

# PENGUNAAN ANALISIS KLASSTER K-MEANS DALAM PEMODELAN REGRESI SPASIAL PADA KASUS TUBERKULOSIS DI JAWA TIMUR TAHUN 2017\*

Hardani Prisma Rizky<sup>1</sup>, Wara Pramesti<sup>2</sup>, and Gangga Anuraga<sup>3</sup>

<sup>1</sup>Program Studi Statistika, FMIPA, Universitas PGRI Adi Buana Surabaya, Indonesia, hardaniprisky@gmail.com

<sup>2</sup>Program Studi Statistika, FMIPA, Universitas PGRI Adi Buana Surabaya, Indonesia, warapra@unipasby.ac.id

<sup>3</sup>Program Studi Statistika, FMIPA, Universitas PGRI Adi Buana Surabaya, Indonesia, g.anuraga@unipasby.ac.id

**Indonesian Journal of Statistics and Its Applications (eISSN:2599-0802)**

**Vol 4 No 1 (2020), 164 - 178**

Copyright © 2020 Hardani Prisma Rizky, Wara Pramesti, and Gangga Anuraga. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## **Abstract**

Tuberculosis (TB) is a contagious infectious disease caused by the bacterium *Mycobacterium tuberculosis* which can attack various organs, especially the lungs. TB if left untreated or incomplete treatment can cause dangerous complications to death. East Java Province has the second-highest TB case after West Java Province. Therefore, we need statistical modeling to analyze the factors that influence TB in East Java Province. The data used in this study were sourced from data from BPS and East Java Provincial Health Offices in 38 districts/cities in East Java Province in 2017. Analysis of data using the OLS regression approach only looked at variable factors but was unable to know the effects of territory. So to overcome this, a spatial regression approach is used by comparing the weight of Queen Contiguity and the results of the k-means cluster analysis to obtain the best model. Based on the results of the analysis, the spatial aspects of the data have met the assumptions of spatial dependencies using the Moran's I test with a p-value of 0.000001295. The weighting matrix used is the k-means cluster weighting matrix  $k = 2$ . The test results obtained by the Spatial Autoregressive Moving Average (SARMA) model selected as the best model with the value of the deterrence coefficient (R<sup>2</sup>) and Akaike Info Criterion (AIC), 87.10% and 586.69. The factors that significantly influence the number of Tuberculosis patients in each district/city in East Java are population density (X<sub>2</sub>) and the number of healthy houses (X<sub>9</sub>).

**Keywords:** moran's i, multiple linear regression, sarma, tuberculosis.

---

\* Received Aug 2019; Accepted Feb 2020; Published online on Feb 2020

## 1. Pendahuluan

Tuberkulosis (TB) adalah penyakit infeksi menular yang disebabkan bakteri *Mycobacterium tuberculosis* yang dapat menyerang berbagai organ, terutama paru-paru. Penyakit TB apabila tidak diobati atau pengobatannya tidak tuntas dapat menimbulkan komplikasi berbahaya hingga kematian. Menurut Dinas Kesehatan Provinsi Jawa Timur 2017, Jawa Timur menempati urutan kedua tertinggi di Indonesia dalam jumlah penemuan penderita TB. Jumlah penemuan kasus baru sebanyak 26.152 kasus dan jumlah penemuan semua kasus TB sebanyak 54.811 kasus. Menurut Suherni & Maduratna (2013), penelitian yang telah dilakukan menyatakan bahwa faktor-faktor yang berpengaruh terhadap terjadinya penyakit TB antara lain yaitu faktor lingkungan, karakteristik individu, lingkungan sosial, dan kondisi sosial ekonomi. Diantaranya yaitu kepadatan penduduk, jumlah masyarakat miskin, jumlah balita gizi buruk, jumlah bayi mendapat cakupan imunisasi BCG, angka harapan hidup, persentase rumah tangga berperilaku hidup sehat dan lain sebagainya.

Regresi spasial merupakan hasil pengembangan dari metode regresi linear klasik (LeSage & Pace, 2009). Pengembangan itu berdasarkan adanya pengaruh tempat atau spasial pada data yang dianalisis. Salah satu ciri khas dalam model regresi spasial yaitu adanya dependensi (ketergantungan) antar lokasi yang menyebabkan pendugaan model bisa menjadi lebih kompleks. Menurut Anselin (2005; 2013), regresi spasial memiliki beberapa model utama yaitu *Spatial Autoregressive Model* (SAR), *Spatial Error Model* (SEM) dan *Spatial Autoregressive Moving Average* (SARMA).

Komponen yang mendasar pada regresi spasial adalah matriks pembobot spasial (*spatial weighted matrices*). Beberapa pendekatan yang dapat dilakukan untuk menampilkan hubungan spasial antar lokasi, diantaranya adalah konsep persinggungan (*contiguity*). Jenis persinggungan ada 3, yaitu *Rook Contiguity*, *Bishop Contiguity* dan *Queen Contiguity*. Salah satu permasalahan yang biasanya mengandung efek spasial diantaranya adalah masalah kedokteran seperti penyebaran suatu penyakit. Dalam penelitian ini ingin memodelkan persebaran penderita penyakit Tuberkulosis di Jawa Timur menggunakan matriks pembobot dari hasil *cluster* atau pengelompokan wilayah berdasarkan faktor-faktor yang diduga mempengaruhi jumlah penderita Tuberkulosis di Jawa Timur.

