

Madani: Jurnal Ilmiah Multidisiplin

Volume 3, Nomor 3, April 2025, P. 219-229

E-ISSN: 2986-6340

Licensed by CC BY-SA 4.0

DOI: <https://doi.org/10.5281/zenodo.15276145>

Analisis Spasial Persebaran COVID-19 di Indonesia Menggunakan Metode K-Means Clustering dan ESD

Christoffel Mario¹, Linda Natasya Siahaan², Junitro Andreas Simanullang³, Tabita Paulina Simamora^{4*}

¹²³⁴Program Studi Statistika, FMIPA, Universitas Negeri Medan

Corresponding: tabitapaulsimamora@gmail.com

Abstrak

Pandemi COVID-19 yang terjadi sejak Maret 2020 berdampak besar terhadap kesehatan masyarakat di Indonesia. Meskipun jumlah kasus mulai menurun, pemahaman terhadap pola penyebaran virus tetap diperlukan untuk penanggulangan yang lebih efektif. Analisis konvensional yang hanya menggunakan statistik deskriptif belum mampu mengungkap keterkaitan antarwilayah secara geografis. Penelitian ini menggabungkan metode *Exploratory Spatial Data Analysis* (ESDA) dan *K-Means Clustering* untuk menganalisis pola spasial dan mengelompokkan provinsi-provinsi di Indonesia berdasarkan jumlah kasus COVID-19, tingkat kesembuhan, dan tingkat kematian. Data yang digunakan berasal dari shapefile provinsi Indonesia (GADM) dan data kasus COVID-19 dari Data Wrapper. Hasil analisis menunjukkan tiga kluster utama. Kluster pertama terdiri dari DKI Jakarta, Jawa Barat, dan Jawa Tengah dengan kasus dan kematian tinggi serta kesembuhan rendah. Kluster kedua mencakup Jawa Timur, Sumatera Utara, dan Sulawesi Selatan dengan kasus lebih rendah, tetapi kesembuhan sangat rendah dan kematian tinggi. Kluster ketiga mencakup 26 provinsi lainnya yang memiliki kasus lebih rendah, kesembuhan tinggi, dan kematian rendah. Hasil ini menunjukkan bahwa penyebaran COVID-19 tidak merata secara spasial, sehingga penting untuk memberikan perhatian khusus pada wilayah-wilayah yang tergolong kluster risiko tinggi.

Kata kunci: COVID-19, clustering, K-means, LISA, Spasial

Abstract

The COVID-19 pandemic, which began in March 2020, has had a significant impact on public health in Indonesia. Although case numbers have started to decline, understanding the spatial spread of the virus remains crucial for effective response efforts. Conventional analyses that rely solely on descriptive statistics often overlook spatial relationships between regions. This study combines *Exploratory Spatial Data Analysis* (ESDA) and *K-Means Clustering* to examine the spatial distribution of COVID-19 cases and group Indonesian provinces based on the number of cases, recovery rates, and mortality rates. The data used include Indonesia's provincial shapefiles from GADM and COVID-19 case data from Data Wrapper. The analysis reveals three main clusters. Cluster one includes DKI Jakarta, West Java, and Central Java, characterized by high case numbers and mortality rates, with below-average recovery rates. Cluster two consists of East Java, North Sumatra, and South Sulawesi, with relatively low case numbers, very low recovery rates, and high mortality rates. Cluster three comprises 26 other provinces with lower case numbers, high recovery rates, and low mortality rates. These findings indicate that COVID-19 transmission in Indonesia is not spatially uniform, highlighting the need for targeted intervention in high-risk areas.

Kata Kunci: COVID-19, clustering, K-means, LISA, Spasial

Article Info

Received date: 27 March 2025

Revised date: 30 March 2025

Accepted date: 15 April 2025

PENDAHULUAN

Pada bulan Desember 2019, kasus yang mirip seperti pneumonia dilaporkan pertama kali terdapat di Wuhan, Hubei, Cina. Pada saat itu kasus tersebut masih dianggap misterius karena belum diketahui pasti sumber penyakit tersebut, namun diduga ada hubungannya dengan keberadaan pasar ikan di Wuhan. Semenjak itu, banyak peneliti berlomba-lomba untuk menemukan etiologi penyakit tersebut. Penelitian yang dilakukan pada sampel menunjukkan etiologi Coronavirus baru (Ren et al., 2020). Awalnya penyakit diberi nama 2019 Novel Coronavirus (2019-nCoV), namun pada tanggal 11 Februari 2020 World Health Organization (WHO) memberikan nama baru Coronavirus Disease (Covid-19). Penyakit tersebut disebabkan oleh virus Severe Acute Respiratory Syndrome Coronavirus-2 (SARS CoV-2). Selain itu, WHO juga mengumumkan bahwa penyakit tersebut telah menyebar secara

