regresi Ensemble Non-Hybrid pada Data Kemiskinan di Provinsi Jawa

Tengah dengan menggunakan R^2 sebagai pengukuran kebaikan model, dengan hasil mdel regresi spasial *ensemble non-hybrid* terbukti dapat mengurangi keragaman dan meningkatkan prediksi dengan R^2 sebesar 73,59%. Rohmawati (2015) melakukan penelitian tentang Aplikasi Analisis Regresi Spasial Ensemble Pada Data Kemiskinan di Pulau Jawa dengan menggunakan RMSE sebagai pengukuran kebaikan model. Hasil yang diperoleh yaitu model regresi spasial otoregresif dengan teknik *ensemble* lebih baik daripada model regresi spasial otoregresif tanpa *ensemble*, serta dapat dapat meningkatkan kemampuan pendugaan menjadi lebih stabil dan model yang dihasilkan bersifat kekar.

Berdasarkan uraian latar belakang diatas, metode yang akan diterapkan pada kasus Indeks Pembangunan Manusia di Provinsi Jawa Tengah tahun 2018 yaitu menggunakan pendekatan statistik regresi spasial *ensemble non-hybrid*.

2. TINJAUAN PUSTAKA

1. Analisis Regresi Linier Berganda

Analisis regresi merupakan suatu metode statistika yang digunakan untuk mengetahui pengaruh variabel prediktor (X) terhadap variabel respon (Y). Secara umum model regresi dapat dinyatakan dalam bentuk persamaan sebagai berikut (Seber dan Lee, 2003):

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$$

$$i = 1, 2, \dots, n$$

(1)

dimana, Y_i sebagai variabel respon, $\beta_0 + \beta_1 + \dots + \beta_{p-1}$ merupakan parameter regresi yang tidak diketahui variabel prediktor adalah $X_{i1} + X_{i2} + \dots + X_{i,p-1}$, dan ε_i merupakan sisa (*error*) regresi untuk pngamatan ke- i dengan syarat asumsi yang harus terpenuhi yaitu $\varepsilon \sim N(0, \sigma^2 I)$, yang artinya ε_i berdistribusi normal, $E(\varepsilon_i) = 0$, dan varian $\varepsilon_i = \sigma^2$.

2. Analisis Regresi Spasial

Analisis regresi spasial merupakan suatu metode statistik untuk menganalisis pengaruh variabel prediktor terhadap variabel respon dengan memperhatikan aspek keterkaitan wilayah (spasial) (Ernawati, 2016). Secara

umum model spasial dapat dinyatakan dalam bentuk persamaan berikut (LeSage, 1999) dan (Anselin, 1988):

$$\mathbf{y} = \rho \mathbf{W}_1 \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \mathbf{u} \quad (2)$$

dengan,

$$\mathbf{u} = \lambda \mathbf{W}_2 \mathbf{u} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}) \quad (3)$$

dimana, \mathbf{y} adalah vektor variabel respon berukuran $n \times 1$, ρ merupakan koefisien spasial lag variabel respon, \mathbf{W}_1 dan \mathbf{W}_2 merupakan matriks pembobot berukuran $n \times n$, \mathbf{X} adalah matriks variabel prediktor berukuran $n \times (p + 1)$, $\boldsymbol{\beta}$ adalah vektor parameter koefisien regresi berukuran $(p + 1) \times 1$, sedangkan \mathbf{u} merupakan vektor *error* pada persamaan (2) berukuran $n \times 1$. Untuk λ merupakan parameter koefisien spasial lag pada *error*, dan $\boldsymbol{\varepsilon}$ merupakan vektor *error* pada persamaan (3) berukuran $n \times 1$ dengan \mathbf{I} adalah matriks identitas berukuran $n \times n$. n merupakan banyak amatan atau lokasi ($i = 1, 2, \dots, n$) dan p adalah banyak variabel prediktor $k = 1, 2, \dots, p$.

a. Matriks Pembobot

Matriks pembobot spasial dinyatakan sebagai matriks yang menggambarkan kekuatan hubungan antar wilayah (Sulistyaningsih, *et.al*, 2017). Untuk mendapatkan matriks pembobot spasial (\mathbf{W}) dengan menggunakan informasi jarak dari ketetanggaan (*neighborhood*) atau kedekatan antar wilayah penelitian (Ernawati, 2016). Pemberian kode pembobot spasial dibuat *Row Standardization* yaitu berdasarkan pada matriks pembobot jumlah tetangga pada satu baris adalah sama (Lee & Wong, 2011).