Penelitian yang telah dilakukan terkait dengan Tuberkulosis (TB) maupun metode Analisis Spasial diantaranya penelitian tentang Analisis Pengelompokan Kecamatan di Kota Surabaya Berdasarkan Faktor Penyebab Terjadinya Penyakit Tuberkulosis yang dilakukan Suherni & Maduratna (2013). Hasil penelitian menunjukkan bahwa berdasarkan faktor-faktor penyebab terjadinya penyakit tuberkulosis, kecamatan Tandes, Gubeng, Rungkut, Sukolilo, Wonokromo, Wonocolo, Genteng, Tambak Sari, Sawahan, Kenjeran, Semampir, Pabean Cantikan, dan Krembangan, merupakan kecamatan-kecamatan dengan daerah kerawanan penyebaran penyakit tuberkulosis yang tinggi. Adapula penelitian yang dilakukan oleh Angisna (2018) yaitu Aplikasi Regresi Spasial untuk Mengetahui Pengaruh Faktor-Faktor Persebaran TB Paru di Kabupaten Magetan. Hasil dari penelitian diperoleh faktor spesifik yang mempengaruhi persebaran TB Paru yaitu kepadatan hunian dan fasilitas kesehatan

dengan menggunakan *Spatial Error Model* (SEM). Berdasarkan uraian di atas tujuan peneliti ini untuk mengetahui faktor-faktor yang mempengaruhi kasus tuberkulosis di Jawa Timur Tahun 2017 dengan menggunakan regresi spasial pembobot klaster k-means.

## 2. Metodologi

### 2.1 Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder yang diambil dari Publikasi Badan Pusat Statistik (BPS) Jawa Timur (BPS, 2018) dan Dinas Kesehatan Jawa Timur tahun 2017 (DINKES, 2018). Data yang diambil pada 38 kabupaten/kota yang ada di Provinsi Jawa Timur. Variabel yang digunakan yaitu variabel jumlah gizi buruk ( $X_1$ ), kepadatan penduduk ( $X_2$ ), persentase imunisasi dasar lengkap ( $X_3$ ), persentase imunisasi BCG ( $X_4$ ), jumlah bayi berat lahir rendah ( $X_5$ ), persentase tempat umum memenuhi syarat kesehatan ( $X_6$ ), persentase rumah tangga berperilaku hidup bersih dan sehat ( $X_7$ ), persentase kemiskinan ( $X_8$ ), jumlah rumah sehat ( $X_9$ ), dan jumlah penderita HIV ( $X_{10}$ ).

### 2.2 Metode Penelitian

Metode analisis data yang digunakan dalam penelitian ini adalah analisis spasial dengan menggunakan pembobot dari hasil analisis klaster k-means. Langkah-langkah analisis yang dilakukan dalam penelitian ini adalah sebagai berikut.

1. Mendeskripsikan persebaran penderita penyakit Tuberkulosis (TB) di Jawa Timur tahun 2017.
2. Melakukan analisis klaster k-means berdasarkan faktor-faktor yang diduga mempengaruhi jumlah penderita Tuberkulosis di Jawa Timur.
3. Memodelkan TB dengan regresi linear berganda.
4. Menguji asumsi residual model regresi linear berganda (asumsi residual identik, independen dan berdistribusi normal).
5. Menyusun pembobot spasial *queen contiguity* dan pembobot dari hasil analisis klaster.
6. Menguji dependensi spasial dengan uji *Moran's I*.
7. Memilih model yang paling signifikan dari hasil pembobot *queen contiguity* dan dari hasil analisis klaster k-means untuk dilakukan pemilihan model spasial selanjutnya.
8. Melakukan uji *Lagrange Multiplier* (LM) untuk menentukan model yang sesuai.
9. Memodelkan TB berdasarkan hasil pemilihan model yang sesuai dari uji LM.
10. Menguji asumsi residual pada model regresi spasial yang terbentuk (asumsi residual identik, independen dan berdistribusi normal).
11. Pemilihan model terbaik dengan kriteria  $R^2$  dan AIC.
12. Interpretasi model.