luas di Cina dan lebih dari 190 negara lainnya dengan metode penularan dari manusia ke manusia. Oleh karena itu, tanggal 12 Maret 2020 WHO menetapkan Covid-19 sebagai pandemik. Hingga tanggal 18 Juni 2020, terdapat 8.352.940 kasus dan 449.785 jumlah kematian di seluruh dunia (World Health Organization, 2020).

Pandemi COVID-19 yang melanda Indonesia sejak awal Maret 2020 memberikan dampak yang luar biasa terhadap seluruh sektor kehidupan, khususnya kesehatan masyarakat. Penyebaran kasus COVID-19 menunjukkan variasi yang signifikan antarprovinsi di Indonesia, dipengaruhi oleh berbagai faktor seperti kepadatan penduduk, mobilitas masyarakat, dan kesiapan infrastruktur kesehatan di masing-masing wilayah (susilawati, dkk. 2020.). Meskipun angka kasus COVID-19 semakin menurun, namun penanggulangan yang tepat memerlukan pemahaman yang mendalam terhadap pola penyebaran penyakit ini.

Analisis konvensional terhadap data COVID-19 cenderung terbatas pada statistik deskriptif dan tidak memperhitungkan aspek spasial dalam penyebarannya. Padahal, interaksi antarwilayah dan mobilitas penduduk memainkan peran penting dalam mempercepat atau memperlambat penyebaran virus (duman care khrisne, dkk. Labeling Indonesia COVID-19 Data Using K-Means Clustering with Optimization). Sebagai upaya pencegahan perkembangan Covid-19 di Indonesia, pada penelitian ini melakukan clustering untuk mengetahui daerah mana saja yang terdampak. Dengan pengelompokan ini juga dapat mengkategorikan kelompok daerah yang potensial penyebaran Covid-19 dan kelompok daerah yang memiliki kasus Covid-19 tertinggi. Clustering bertujuan pada pengelompokan seperti record, pengamatan serta membentuk kelas dari objek yang memiliki kemiripan dengan objek lainnya. K-Means merupakan salah satu algoritma clustering yang masuk dalam gerombolan Unsupervised learning yang dipergunakan mengelompokkan data atau objek kedalam kelompok dengan sistem partisi (Zulfa Nabila, 2021).

Pendekatan berbasis spasial, seperti *Exploratory Spatial Data Analysis* (ESDA), dapat mengungkapkan keterkaitan antarwilayah secara geografis, sehingga dapat membantu dalam perencanaan penanggulangan yang lebih efektif. Beberapa penelitian telah menggunakan metode ESDA untuk memeriksa autokorelasi spasial dalam data COVID-19. Misalnya, Rachmawati (2024) mengaplikasikan Moran's I dan Local Indicators of Spatial Association (LISA) untuk menganalisis konsentrasi kasus COVID-19 di Indonesia. Hasilnya menunjukkan bahwa terdapat klaster-klaster provinsi dengan jumlah kasus yang lebih tinggi di wilayah Indonesia bagian barat dan tengah (annisa rachmawati, dkk. 2024). LISA adalah pengukuran autokorelasi spasial secara lokal yang digunakan untuk menganalisis hubungan spasial antara suatu lokasi pengamatan terhadap lokasi pengamatan lainnya. LISA sebagai suatu statistik yang dapat memberikan petunjuk adanya pengelompokan hubungan spasial yang signifikan dari nilai yang sama di sekitar daerah tersebut dan jumlah dari nilai LISA untuk seluruh wilayah sebanding dengan nilai indeks Moran (abimanyu kurniadi, 2018). Indeks Moran adalah pengukuran autokorelasi spasial secara global yang digunakan untuk menganalisis hubungan spasial pada suatu kejadian. Analisis ini menggambarkan autokorelasi spasial di suatu wilayah secara keseluruhan (Anuraga Gangga, 2017). Indeks Moran tidak memberikan informasi pola spasial pada wilayah tertentu. Oleh karena itu, diperlukan kecenderungan adanya hubungan spasial di setiap lokasi dengan menggunakan LISA (referensi proyek spasial 7). Selain itu, penelitian oleh Fitriyani & Abdulloh (2022) mengungkapkan bahwa klasterisasi spasial dapat membantu menentukan wilayah yang memerlukan perhatian khusus dalam pengendalian pandemi (nurul khasanah fitriyani, 2021).