$$\mathbf{W}_{ij(std)} = \frac{w_{ij}}{\sum_{j=1}^n w_{ij}} \quad (4)$$

Pada penelitian ini matriks pembobot yang digunakan adalah *Queen Contiguity* (persinggungan sisi dan sudut) dimana $\mathbf{W}_{ij} = 1$ untuk sisi (*entity*) yang bersisian (*common side*) dan titik sudutnya (*common vertex*) bertemu

dengan wilayah yang menjadi perhatian, $W_{ij} = 0$ untuk wilayah lainnya.

b. Uji Autokorelasi Spasial

Autokorelasi spasial merupakan taksiran korelasi antar nilai amatan yang berkaitan dengan wilayah spasial pada variabel yang sama (Kusumawardani, et.al, 2014). Regresi spasial memiliki dua efek spasial yaitu kergaman spasial (*beterogeneity spatial*) dan ketergantungan spasial (*dependence spatial*).

Indeks Moran atau Morans'I merupakan teknik dalam analisis spasial untuk menghitung hubungan spasial yang terjadi dalam suatu ruang (wilayah) (Gittleman dan Kot, 1990). Hipotesis yang digunakan yaitu (Lee dan Wong, 2001):

$$H_0: \hat{I} = I_0 \text{ (tidak ada autokorelasi spasial)}$$

$$H_1: \hat{I} \neq I_0 \text{ (terdapat autokorelasi spasial)}$$

Statistik uji yang digunakan sebagai berikut:

$$Z = \frac{\hat{I} - I_0}{\sqrt{\widehat{var}(\hat{I})}} \tag{5}$$

dengan, Z merupakan nilai statistik uji Morans'I, \hat{I} adalah estimasi indeks Morans'I, $\hat{E}(\hat{I}) = I_0$ adalah nilai ekspektasi Morans'I, dan $\widehat{var}(\hat{I})$ merupakan varians Morans'I.

Uji *Breusch-Pagan Test* (*BP Test*) merupakan salah satu uji yang digunakan untuk melihat adanya keragaman spasial (Anselin, 1988), dengan menggunakan hipotesis sebagai berikut:

$$H_0: \sigma_1^2 = \dots = \sigma_2^2 = \sigma$$

$$H_1: \text{minimal ada satu } \sigma_1^2 \neq \sigma$$

Dengan menggunakan statistik uji:

$$BP = \left(\frac{1}{2}\right) f' Z(Z'Z)^{-1} Z' f \sim \chi_p^2 \tag{6}$$

Pengambilan keputusan, jika $BP > \chi_{\alpha;p}^2$ atau $p - value < \alpha$ maka tolak H_0 .

Uji dependensi spasial atau uji ketergantungan spasial digunakan untuk menunjukkan bahwa pengamatan disuatu wilayah bergantung pada pengamatan diwilayah lain yang berdekatan (Nisa, 2017). Efek ketergantungan spasial merupakan

ketergantungan lag dan sisaan spasial yang dapat diuji dengan menggunakan uji *Lagrange Multiplier* (*LM Test*) (Amelia, 2012).

- **Lagrange Multiplier Lag (LM Lag)**

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

Dengan statistik uji:

$$LM_\rho = \frac{\left[\frac{\varepsilon' W y}{\varepsilon' \varepsilon} \right]^2}{D} \tag{7}$$

Dengan pengambilan keputusan H_0 ditolak jika nilai $LM > \chi_p^2$, dimana p adalah banyaknya parameter spasial. Apabila H_0 ditolak maka model regresi spasial yang sesuai yaitu model *Spatial Autoregressive Model* (SAR).

- **Lagrange Multiplier Error (LM Error)**

$$H_0: \lambda = 0$$

$$H_1: \lambda \neq 0$$

Dengan statistik uji:

$$LM_\lambda = \frac{\left[\frac{\varepsilon' W \varepsilon}{\left(\frac{\varepsilon' \varepsilon}{N}\right)} \right]^2}{tr[W'W + WW]} \tag{8}$$

Dengan pengambilan keputusan H_0 ditolak jika nilai $LM > \chi_p^2$, dimana p adalah banyaknya parameter spasial. Apabila H_0 ditolak maka model regresi spasial yang sesuai yaitu model *Spatial Error Model* (SEM).

c. Spatial Autoregressive Model (SAR)

Model SAR adalah model regresi linier yang memiliki korelasi spasial pada varibel responnya (Anselin, 1999). model SAR dengan persamaan:

$$y = \rho W y + X \beta + \varepsilon \tag{9}$$

$$\varepsilon \sim N(0, \sigma^2 I)$$

ε_i diasumsikan menyebar normal, bebas stokastik, identik, dengan nilai tengah nol dan ragam σ^2 , ε_i merupakan *error* pada lokasi ke- i (Arisanti, 2011).