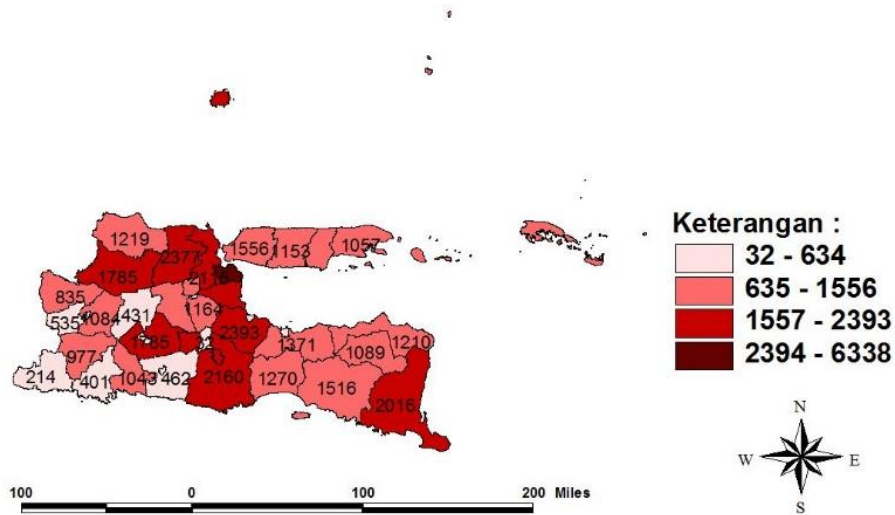
### 3. Hasil dan Pembahasan

#### 3.1 Karakteristik Persebaran Penyakit Tuberkulosis di Jawa Timur Tahun 2017

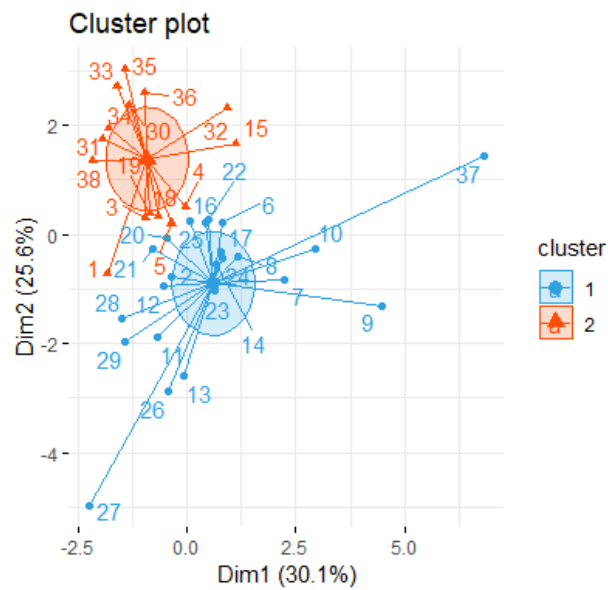
Persebaran jumlah kasus tuberkulosis di Provinsi Jawa Timur dalam bentuk peta disajikan pada Gambar 1. Berdasarkan Gambar 1 dapat diketahui bahwa terdapat empat kategori yang dibedakan dengan warna merah, semakin pekat warna merah maka semakin banyak jumlah kasus tuberkulosis. Warna merah paling terang menunjukkan jumlah penderita tuberkulosis pada wilayah tersebut dikatakan rendah yaitu antara 32-634 kasus. Terdapat 12 daerah yang termasuk dalam kategori rendah yaitu Kabupaten Pacitan, Kabupaten Trenggalek, Kabupaten Blitar, Kabupaten Nganjuk, Kabupaten Magetan, Kota Madiun, Kota Mojokerto, Kota Pasuruan, Kota Probolinggo, Kota Batu, Kota Kediri dan Kota Blitar. Warna merah muda menunjukkan jumlah penderita tuberkulosis pada wilayah tersebut dikatakan cukup rendah yaitu antara 635-1556 kasus. Terdapat 16 daerah yang termasuk dalam kategori cukup rendah yaitu Kabupaten Ngawi, Kabupaten Madiun, Kabupaten Ponorogo, Kabupaten Tulungagung, Kabupaten Tuban, Kabupaten Jombang, Kabupaten Mojokerto, Kabupaten Lumajang, Kabupaten Jember, Kabupaten Bondowoso, Kabupaten Probolinggo, Kabupaten Situbondo, Kabupaten Bangkalan, Kabupaten Sampang, Kabupaten Pamekasan dan Kabupaten Sumenep. Warna merah menunjukkan jumlah penderita tuberkulosis di wilayah tersebut dikatakan cukup tinggi yaitu 1557-2393 kasus. Terdapat 9 daerah yang termasuk dalam kategori cukup tinggi yaitu Kabupaten Bojonegoro, Kabupaten Lamongan, Kabupaten Gresik, Kabupaten Sidoarjo, Kabupaten Pasuruan, Kabupaten Malang, Kabupaten Kediri dan Kabupaten Banyuwangi. Sedangkan untuk warna merah pekat menunjukkan bahwa jumlah penderita tuberkulosis di wilayah tersebut dikatakan tinggi yaitu antara 2394-6338 kasus. Terdapat 1 kasus yang termasuk dalam kategori sangat tinggi yaitu Kota Surabaya. Dengan begitu dapat diketahui bahwa daerah dengan jumlah penderita tuberkulosis tertinggi terletak pada Kota Surabaya sedangkan daerah yang jumlah penderita tuberkulosis yang paling rendah yaitu Kota Batu.

#### 3.2 Analisis Kluster K-Means

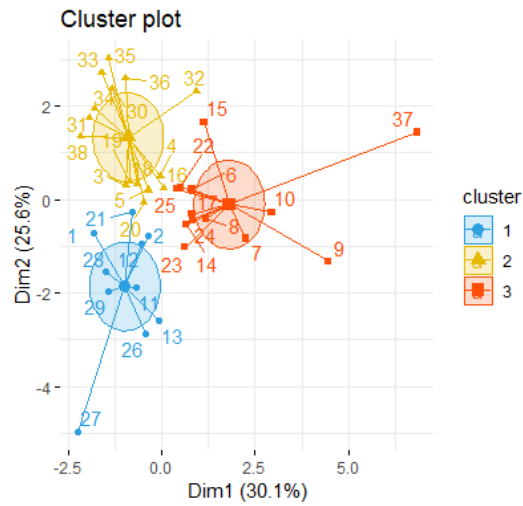
Pada penelitian ini menggunakan analisis kluster k-means dengan jumlah kluster dua, tiga, empat dan lima untuk mengelompokkan Kabupaten/Kota di Jawa Timur berdasarkan variabel yang diduga mempengaruhi jumlah penyakit tuberkulosis. Setiap anggota kluster disajikan pada Gambar 2 – Gambar 5. Rincian anggota tiap kluster dapat dilihat pada Lampiran 1.