Selanjutnya, metode *K-Means Clustering* sering digunakan dalam mengelompokkan data berdasarkan kesamaan karakteristik. Data – data dipilih menjadi beberapa kelompok dengan kriteria yang telah ditentukan lalu dikumpulkan menjadi satu dalam sebuah Cluster. Dimana setiap Cluster memiliki titik pusat yang disebut Centroid (dwitri, nayuni, dkk. 2020). Metode ini terbukti efektif dalam mengidentifikasi pola risiko di berbagai daerah. Dalam konteks COVID-19, K-Means digunakan untuk mengelompokkan provinsi berdasarkan jumlah kasus, tingkat kesembuhan, dan tingkat kematian. K-Means clustering memungkinkan pengguna melakukan pengelompokan data kedalam cluster berdasarkan variabel-variabel yang ada tanpa harus melalui proses training data terlebih dahulu. Tujuan utama dari algoritma K Means clustering ini untuk meminimalisir jarak antara titik data dengan cluster yang sesuai (Yudhistira, aditia. 2023). Taqiyyuddin (2022) menyimpulkan bahwa penggunaan K-Means membantu dalam mengelompokkan provinsi dengan pola penyebaran yang serupa, sehingga memudahkan dalam merancang kebijakan yang sesuai dengan karakteristik masing-masing wilayah (teguh amar taqiy).

Penelitian ini menggabungkan dua pendekatan yaitu ESDA dan K-Means Clustering. Dengan tujuan untuk menganalisis persebaran spasial kasus COVID-19 antarprovinsi di Indonesia untuk melihat pola keterkaitan wilayah, mengelompokkan provinsi berdasarkan karakteristik jumlah kasus, tingkat kesembuhan, dan tingkat kematian menggunakan metode K-Means Clustering, serta menginterpretasikan hasil analisis spasial dan clustering sebagai dasar identifikasi wilayah yang berisiko tinggi terhadap penyebaran COVID-19.

METODE PENELITIAN

Jenis Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan analisis spasial. Penelitian ini bertujuan untuk mengidentifikasi pola penyebaran COVID-19 di Indonesia dengan menggunakan metode K-Means Clustering dan Exploratory Spatial Data Analysis (ESDA). K-Means Clustering digunakan untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan karakteristik penyebaran COVID-19, sementara ESDA digunakan untuk mengidentifikasi adanya autokorelasi spasial pada distribusi kasus COVID-19.

Subjek dan Objek Penelitian

- Subjek Penelitian: Provinsi-provinsi di Indonesia yang terpapar COVID-19.
- Objek Penelitian: Persebaran kasus COVID-19, tingkat kesembuhan, dan tingkat kematian di setiap provinsi di Indonesia.

Data

Data yang digunakan dalam penelitian ini adalah sebagai berikut:

- 1). Data Spasial Provinsi Indonesia: Data ini diperoleh dari shapefile provinsi Indonesia yang disediakan oleh GADM (Geospatial Data for Administrative Mapping), yang mencakup batas administratif dan koordinat geografis masing-masing provinsi.
- 2). Data Kasus COVID-19: Data jumlah kasus COVID-19, tingkat kesembuhan, dan tingkat kematian per provinsi di Indonesia diperoleh dari situs Datawrapper (<https://datawrapper.dwcdn.net/MwHOx/5/>).

Metode Pengumpulan Data

Pengumpulan data dilakukan melalui dua langkah utama:

- 1). Pengunduhan Data Spasial: Data batas provinsi Indonesia diperoleh melalui platform GADM dalam format shapefile.
- 2). Pengunduhan Data COVID-19: Data jumlah kasus, jumlah sembuh, dan jumlah kematian COVID-19 di setiap provinsi diambil dari Datawrapper.