d. Spatial Error Model (SEM)

Model SEM merupakan model regresi linier yang memiliki korelasi spasial pada *error* (Amelia, 2012). model SEM dengan persamaan:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \quad (10)$$

$$\mathbf{u} = \lambda \mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon} \quad (11)$$

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$$

3. Model Regresi Spasial *Ensemble*

Ensemble merupakan suatu teknik dalam memprediksi dengan menggabungkan beberapa model yang dihasilkan dari suatu metode atau beberapa metode (R, 2015). Teknik *ensemble* adalah menggabungkan hasil prediksi dari banyak model menjadi satu buah prediksi akhir, bukan memilih satu model terbaik dari sekian banyak model dan kemudian melakukan prediksi dari model terbaik yang terpilih tersebut (Fransiska, 2014).

$$\hat{\mathbf{Y}} = \frac{1}{Q} \sum_{p=1}^Q \hat{\mathbf{Y}}_p \quad (12)$$

dimana $\hat{\mathbf{Y}}$ merupakan rata-rata estimasi parameter, Q adalah banyaknya resampling, $\hat{\mathbf{Y}}_p$ merupakan model regresi spasial ke- p , dan p adalah banyaknya model regresi spasial.

Teknik *hybrid ensemble* menggunakan beberapa model dengan menggabungkan pendugaan parameter yang dihasilkan oleh masing-masing model menjadi hasil akhir. Sedangkan teknik *non-hybrid ensemble* menggunakan satu model dan digunakan berulang kali untuk mendapatkan model yang berbeda dan hasil pendugaan parameter dari masing-masing model tersebut digabungkan menjadi satu. Batasan pada penelitian ini menggunakan teknik *non-hybrid ensemble*, dimana model regresi spasial yang didapatkan dengan cara menambahkan *noise* pada variabel respon. Menurut Wu dan Huang, (2009) *Noise* merupakan gangguan yang tidak beraturan pada data. *Additive noise* didapatkan dari bangkitan berdistribusi normal $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma)$ yang dibangkitkan sebanyak q ulangan dan menghasilkan set data resampling sebanyak q (Sulistyaningsih, *et.al*, 2017). Persamaan *additive noise* dapat dituliskan sebagai berikut:

$$\mathbf{m} = \mathbf{y} + \boldsymbol{\varepsilon}_z$$

(13)

dimana \mathbf{m} merupakan vektor variabel respon setelah ditambahkan *noise*, \mathbf{y} adalah vektor variabel respon sebelum ditambahkan *noise*, dan $\boldsymbol{\varepsilon}_z \sim N(\mathbf{0}, \sigma)$. Untuk σ merupakan simpangan baku dengan nilai kecil dari nilai rentang data.

4. Pemilihan Model Terbaik

Untuk mendapatkan model yang baik diperlukan sebuah metode dalam mengukur kebaikan model. Pemilihan model terbaik menggunakan *Akaike's Information Criterion* (AIC) dan *Root Mean Square Error* (RMSE).

Berikut adalah rumus yang digunakan dalam menghitung dilai AIC (Fathurahman, 2009):

$$AIC = e \frac{2k}{n} \frac{\sum_{i=1}^n \hat{u}_i^2}{n} \quad (14)$$

dimana,

k merupakan jumlah parameter yang diestimasi dalam model regresi, n adalah jumlah observasi, $e = 2,718$, dan u adalah *error*.

RMSE merupakan akar dari rata-rata jumlah kuadrat *error* antara selisih nilai prediksi dengan nilai aktual. Berikut adalah rumus dari RMSE:

$$RMSE = \sqrt{\frac{\sum (Y_t - \hat{Y}_t)^2}{n}} \quad (15)$$

Semakin kecil nilai AIC dan nilai RMSE yang dihasilkan maka semakin bagus model yang didapatkan.

5. Indeks Pembangunan Manusia

Pada tahun 1990, UNDP memperkenalkan suatu indikator yang telah dikembangkannya yaitu indikator yang dapat menggambarkan perkembangan pembangunan manusia secara terukur dan representatif, yang dinamakan *Human Development Index* (HDI) atau Indeks Pembangunan Manusia (IPM) (Pratowo, 2013). Indeks ini pertama kali dikembangkan oleh seorang ekonom Pakistan pemenang Nobel India Amartya Sen dan Mahbub Ul Haq dibantu oleh Gustav Ranis dari Yale University dan Lord Meghnad Desai (Arifli, 2017). Capaian pembangunan manusia di suatu

wilayah pada waktu tertentu dapat dikelompokkan ke dalam empat kelompok (BPS, 2015) yaitu: Sangat tinggi $IPM \geq 80$, Tinggi $70 \leq IPM < 80$, Sedang $60 \leq IPM < 70$, Rendah $IPM < 60$.