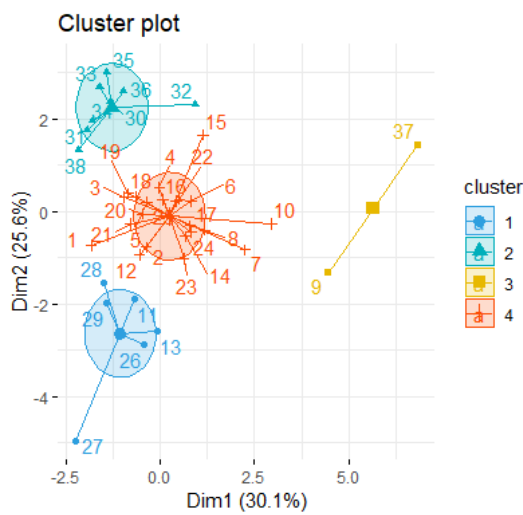
Gambar 1: Peta Persebaran Jumlah Kasus Tuberkulosis di Provinsi Jawa Timur.



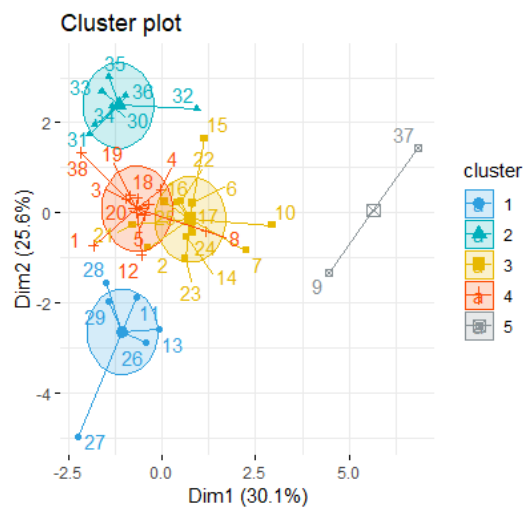
Gambar 2: Anggota Kluster K-Means k=2



Gambar 3: Anggota Kluster K-Means k=3



Gambar 4: Anggota Kluster K-Means k=4



Gambar 5: Anggota Kluster K-Means k=5

### 3.3 Pemeriksaan Multikolinieritas

Berdasarkan hasil pemeriksaan multikolinieritas dengan melihat nilai VIF yang disajikan pada Tabel 1 diperoleh hasil bahwa nilai VIF pada setiap variabel independen kurang dari 10. Sehingga dapat disimpulkan bahwa model terbebas dari multikolinieritas.

Tabel 1: Nilai VIF dari model yang digunakan.

Variabel	Nilai VIF
X <sub>1</sub>	1.46
X <sub>2</sub>	3.45
X <sub>3</sub>	1.21
X <sub>4</sub>	2.68
X <sub>5</sub>	3.15
X <sub>6</sub>	1.81
X <sub>7</sub>	1.48
X <sub>8</sub>	4.17
X <sub>9</sub>	3.31
X <sub>10</sub>	4.65

### 3.4 Regresi Linear Berganda

Berdasarkan hasil pengujian parameter secara serentak diperoleh nilai  $F_{hitung}$  sebesar 12.15 yang berarti lebih dari  $F_{tabel}$  (2.19) dan  $p$ -value sebesar 0.00 yang berarti kurang dari  $\alpha = 5\%$  sehingga tolak  $H_0$ . Maka dapat disimpulkan bahwa variabel independen berpengaruh secara serentak terhadap variabel dependen.

Setelah pengujian parameter secara serentak selanjutnya dilakukan pengujian parameter secara parsial atau individu untuk mengetahui variabel apa saja yang berpengaruh secara signifikan terhadap variabel dependen. Hasil pengujian parameter secara parsial disajikan pada Tabel 2.

Berdasarkan Tabel 2 diketahui bahwa terdapat dua variabel independen yang berpengaruh secara signifikan terhadap variabel dependen karena memiliki nilai  $p$ -value  $< \alpha$  yaitu variabel kepadatan penduduk ( $X_2$ ) dan variabel jumlah rumah sehat ( $X_9$ ). Dari hasil pengujian diperoleh persamaan regresi linier berganda sebagai berikut.

$$\hat{y} = 670.7 + 0.48X_{i1} + 0.23X_{i2} + 5.61X_{i3} - 25.87X_{i4} - 0.33X_{i5} + 0.80X_{i6} - 1.84X_{i7} + 60.68X_{i8} + 0.01X_{i9} + 0.76X_{i10}$$

Berdasarkan model yang diperoleh dapat diinterpretasikan bahwa setiap kepadatan penduduk naik sebesar satu satuan dan variabel lainnya konstan maka jumlah penderita tuberkulosis di Jawa Timur naik sebesar 0,23. Dan setiap jumlah rumah sehat bertambah sebesar satu satuan dan variabel lainnya konstan maka jumlah penderita tuberkulosis di Jawa Timur naik sebesar 0.01.

Tabel 2: Hasil pengujian parameter secara parsial.