Instrumen Penelitian

Instrumen yang digunakan dalam penelitian ini meliputi:

- 1). Shapefile Provinsi: Digunakan untuk mendefinisikan batas wilayah administratif Indonesia.
- 2). Data COVID-19: Digunakan untuk analisis klusterisasi dan autokorelasi spasial.
- 3). Alat Pengolahan dan Analisis: Pengolahan data dan analisis spasial dilakukan menggunakan R dengan berbagai paket seperti sf, spdep, dan factoextra.

Analisis Data

- 1). K-Means Clustering: Digunakan untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan jumlah kasus COVID-19, tingkat kesembuhan, dan tingkat kematian. Metode ini bertujuan untuk mengidentifikasi kluster provinsi dengan karakteristik yang serupa.
 - a. Langkah pertama adalah menentukan jumlah kluster yang optimal menggunakan Metode Elbow.
 - b. Setelah jumlah kluster ditentukan, dilakukan analisis K-Means untuk membagi provinsi-provinsi dalam kluster berdasarkan variabel-variabel tersebut.
- 2). Exploratory Spatial Data Analysis (ESDA): Analisis ini digunakan untuk mengetahui adanya autokorelasi spasial pada persebaran COVID-19 di Indonesia. Salah satu teknik yang digunakan adalah Local Moran's I untuk mendeteksi pola spasial pada distribusi kasus COVID-19 antarprovinsi. Hasil analisis Local Moran's I ini menunjukkan apakah provinsi memiliki pola yang serupa dengan tetangganya atau tidak.

Validitas Data

Validitas data dalam penelitian ini diperoleh dengan cara memastikan bahwa data yang digunakan berasal dari sumber yang terpercaya dan valid. Data kasus COVID-19 yang diambil dari

Datawrapper diperoleh melalui sumber yang sah dan terupdate. Data spasial provinsi Indonesia yang diperoleh dari GADM juga merupakan sumber data yang diakui dalam penelitian geospasial.

Tahapan Penelitian

Tahapan penelitian ini terdiri dari beberapa langkah berikut:

1. Pengumpulan Data: Mengambil shapefile provinsi dan data COVID-19.
2. Pengolahan Data: Menggabungkan data, menghitung tingkat kesembuhan dan kematian, serta menstandarisasi data.
3. Penentuan Kluster: Menentukan jumlah kluster dengan Metode Elbow.
4. K-Means Clustering: Mengelompokkan provinsi berdasarkan data COVID-19.
5. Analisis Spasial: Menghitung Local Moran's I untuk menganalisis pola spasial.
6. Visualisasi Hasil: Membuat peta kluster dan peta spasial.
7. Penyimpanan Hasil: Menyimpan hasil analisis dan peta dalam format shapefile dan gambar.