Sudah banyak peneliti yang melakukan penelitian terkait faktor yang mempengaruhi IPM. Sehingga pada penelitian ini akan digunakan beberapa variabel yang berpengaruh terhadap IPM yang dirujuk pada beberapa penelitian terdahulu yaitu: kepadatan penduduk, kemiskinan, produk domestik regional bruto, tingkat partisipasi Angkatan kerja, angka partisipasi sekolah, angka putus sekolah, dan rata-rata pengeluaran per kapita per bulan makanan dan non makanan.

3. METODE

1. Sumber Penelitian

Data yang digunakan pada penelitian ini merupakan data sekunder yang diambil dari hasil publikasi Badan Pusat Statistik Provinsi Jawa Tengah Tahun 2017.

2. Variabel Penelitian

Variabel yang digunakan pada penelitian ini adalah variabel respon (Y) dan variabel prediktor (X). Variabel respon yaitu indeks pembangunan manusia, sedangkan variabel prediktor yaitu variabel-variabel yang diduga mempengaruhi indeks pembangunan manusia yang terdiri dari tujuh variabel yaitu: kepadatan penduduk, kemiskinan, produk domestik regional bruto, tingkat partisipasi Angkatan kerja, angka partisipasi sekolah, angka putus sekolah, dan rata-rata pengeluaran per kapita per bulan makanan dan non makanan.

3. Langkah-langkah penelitian

Adapun langkah-langkah penelitian ini adalah:

1. Melakukan eksplorasi data
2. Melakukan analisis regresi linier berganda menggunakan *Ordinary Least Square* (OLS) untuk mengetahui pelanggaran asumsi pada residual.
3. Menentukan dan menghitung matriks pembobot antar wilayah dengan menggunakan pembobot spasial *Queen*.
4. Melakukan uji Indeks Moran untuk identifikasi awal adanya efek spasial.
5. Melakukan uji *Lagrange Multiplier* (LM) untuk mengetahui ketergantungan spasial.
6. Melakukan analisis regresi spasial

7. Melakukan prediksi model spasial *ensemble non-hybrid* dengan langkah:

- Membangkitkan *noise* dengan $\epsilon \sim N(0, \sigma)$
- Menambahkan *noise* pada data IPM (Y)
- Melakukan pengujian pada data yang telah ditambahkan *noise* dengan model regresi spasial yang didapatkan
- Mengulangi langkah a sampai c sebanyak k kali dengan *noise* yang berbeda pada setiap iterasi

8. Membuat model *ensemble non-hybrid* dari k model regresi spasial dengan menghitung rata-rata koefisien dari model tersebut

9. Mengukur kebaikan model dan interpretasi model

4. HASIL DAN PEMBAHASAN

1. Eksplorasi data

Eksplorasi data yang dilakukan pada penelitian ini untuk mengetahui karakteristik dari data secara umum dengan menggunakan peta tematik, dimana semakin pekatnya warna pada peta maka semakin tinggi nilai IPM pada wilayah tersebut.



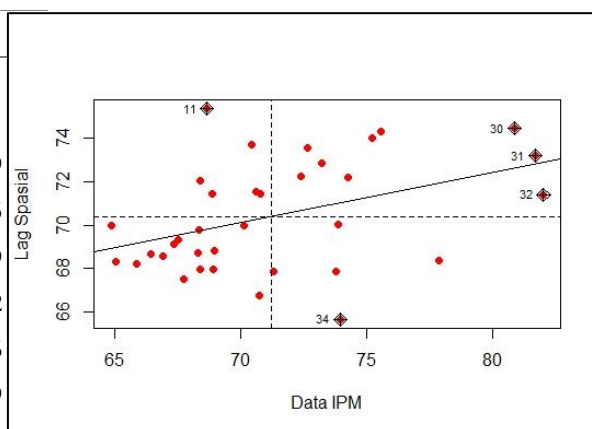
Gambar 1 Penyebaran Indeks Pembangunan Manusia (IPM) Jawa Tengah

Gambar 1 menunjukkan penyebaran IPM di Provinsi Jawa Tengah tahun 2017. Adanya pengaruh efek spasial atau keterkaitan antar wilayah dapat dilihat dari antar kabupaten/kota yang berdekatan memiliki angka IPM yang tidak jauh berbeda.