Parameter	Koefisien	t-value	p-value
$\hat{\beta}_0$	670.7	0.31	0.75
$\hat{\beta}_1$	0.48	0.55	0.58
$\hat{\beta}_2$	0.23	3.06	*0.00
$\hat{\beta}_3$	5.61	0.49	0.62
$\hat{\beta}_4$	-25.87	-1.44	0.15
$\hat{\beta}_5$	-0.33	-0.82	0.41
$\hat{\beta}_6$	0.80	0.09	0.92
$\hat{\beta}_7$	-1.84	-0.26	0.79
$\hat{\beta}_8$	60.68	1.60	0.12
$\hat{\beta}_9$	0.01	5.66	*0.00
$\hat{\beta}_{10}$	0.76	0.77	0.44

\*signifikan pada  $\alpha = 5\%$

Selain itu dari hasil pengujian diperoleh  $R^2$  sebesar 0.81 yang dapat diartikan bahwa sebesar 81.81% variasi jumlah kasus tuberkulosis di Jawa Timur dapat dijelaskan oleh variabel independen, dan sisanya sebesar 18.19% dijelaskan oleh variabel lain yang tidak dimasukkan kedalam model regresi dalam penelitian ini.

Setelah diperoleh model, langkah selanjutnya yaitu menguji asumsi residual pada model regresi linier berganda. Pengujian yang dilakukan adalah pengujian asumsi residual identik, independen, dan berdistribusi normal. Pengujian asumsi residual disajikan pada Tabel 3.

Tabel 3: Pengujian Asumsi Residual

Uji	Nilai	p-value
<i>Breusch-Pagan</i> (Identik)	13.01	0.22
<i>Durbin-Watson</i> (Independen)	1.92	0.08
<i>Shapiro-Wilk</i> (Normalitas)	0.96	0.27

Berdasarkan hasil pengujian asumsi residual diketahui bahwa semua asumsi telah terpenuhi karena semua *p-value* dalam pengujian lebih dari  $\alpha=5\%$  yang berarti gagal tolak  $H_0$ . Sehingga dapat disimpulkan bahwa asumsi residual identik, independen, dan berdistribusi normal telah terpenuhi.

### 3.5 Moran's I

Berdasarkan pengujian *Moran's I* dengan pemberian pembobot *queen contiguity* dan pembobot spasial dari hasil analisis kluster k-means k=2, k=3, k=4 dan k=5 yang disajikan pada Tabel 4 dapat diketahui bahwa pembobot *queen contiguity* dan kluster k-means k=3 tidak signifikan karena memiliki nilai *p-value* lebih dari  $\alpha =5\%$  sedangkan



untuk pembobot kluster k-menas k=2, k=4, dan k=5 signifikan karena memiliki nilai *p-value* kurang dari  $\alpha = 5\%$ .

Tabel 4: Perbandingan *Moran's I* Pembobot *Queen* dan Kluster

Pembobot	<i>Moran's I</i>	<i>p-value</i>
<i>Queen</i>	-0.003	0.52
<i>Contiguity</i>		
K-Means k=2	0.04	* $1.3 \times 10^{-6}$
K-Means k=3	-0.06	0.97
K-Means k=4	-0.18	* $2.8 \times 10^{-5}$
K-Means k=5	-0.17	*0.03

\*signifikan pada  $\alpha = 5\%$

### 3.6 Lagrange Multiplier (LM)

Berdasarkan hasil pengujian LM menggunakan pembobot *queen contiguity* dan pembobot kluster k-means k=2, k=3, k=4 dan k=5 yang disajikan pada Tabel 5 diperoleh hasil bahwa pada pembobot *queen contiguity*, pembobot kluster k-means k=3 dan pembobot kluster k-means k=5 tidak ada yang signifikan karena semua *p-value* lebih dari  $\alpha = 5\%$ . Pada pembobot kluster k-means k=2 signifikan pada model SAR dan SARMA sedangkan pada pembobot kluster k-means k=4 signifikan pada semua model yaitu SAR, SEM, dan SARMA.

Tabel 5: Uji *Lagrange Multiplier* (LM).

Pembobot	Uji LM	Nilai	<i>p-value</i>
<i>Queen</i>	SAR	0.72	0.39
<i>Contiguity</i>	SEM	0.00	0.97
	SARMA	1.26	0.53
K-Means k=2	SAR	9.66	*0.00
	SEM	0.57	0.44
	SARMA	11.56	*0.00
K-Means k=3	SAR	0.31	0.57
	SEM	0.87	0.34
	SARMA	2.04	0.35
K-Means k=4	SAR	5.38	*0.02
	SEM	4.49	*0.03
	SARMA	6.32	*0.04
K-Means k=5	SAR	2.13	0.14
	SEM	3.25	0.07
	SARMA	3.42	0.18

\*signifikan pada  $\alpha = 5\%$

Berdasarkan hasil uji dependensi dapat disimpulkan bahwa pembobot kluster k-means k=2 dengan model SAR dan SARMA terpilih untuk pemodelan jumlah penderita

tuberkulosis di Jawa Timur karena memiliki nilai *p-value* yang paling signifikan dibandingkan dengan yang lainnya.

### 3.7 Spatial Autoregressive Model (SAR) Pembobot Klaster K-Means k=2

Hasil estimasi dan pengujian parameter *Spatial Autoregressive Model* (SAR) disajikan pada Tabel 6.