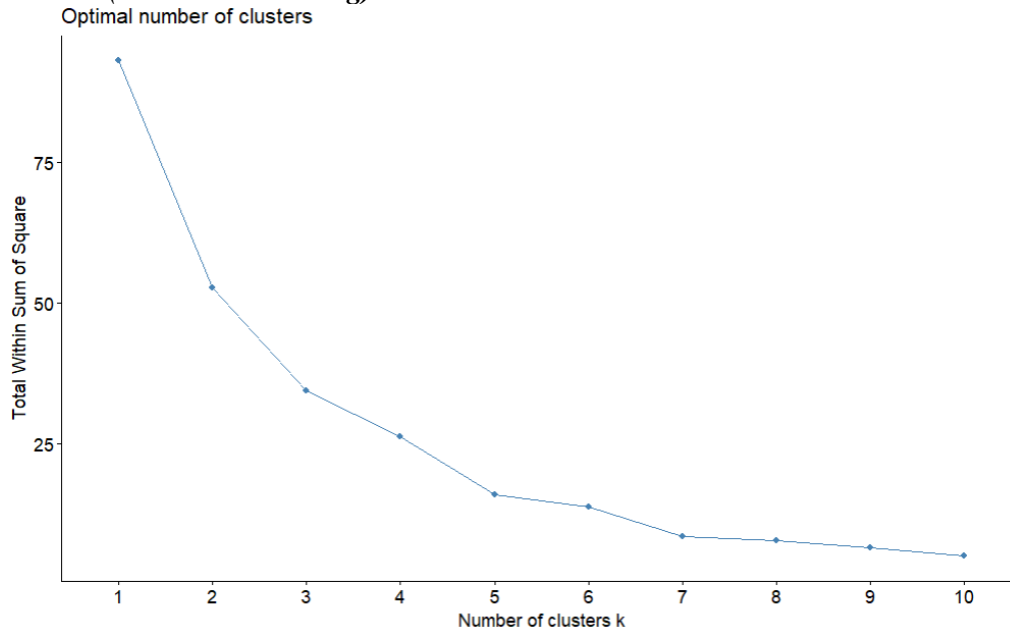
HASIL DAN PEMBAHASAN

Preprocessing Data

No	Provinsi Asal	Kasus	Sembuh	Kematian	Latitude	Longitude
1	Aceh	37.733	34.027	1.927	4,695135	96,7493993
2	Bali	112.582	107.627	3.943	-8,4095178	115,188916
3	Banten	131.475	128.002	2.670	-6,4058172	106,0640179
4	Bangka Belitung	51.209	48.924	1.397	-2,7410513	106,4405872
5	Bengkulu	23.021	22.463	466	-3,5778471	102,3463875
6	DI Yogyakarta	154.829	148.125	5.190	-7,8753849	110,4262088
7	DKI Jakarta	857.765	842.510	13.539	-6,211544	106,845172
8	Jambi	29.547	28.366	767	-1,4851831	102,4380581
9	Jawa Barat	702.722	685.219	14.624	-7,090911	107,668887
10	Jawa Tengah	482.009	448.597	29.894	-7,150975	110,1402594
11	Jawa Timur	395.475	364.251	29.413	-7,5360639	112,2384017
12	Kalimantan Barat	39.961	38.711	1.045	-0,2787808	111,4752851
13	Kalimantan Timur	156.811	150.282	5.381	1,6406296	116,419389
14	Kalimantan Tengah	45.162	42.338	1.367	-1,6814878	113,3823545
15	Kalimantan Selatan	69.425	66.531	2.359	-3,0926415	115,2837585
16	Kalimantan Utara	35.007	31.202	773	3,0731	116,0414
17	Kepulauan Riau	53.633	51.435	1.737	3,9456514	108,1428669
18	Nusa Tenggara Barat	27.423	26.380	799	-8,6529334	117,3616476
19	Sumatera Selatan	59.615	56.184	3.046	-3,3194374	103,914399
20	Sumatera Barat	89.276	86.244	2.121	-0,7399397	100,8000051
21	Sulawesi Utara	34.179	32.317	1.028	0,6246932	123,9750018
22	Sumatera Utara	104.706	100.076	2.837	2,1153547	99,5450974
23	Sulawesi Tenggara	20.030	19.216	523	-4,14491	122,174605
24	Sulawesi Selatan	108.584	104.498	2.206	-3,6687994	119,9740534
25	Sulawesi Tengah	46.341	43.858	1.561	-1,4300254	121,4456179
26	Lampung	49.064	43.964	3.764	-4,5585849	105,4068079
27	Riau	127.735	122.629	4.064	0,2933469	101,7068294
28	Maluku Utara	11.956	11.567	302	1,5709993	127,8087693
29	Maluku	14.524	14.113	258	-3,2384616	130,1452734

30	Papua Barat	22.899	22.461	352	-1,3361154	133,1747162
31	Papua	33.746	31.441	496	-4,269928	138,0803529
32	Sulawesi Barat	12.159	11.567	336	-2,8441371	119,2320784
33	Nusa Tenggara Timur	62.716	60.747	1.296	-8,6573819	121,0793705
34	Gorontalo	11.735	11.152	458	0,6999372	122,4467238

Analisis Kluster (K-Means Clustering)