2. Analisis Regresi Linier Berganda dengan OLS

Tabel 1 Estimasi Parameter dengan OLS

Parameter	Estimasi	t _{hitung}	p-value	VIF
β_0	17,92	1,316	0,199	
β_1	0,0004	2,281	0,031*	2,201
β_2	-0,2256	-2,430	0,022*	1,739
β_3	0,0000	1,088	0,286	1,405
β_4	0,1538	1,423	0,166	1,199
β_5	0,146	4,335	0,000*	1,732
β_6	-0,0165	-0,840	0,408	1,816
β_7	0,0000	3,092	0,004*	2,899
F _{hitung}	30,05			



Gambar 2 Morans'I Scatterplot IPM

Ket: *) signifikan pada $\alpha = 5\%$ $F_{\text{tabel}} = 2,40$

Nilai F_{hitung} pada Tabel 1 lebih besar dari F_{tabel} maka H_0 ditolak, artinya bahwa variabel prediktor secara serentak berpengaruh terhadap variabel respon. Model regresi dengan OLS yang terbentuk adalah:

$$\hat{y}_i = 17,92 + 0,000401X_{i1} - 0,2256X_{i2} + 0,146X_{i5} + 0,0000X_{i7}$$

Setelah mendapatkan model regresi selanjutnya melakukan uji asumsi residual.

3. Pemodelan Regresi Spasial

Sebelum melakukan pemodelan dilakukan identifikasi pola hubungan menggunakan uji autokorelasi spasial dengan Morans'I test. Berdasarkan hasil analisis uji Morans'I menunjukkan bahwa nilai Morans'I 0,2307 dengan nilai p-value $0,0115 < 0,05$ maka H_0 ditolak, artinya terdapat autokorelasi spasial. Nilai I sebesar $0,2307 > I_0 -0,0294$ menunjukkan bahwa data berautokorelasi positif, artinya lokasi yang berdekatan cenderung berkelompok dan mempunyai nilai yang hampir mirip.

Gambar 2 menunjukkan variabel IPM terdapat pengelompokkan pada kuadran III (*low-low*), artinya kabupaten/kota dengan angka IPM yang rendah cenderung berdekatan dengan kabupaten/kota yang memiliki angka IPM yang rendah pula. Selanjutnya pada kuadran I (*high-high*) menunjukkan bahwa kabupaten/kota yang memiliki angka IPM tinggi cenderung berdekatan dengan kabupaten/kota yang memiliki angka IPM tinggi pula.

Sebelum melakukan pemodelan spasial, terlebih dahulu melakukan uji identifikasi model spasial dengan menggunakan uji *Lagrange Multiplier* (LM test).

Tabel 2 Uji Lagrange Multiplier

Uji dependensi spasial	p-value
<i>Lagrange Multiplier</i> (error)	0.754
<i>Lagrange Multiplier</i> (lag)	0.036*
SARMA	0.070

Ket: *) signifikan pada $\alpha = 5\%$

Tabel 2 menunjukkan bahwa nilai p-value pada *Lagrange Multiplier* (lag) sebesar $0,036 < 0,05$ sehingga H_0 ditolak. Artinya terdapat dependensi spasial pada lag, sehingga perlu dilanjutkan melakukan pemodelan dengan menggunakan metode *Spatial Autoregressive Model* (SAR).

Berdasarkan hasil uji dependensi spasial maka dilanjutkan dengan melakukan

pemodelan regresi spasial menggunakan metode SAR.

Tabel 3 Estimasi Parameter SAR

Parameter	Estimasi	Z _{hitung}	p-value
β_0	12,451	1,032	0,302
β_1	0,0004	2,866	0,004*
β_2	-0,1934	-2484	0,012*
β_3	0,0000	1,281	0,200
β_4	0,0931	1,022	0,306
β_5	0,119	4,129	0,000*
β_6	-0,0196	-1,210	0,226
β_7	0,0000	3,887	0,000*
ρ	0,234		0,040

Ket: *) signifikan pada $\alpha = 5\%$

Tabel 3 menunjukkan bahwa nilai ρ (rho) sebesar 0,234 dengan p-value 0,040 yang signifikan pada 0,05 maka H_0 ditolak yang artinya terdapat pengaruh spasial pada lag. Berdasarkan hasil estimasi parameter SAR pada Tabel 3 maka diperoleh model:

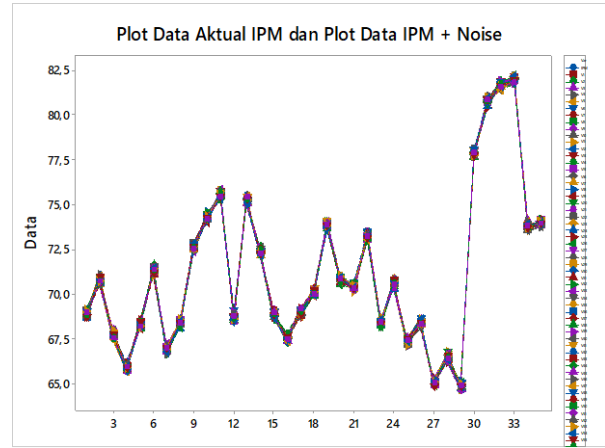
$$\hat{y}_i = 0,234 \sum_{j=1}^{35} W_{ij}y_j + 12,451 + 0,0004X_{i1} - 0,1934X_{i2} + 0,119X_{i5} + 0,0000X_{i7}$$

$\sum_{j=1}^n W_{ij}y_j$ menunjukkan adanya pengaruh spasial terhadap angka IPM antara kabupaten/kota yang diamati (i) dengan kabupaten/kota yang letaknya berdekatan (j).

Selanjutnya uji asumsi IIDN dengan hasil bahwa semua uji memenuhi.

4. Pemodelan Regresi Spasial Ensemble

Model spasial ensemble yang digunakan adalah non-hybrid, karena pengaruh spasial hanya pada lag sehingga model spasial yang terpilih yaitu Spatial Autoregressive Model (SAR). Penambahan noise yang berbeda sebanyak 100 kali pada variabel Y dengan menggunakan nilai standar deviasi yang dicobakan 0,11 berdasarkan penelitian dari (R, 2015).



Gambar 3 Plot Data IPM Aktual dan Plot Data IPM + Noise

Model prediksi SAR dengan teknik ensemble yang dihasilkan sebanyak 100 model sebagai berikut:

$$\begin{aligned} \hat{y}_1 &= 0,2504 \sum_{j=1}^{35} W_{ij}y_j + 11,7781 + 0,0003X_{i1} \\ &\quad - 0,2219X_{i2} + 0,1275 X_{i5} + 0,0000X_{i7} \\ \hat{y}_2 &= 0,2426 \sum_{j=1}^{35} W_{ij}y_j + 12,1676 + 0,0003X_{i1} \\ &\quad - 0,2352X_{i2} + 0,1273X_{i5} + 0,0000X_{i7} \\ \hat{y}_3 &= 0,2777 \sum_{j=1}^{35} W_{ij}y_j + 13,1828 + 0,0003X_{i1} \\ &\quad - 0,2442X_{i2} + 0,1262X_{i5} + 0,0000X_{i7} \\ &\quad \vdots \\ \hat{y}_{100} &= 0,2633 \sum_{j=1}^{35} W_{ij}y_j + 12,5751 \\ &\quad + 0,0003X_{i1} - 0,2347X_{i2} + 0,1269X_{i5} \\ &\quad + 0,0000X_{i7} \end{aligned}$$

Model regresi spasial lag ensemble didapatkan dengan merata-ratakan koefisien parameter 100 model, sehingga dihasilkan satu model tunggal regresi spasial sebagai berikut:

$$\hat{y} = 0,256 \sum_{j=1}^{35} W_{ij}y_j + 13,0 + 0,0003X_{i1} - 0,232 X_{i2} + 0,128X_{i5} + 0,0000X_{i7}$$

Nilai penduga ρ sebesar 0,256 yang artinya kabupaten/kota yang memiliki angka IPM yang tinggi diduga dipengaruhi oleh kabupaten/kota yang ada disekelilingnya sebesar 0,256. Interpretasi dari model SAR *ensemble* dengan asumsi variabel lain dianggap konstan yang menunjukkan bahwa setiap kenaikan satu persen kepadatan penduduk akan meningkatkan angka IPM sebesar 0,0003 kali. Selanjutnya, setiap kenaikan satu persen angka kemiskinan akan menurunkan angka IPM sebesar -0,232 kali. Kemudian setiap kenaikan satu persen angka partisipasi sekolah akan meningkatkan nilai IPM sebesar 0.128 kali, dan setiap kenaikan satu persen rata-rata pengeluaran per kapita per bulan makanan dan non makanan maka akan meningkatkan angka IPM sebesar 0,0000 kali.

Sedangkan $\sum_{j=1}^n W_{ij}y_j$ menunjukkan adanya pengaruh spasial terhadap angka IPM antara kabupaten/kota yang diamati (i) dengan kabupaten/kota yang letaknya berdekatan (j).