Tabel 6: Hasil Estimasi dan Pengujian Parameter SAR

Parameter	Koefisien	<i>p-value</i>
$\rho$	0.61	*0.01
$\hat{\beta}_0$	98.24	0.95
$\hat{\beta}_1$	-0.54	0.51
$\hat{\beta}_2$	0.23	*0.00
$\hat{\beta}_3$	5.16	0.59
$\hat{\beta}_4$	-22,52	0.14
$\hat{\beta}_5$	-0.59	0.11
$\hat{\beta}_6$	2.39	0.74
$\hat{\beta}_7$	-4.44	0.47
$\hat{\beta}_8$	43.77	0.17
$\hat{\beta}_9$	0.01	*0.00
$\hat{\beta}_{10}$	1,01	0.22

\*signifikan pada  $\alpha=5\%$

Berdasarkan Tabel 6 dapat dituliskan persamaan model regresi *Spatial Autoregressive Model* (SAR) sebagai berikut.

$$\hat{y}_i = 98.24 + 0.61 \sum_{j=1, i \neq j}^n w_{ij} y_j - 0.54X_{i1} + 0.23X_{i2} + 5.16X_{i3} - 22.52X_{i4} - 0.59X_{i5} + 2.39X_{i6} - 4.44X_{i7} + 43.77X_{i8} + 0.01X_{i9} + 1.01X_{i10}$$

Hasil pengujian pada Tabel 6 menunjukkan bahwa terdapat dependensi spasial pada lag ( $\rho$ ) yang memiliki *p-value* kurang dari  $\alpha=5\%$  yang artinya ada keterkaitan jumlah penderita tuberkulosis pada suatu wilayah dengan wilayah lainnya yang berada pada satu kelompok. Selain itu terdapat dua variabel yang signifikan pada  $\alpha = 5\%$  yaitu variabel kepadatan penduduk ( $X_2$ ) dan variabel jumlah rumah sehat ( $X_9$ ). Maka dapat disimpulkan bahwa kedua variabel tersebut berpengaruh secara signifikan terhadap jumlah penderita tuberkulosis di Jawa Timur. Hasil pengujian menghasilkan nilai  $R^2$  sebesar 0.8515 yang artinya sebesar 85.15% variasi jumlah penderita tuberkulosis di Jawa Timur dapat dijelaskan oleh variabel independen, sedangkan sisanya 14.844% dijelaskan oleh variabel lain yang tidak dimasukkan kedalam model.

Setelah diperoleh persamaan model regresi SAR dengan pembobot kluster k-means k=2 langkah selanjutnya yaitu pengujian asumsi residual pada model. Pengujian asumsi residual yang dilakukan adalah uji residual identik, uji residual independen, dan uji residual berdistribusi normal. Hasil pengujian disajikan pada Tabel 7

Tabel 7: Pengujian Asumsi Residual Model SAR

Uji	Nilai	<i>p-value</i>
<i>Breusch-Pagan</i> (Identik)	14.79	0.13
<i>Durbin-Watson</i> (Independen)	1.30	-
<i>Shapiro-Wilk</i> (Normalitas)	0.97	0.61

Berdasarkan hasil pengujian asumsi residual model SAR pada Tabel 17 diketahui bahwa semua asumsi telah terpenuhi karena semua *p-value* dalam pengujian lebih dari  $\alpha=5\%$  yang berarti gagal tolak  $H_0$ . Pada uji DW diwakilkan dengan nilai DW yang berada diantara nilai  $dU = 2,1803$  dan  $4-dU = 1,8197$  yang berarti gagal tolak  $H_0$  sehingga dapat disimpulkan bahwa tidak terjadi autokorelasi Sehingga dapat disimpulkan bahwa asumsi residual identik, independen, dan berdistribusi normal telah terpenuhi.

### 3.8 *Spatial Autoregressive Moving Average (SARMA) Pembobot Kluster K-Means k=2*

Hasil estimasi dan pengujian parameter *Spatial Autoregressive Moving Average (SARMA)* disajikan pada Tabel 8.

Tabel 8: Hasil Estimasi dan Pengujian Parameter SARMA

Parameter	Koefisien	<i>p-value</i>
$\rho$	0.69	*0.00
$\lambda$	-5.62	0.08
$\hat{\beta}_0$	-490.17	0.80
$\hat{\beta}_1$	-0.43	0.59
$\hat{\beta}_2$	0.25	*0.00
$\hat{\beta}_3$	6.61	0.54
$\hat{\beta}_4$	-20.08	0.21
$\hat{\beta}_5$	-0.56	0.11
$\hat{\beta}_6$	1.78	0.81
$\hat{\beta}_7$	-4.18	0.49
$\hat{\beta}_8$	51.63	0.14
$\hat{\beta}_9$	0,01	*0.00
$\hat{\beta}_{10}$	1.02	0.24

\*signifikan pada  $\alpha =5\%$

Berdasarkan Tabel 8 dapat dituliskan persamaan model regresi *Spatial Autoregressive Moving Average* (SARMA) sebagai berikut.

$$\hat{y}_i = -490.17 + 0.69 \sum_{j=1, i \neq j}^n w_{ij} y_j - 0.43X_{i1} + 0.25X_{i2} + 6.61X_{i3} - 20.08X_{i4} - 0.56X_{i5} +$$

$$1.78X_{i6} - 4.18X_{i7} + 51.63X_{i8} + 0,01X_{i9} + 1.02X_{i10} + u_i$$

$$u_i = -5.62 \sum_{j=1, i \neq j}^n w_{ij} u_j$$

Hasil pengujian pada Tabel 8 menunjukkan bahwa terdapat dependensi spasial pada *lag* ( $\rho$ ) karena memiliki nilai *p-value* kurang dari  $\alpha = 5\%$ , sedangkan pada *error* ( $\lambda$ ) tidak signifikan karena memiliki *p-value* lebih dari  $\alpha = 5\%$  yang artinya ada keterkaitan jumlah penderita tuberkulosis pada suatu wilayah dengan wilayah lainnya yang berada pada satu kelompok. Selain itu terdapat dua variabel yang signifikan pada  $\alpha = 5\%$  yaitu variabel kepadatan penduduk ( $X_2$ ) dan variabel jumlah rumah sehat ( $X_9$ ). Maka dapat disimpulkan bahwa kedua variabel tersebut berpengaruh secara signifikan terhadap jumlah penderita tuberkulosis di Jawa Timur. Hasil pengujian menghasilkan nilai  $R^2$  sebesar 0.871 yang artinya sebesar 87.10 % variasi jumlah penderita tuberkulosis di Jawa Timur dapat dijelaskan oleh variabel independen, sedangkan sisanya 12.9% dijelaskan oleh variabel lain yang tidak dimasukkan kedalam model.

Setelah diperoleh persamaan model regresi SARMA dengan pembobot kluster k-menas k=2 langkah selanjutnya yaitu pengujian asumsi residual pada model tersebut. Pengujian asumsi residual yang dilakukan adalah uji residual identik, uji residual independen, dan uji residual berdistribusi normal. Hasil pengujian disajikan pada tabel 9.

Tabel 9: Pengujian Asumsi Residual Model SARMA

Uji	Nilai	<i>p-value</i>
<i>Breusch-Pagan</i> (Identik)	18.11	0.05
<i>Durbin-Watson</i> (Independen)	1.29	-
<i>Shapiro-Wilk</i> (Normalitas)	0.97	0.51

Berdasarkan hasil pengujian asumsi residual model SARMA pada Tabel 9 diketahui bahwa semua asumsi telah terpenuhi karena semua *p-value* dalam pengujian lebih dari  $\alpha=5\%$  yang berarti gagal tolak  $H_0$ . Pada uji DW diwakili dengan nilai DW yang berada diantara nilai  $dU = 2.1803$  dan  $4-dU = 1.8197$  yang berarti gagal tolak  $H_0$  sehingga dapat disimpulkan bahwa tidak terjadi autokorelasi Sehingga dapat disimpulkan bahwa asumsi residual identik, independen, dan berdistribusi normal telah terpenuhi.

### 3.9 Pemilihan Model Terbaik

Pemilihan model terbaik pada penelitian ini menggunakan kriteria pengujian  $R^2$  dan nilai AIC untuk membandingkan antara model Regresi Linear Berganda (OLS), *Spatial Autoregressive Model* (SAR) dan *Spatial Autoregressive Moving Average* (SARMA). Hasil pengujian disajikan pada Tabel 10

Tabel 10: Nilai  $R^2$  dan AIC

Model	$R^2$	AIC
OLS	81.81%	595.75
SAR	85.15%	590.04
SARMA	87.10%	586.69

Berdasarkan Tabel 10 menunjukkan nilai  $R^2$  pada model SARMA dengan pembobot klaster k-means k=2 lebih besar dibandingkan dengan model lainnya yang menunjukkan bahwa model SARMA dengan pembobot klaster k-means k=2 lebih baik dalam menjelaskan variabel dependen. Selain itu jika dilihat dari nilai AIC, model SARMA pembobot klaster k-means k=2 memiliki nilai AIC yang lebih kecil dibandingkan model lainnya. Maka dapat disimpulkan bahwa model yang terbaik untuk memodelkan jumlah penderita tuberkulosis di Jawa Timur adalah model SARMA dengan pembobot klaster k-means k=2.

## 4. Simpulan dan Saran

### 4.1 Simpulan

Berdasarkan pembahasan mengenai aplikasi regresi spasial menggunakan pembobot dari hasil klaster k-means untuk pemodelan Tuberkulosis di Jawa Timur, maka dapat diambil kesimpulan bahwa pemodelan Tuberkulosis di Jawa Timur dipilih model *Spatial Autoregressive Moving Average* (SARMA) dengan pembobot klaster k-means k=2. Hasil analisis menunjukkan terdapat dua variabel yang berpengaruh secara signifikan terhadap jumlah penderita tuberkulosis di Jawa Timur pada tahun 2017 yaitu variabel kepadatan penduduk ( $X_2$ ) dan variabel jumlah rumah sehat ( $X_9$ ) dengan nilai  $R^2$  dan AIC berturut-turut 87.10% dan 586.69.

### 4.2 Saran

Dalam penulisan penelitian ini menggunakan *Spatial Autoregressive and Moving Average* (SARMA) dengan menggunakan pembobot dari hasil klaster k-means. Pengembangan lebih lanjut menggunakan penelitian menggunakan metode analisis spasial dengan mencoba membuat pembobot dari hasil klaster yang lain seperti *fuzzy c-means* atau klaster nonhirarki, selain itu dapat menambahkan variabel lain dalam penelitian selanjutnya sehingga memungkinkan model spasial dengan pendekatan area yang lainnya seperti *Spatial Error Model* (SEM) atau *Spatial Autoregressive Model* (SAR).