5. Pemilihan Model Terbaik

Pemilihan model terbaik dengan melihat nilai AIC dan nilai RMSE, semakin kecil nilai AIC dan RMSE maka semakin bagus model yang didapatkan.

Tabel 4 Pemilihan Model Terbaik

Model	AIC	RMSE	R ²
OLS	145,109	1,4894	0,8568
SAR	143	1,3932	0,9004
SAR <i>Ensemble</i>	143	1,3899	0,9009

Hasil dari nilai AIC dan RMSE menunjukkan model spasial lag dengan Teknik *ensemble* dapat meningkatkan kemampuan pendugaan menjadi lebih stabil dan dan model yang dihasilkan bersifat kuat. Berdasarkan nilai R² model SAR *ensemble* memiliki nilai R² = 0,9009 lebih besar dari R² model OLS, yang artinya bahwa sebesar 90,09% angka

7. REFERENSI

- Ali, R., & Nugraha, J. (2019). Penerapan Metode Regresi Ridge dalam Mengatasi Masalah Multikolinearitas Pada Kasus Indeks Pembangunan Manusia di Indonesia Tahun 2017. *Prosiding Sendika*.
- Amelia, M. (2012). *Penerapan Regresi Spasial Untuk Data Kemiskinan Kabupaten di Pulau Jawa*. Bogor: IPB.

IPM dapat dijelaskan oleh variabel prediktor dan sisanya sebesar 9,91% dapat dijelaskan oleh variabel lain diluar model.

5. KESIMPULAN

1. Penyebaran angka IPM di Jawa Tengah tahun 2017 pada analisis deskriptif menunjukkan terjadinya pengelompokan pada kabupaten/kota yang memiliki wilayah berdekatan. Terdapat empat wilayah kabupaten/kota yang memiliki angka IPM dalam kategori sangat tinggi yaitu Kota Surakarta, Kota Salatiga, Kota Magelang, dan Kota Semarang. berdasarkan hasil identifikasi awal dengan menggunakan uji Morans'I menunjukkan bahwa data berautokorelasi positif, artinya lokasi yang berdekatan cenderung berkelompok dan mempunyai nilai yang hampir mirip.
2. Dari hasil pemodelan IPM dan variabel-variabel yang mempengaruhinya didapatkan kesimpulan, dengan menggunakan teknik *ensemble* pada regresi spasial memberikan hasil yang lebih baik. Hal ini dapat dilihat dari nilai AIC sebesar 143 dan nilai RMSE sebesar 1,4584. Adapun variabel prediktor yang berpengaruh terhadap angka IPM di Jawa Tengah tahun 2017 yaitu kepadatan penduduk (X₁), kemiskinan (X₂), angka partisipasi sekolah (X₅), dan rata-rata pengeluaran per kapita per bulan makanan dan non makanan (X₇). Dengan model SAR *ensemble* yang didapatkan yaitu:

$$\hat{y} = 0,256 \sum_{j=1}^{35} W_{ij}y_j + 13,0 + 0,0003X_{i1} - 0,232 X_{i2} + 0,128X_{i5} + 0,0000X_{i7}$$

6. UCAPAN TERIMA KASIH

Kami Ucapkan terimakasih kepada BPS Jawa Tengah atas informasi dan data. Laboratorium Prodi Statistika UNIMUS atas dukungannya dalam penelitian ini.

- Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Netherlands: Kluwer Academic Publishers.
- Anselin, L. (1999). *Spatial Econometrics*. Dallas: School of Social Sciences.
- Arifli, A. F. (2017). *Analisis Fktor-Faktor yang Mempengaruhi Indeks Pembangunan Manusia di provinsi Sumatera Barat Periode Tahun 2010-2015*. Yogyakarta: UIN Sunan Kalijaga.
- Arisanti, R. (2011). *Model Regresi Spasial Untuk Deteksi Faktor-Faktor Kemiskinan di Provinsi Jawa Timur*. Bogor: IPB.
- Badan Pusat Statistik. (2006-2007). *Indeks Pembangunan Manusia*. Jakarta, Indonesia.
- Badan Pusat Statistik. (2010). *Indeks Pembangunan Manusia*. Jakarta.
- Badan Pusat Statistik. (2015). *Indeks Pembangunan Manusia 2014: Metode Baru*. BPS.
- Badan Pusat Statistik. (2019). Retrieved from <https://www.bps.go.id/dynamictable/2016/06/16/1211/indeks-pembangunan-manusia-menurut-provinsi-2010-2018-metode-baru.html>
- Canuto, A. M., Abreu, M. C., Oliviera, L., Jr, J. C., & Santos, A. (2007). Investigating the Influence of The Choice of the Ensemble Members in Accuracy and Diversity of Selection-Based and Fusion-Based Methods for Ensembles. *Pattern Recognition Letters*, 472–486.
- Ernawati. (2016). *Analisis Faktor-Faktor yang Memengaruhi Angka Prevalensi Penyakit Kusta di Jawa Timur dengan Pendekatan Spatial Durbin Model*. Surabaya: ITS.
- Fathurahman, M. (2009). Pemilihan Model Regresi Terbaik Menggunakan Metode Akaike's Information Criterion dan Schwarz Information Criterion. *Jurnal Informatika Mulawarman*, 37-41.
- Gittleman, J. L., & Kot, M. (1990). Adaptation: Statistics and A null Model for Estimating Phylogenetic Effects. *Systematic Zoology*, 227 – 241.
- Kusumaningrum, R. A. (2018). *Analisis Faktor-Faktor yang Mempengaruhi Indeks Pembangunan Manusia Provinsi Jawa Tengah Periode Tahun 2006-2016*. Yogyakarta: UII.
- Kusumawardani, D. S., Rahmawati, D., Wuryandari, T., & Hoyyi, A. (2014). Identifikasi Autokorelasi Spasial Pada Jumlah Pengangguran di Jawa Tengah Menggunakan Indeks Moran. *Media Statistika*, 1-10.
- Lee, J., & Wong, D. W. (2011). *Statistical Analysis of Geographic Information with ArcView GIS and ArcGIS*. New York: John Wiley and Sons.
- Lee, J., & Wong, D. W. (2001). *Statistical Analysis with ArcView GIS*. Canada: John Willey & Sons, Inc.
- LeSage, J. P. (1999). *The Theory and Practice of Spatial Econometrics*. Toledo: Departement of Economics University of Toledo.
- Melliana, A., & Zain, I. (2013). Analisis Statistika Faktor yang Mempengaruhi Indeks Pembangunan Manusia di Kabupaten/Kota Provinsi Jawa Timur dengan Menggunakan Regresi Panel. *JURNAL SAINS DAN SENI POMITS*, 2337-3520.
- Milton, J. N., Steinberg, M. H., & Sebastiani, P. (2015). Evaluation of an Ensemble of Genetic Models for Prediction of a Quantitative Trait. *Methods Article*, 5:1-6.
- Nisa, I. I. (2017). *Pemodelan Spatial Durbin Error Model (SDEM) pada Data Indeks Pembangunan Manusia (IPM)*. Semarang: UNIMUS.
- Oktaviani, N. (2018). *Analisis Regresi Spasial Pada Data Jumlah Penduduk Miskin Provinsi Lampung Tahun 2017*. Bandar Lampung: Universitas Lampung.

- Pratiwi, L. P., Srinandi, I. A., & Susilawati, M. (2013). Analisis Kemiskinan dengan Pendekatan Model Regresi Spasial Durbin (Studi Kasus: Kabupaten Gianyar). *E-Jurnal Matematika*, 11-16.
- Pratowo, N. I. (2013). Analisis Faktor-Faktor yang Berpengaruh Terhadap Indeks Pembangunan Manusia. *Jurnal Studi Ekonomi Indonesia*, 15-31.
- R, S. M. (2015). *Ensemble Hybrid Terboboti Pada Model Autokorelasi Spasial Untuk Memprediksi Indeks Pembangunan Manusia di Pulau Jawa*. Bogor: IPB.
- Rohmawati, N. (2015). *Aplikasi Analisis Regresi Spasial Ensemble pada Data Kemiskinan di Pulau Jawa*. Bogor: IPB.
- Savita, C. A., Handajani, S. S., & Winarno, B. (2017). Penerapan Model Regresi Ensemble Non-Hybrid pada Data Kemiskinan di Provinsi Jawa Tengah. *The 6th University Research Colloquium 2017*.
- Seber, G. A., & Lee, A. J. (2003). *Linier Regression Analysis* (2nd ed.). Hoboken: John Wiley & Sons Inc.
- Sulistyaningsih, Saputro, D. R., & Widyaningsih, P. (2017). *Teknik Ensemble dengan Additive Noise pada Estimasi Parameter Model Autoregressive Spasial*. Surakarta: Universitas Sebelas Maret.
- UNDP. (1990). *Human Development Report*. New York: Oxford University Press.
- Wu, Z., & Huang, N. E. (2009). Ensemble Empirical Mode Decomposition: a Noise-Assisted Data Analysis Method. *Advances in Adaptive Data Analysis*, 1-41.