**Daftar Pustaka**

- Angisna, T. (2018). *Aplikasi Regresi Spasial untuk Mengetahui Pengaruh Faktor-Faktor Persebaran TB Paru di Kabupaten Magetan [skripsi]*. Surabaya (ID): Universitas Airlangga.
- Anselin, L. (2005). *Spatial regression analysis in R: a workbook*. University of Illinois, Urbana Champaign.
- Anselin, L. (2013). *Spatial econometrics: methods and models*. Springer Science & Business Media.
- [BPS] Badan Pusat Statistik. (2018). *Provinsi Jawa Timur dalam Angka tahun 2018*. Surabaya (ID): Badan Pusat Statistik Jawa Timur.
- [DINKES] Dinas Kesehatan Jawa Timur. (2018). *Profil Kesehatan Provinsi Jawa Timur Tahun 2017*. Surabaya (ID): Dinas Kesehatan Jawa Timur.
- LeSage, J. P., & Pace, R. K. (2009). *Introduction to Spatial Econometrics*. Boca Raton (US): CRC Press.
- Suherni, N. A. D., & Maduratna, M. (2013). Analisis Pengelompokan kecamatan di kota surabaya berdasarkan faktor penyebab terjadinya penyakit tuberkulosis. *Jurnal Sains Dan Seni ITS*, 2(1): D13–D18.

## Lampiran 1 Rincian Anggota Klaster

1. Rincian Anggota Tiap Klaster K-Means  $k=2$ 

**Klaster 1:** Kab. Ponorogo, Kab. Kediri, Kab. Malang, Kab. Lumajang, Kab. Jember, Kab. Banyuwangi, Kab. Bondowoso, Kab. Situbondo, Kab. Probolinggo, Kab. Pasuruan, Kab. Mojokerto, Kab. Jombang, Kab. Magetan, Kab. Ngawi, Kab. Bojonegoro, Kab. Tuban, Kab. Lamongan, Kab. Gresik, Kab. Bangkalan, Kab. Sampang, Kab. Pamekasan, Kab. Sumenep, Kota Surabaya

**Klaster 2:** Kab. Pacitan, Kab. Trenggalek, Kab. Tulungagung, Kab. Blitar, Kab. Sidoarjo, Kab. Nganjuk, Kab. Madiun, Kota Kediri, Kota Blitar, Kota Malang, Kota Probolinggo, Kota Pasuruan, Kota Mojokerto, Kota Madiun, Kota Batu
2. Rincian Anggota Tiap Klaster K-Means  $k=3$ 

**Klaster 1:** Kab. Ngawi, Kab. Ponorogo, Kab. Situbondo, Kab. Pacitan, Kab. Pamekasan, Kab. Bondowoso, Kab. Sumenep, Kab. Probolinggo, Kab. Bangkalan, Kab. Sampang

**Klaster 2:** Kab. Madiun, Kab. Trenggalek, Kab. Nganjuk, Kab. Magetan, Kab. Blitar, Kab. Mojokerto, Kab. Tulungagung, Kota Mojokerto, Kota Probolinggo, Kota Madiun, Kota Malang, Kota Kediri, Kota Pamekasan, Kota Blitar, Kota Batu

**Klaster 3:** Kab. Sidoarjo, Kab. Bojonegoro, Kab. Kediri, Kab. Gresik, Kab. Jombang, Kab. Pasuruan, Kab. Lumajang, Kab. Lamongan, Kab.
3. Rincian Anggota Tiap Klaster K-Means  $k=4$ 

**Klaster 1:** Kab. Pamekasan, Kab. Bondowoso, Kab. Sumenep, Kab. Probolinggo, Kab. Bangkalan, Kab. Sampang

**Klaster 2:** Kota Mojokerto, Kota Probolinggo, Kota Madiun, Kota Kediri, Kota Malang, Kota Pasuruan, Kota Blitar, Kota Batu

**Klaster 3:** Kab. Jember, Kota Surabaya

**Klaster 4:** Kab. Sidoarjo, Kab. Madiun, Kab. Tulungagung, Kab. Bojonegoro, Kab. Trenggalek, Kab. Mojokerto, Kab. Kediri, Kab. Ngawi, Kab. Blitar, Kab. Gresik, Kab. Banyuwangi, Kab. Jombang, Kab. Lumajang, Kab. Magetan, Kab. Magetan, Kab. Lamongan, Kab. Ponorogo, Kab. Pasuruan, Kab. Malang, Kab. Pacitan, Kab. Situbondo, Kab. Tuban
4. Rincian Anggota Tiap Klaster K-Means  $k=5$ 

**Klaster 1:** Kab. Pamekasan, Kab. Bonodowo, Kab. Sumenep, Kab. Probolinggo, Kab. Bangkalan, Kab. Sampang

**Klaster 2:** Kota Mojokerto, Kota Probolinggo, Kota Madiun, Kota Kediri, Kota Pasuruan, Kota Blitar

**Klaster 3:** Kab. Sidoarjo, Kab. Bojonegoro, Kab. Kediri, Kab. Mojokerto, Kab. Gresik, Kab. Banyuwangi, Kab. Jombang, Kab. Ngawi, Kab. Lamongan, Kab. Pasuruan, Kab. Ponorogo, Kab. Tuban, Kab. Malang

**Klaster 4:** Kota Batu, Kab. Madiun, Kab. Trenggalek, Kab. Tulungagung, Kab. Nganjuk, Kab. Blitar, Kab. Magetan, Kab. Lumajang, Kab. Pacitan, Kab. Situbondo

**Klaster 5:** Kota Surabaya, Kab. Jember