Gambar tersebut menunjukkan hasil Elbow Method untuk menentukan jumlah kluster yang optimal dalam analisis K-Means clustering yang merupakan algoritma klusterisasi yang bertujuan membagi data ke dalam kelompok berdasarkan kedekatan terhadap pusat kluster. (Hartigan & Wong, 1979) Grafik ini memplot jumlah kluster (k) terhadap nilai Total Within Sum of Squares (WSS), yang menggambarkan seberapa rapat data dalam setiap kluster. Terlihat bahwa penurunan nilai WSS sangat tajam dari $k = 1$ hingga $k = 3$, kemudian mulai melandai setelah $k = 3$. Titik siku atau "elbow" pada grafik ini terlihat jelas berada pada $k = 3$, yang menunjukkan bahwa menambahkan lebih banyak kluster setelah titik tersebut hanya akan memberikan sedikit peningkatan (penurunan WSS yang kecil). Oleh karena itu, jumlah kluster yang optimal adalah 3, karena pada titik ini terdapat keseimbangan antara kompleksitas model dan kemampuan untuk menjelaskan variasi data secara efektif. Pemilihan tiga kluster ini didukung oleh perubahan signifikan dalam nilai WSS sebelum dan sesudah titik tersebut, dan dapat dianggap sebagai representasi terbaik dari struktur data dalam konteks penyebaran COVID-19 antarprovinsi di Indonesia.

K-means clustering with 3 clusters of sizes 3, 3, 26

Cluster means:

	Kasus	tingkat_kesembuhan	tingkat_kematian
1	2.8427131	-0.3975356	1.3211153
2	-0.3962699	-2.4264778	1.1621323
3	-0.2822819	0.3258477	-0.2865286

Clustering vector:

	1	2	3	4	5	6	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1	2	3	3	3	3	3	3	1	1	1	3	3	3	3	2	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3
33																															
3																															

Within cluster sum of squares by cluster:

[1] 12.125225 6.426981 15.341554
(between_SS / total_SS = 63.6 %)

Available components:

[1]	"cluster"	"centers"	"totss"	"withinss"	"tot.withinss"	"betweenss"
[7]	"size"	"iter"	"ifault"			

Hasil analisis K-Means clustering menunjukkan bahwa provinsi-provinsi di Indonesia dapat dikelompokkan ke dalam tiga klaster berdasarkan jumlah kasus COVID-19, tingkat kesembuhan, dan tingkat kematian. Klaster pertama terdiri dari tiga provinsi yang ditandai dengan jumlah kasus COVID-19 yang sangat tinggi (nilai $z = +2.84$), tingkat kesembuhan yang sedikit di bawah rata-rata (-0.40), serta tingkat kematian yang tinggi (+1.32). Provinsi dalam klaster ini dapat dianggap sebagai wilayah dengan beban pandemi yang paling berat. Klaster kedua juga terdiri dari tiga provinsi, namun memiliki karakteristik yang berbeda, yakni jumlah kasus relatif rendah (-0.40), tingkat kesembuhan yang sangat rendah (-2.43), dan tingkat kematian yang tinggi (+1.16). Hal ini mengindikasikan bahwa meskipun jumlah kasus tidak tinggi, kualitas penanganan atau kapasitas layanan kesehatan di provinsi-provinsi ini mungkin masih lemah, sehingga menyebabkan angka kematian yang tinggi. Sementara itu, klaster ketiga mencakup mayoritas provinsi (26 dari 32), dengan karakteristik jumlah kasus di bawah rata-rata (-0.28), tingkat kesembuhan yang baik (+0.33), dan tingkat kematian yang rendah (-0.29). Provinsi dalam klaster ini cenderung menunjukkan penanganan COVID-19 yang relatif efektif dan stabil. Secara keseluruhan, model ini berhasil menjelaskan sekitar 63,6% variasi data yang ada, yang menunjukkan bahwa hasil klastering cukup representatif untuk menggambarkan pola persebaran dan karakteristik COVID-19 antarprovinsi.

Berdasarkan hasil klastering K-Means dan data COVID-19 antarprovinsi yang ditampilkan dalam tabel, provinsi-provinsi di Indonesia terbagi ke dalam tiga klaster yang mencerminkan karakteristik berbeda dalam hal jumlah kasus, tingkat kesembuhan, dan tingkat kematian. Klaster 1 mencakup provinsi dengan jumlah kasus COVID-19 yang sangat tinggi, tingkat kesembuhan di bawah rata-rata, dan tingkat kematian yang tinggi. Berdasarkan vektor klastering, provinsi yang masuk ke dalam klaster ini adalah:

- DKI Jakarta, Jawa Barat, dan Jawa Tengah. Provinsi-provinsi ini merupakan wilayah padat penduduk dan menjadi episentrum penyebaran COVID-19 di Indonesia. Tingginya jumlah kasus dan kematian mencerminkan beban berat dalam penanganan pandemi di daerah urban yang memiliki mobilitas penduduk tinggi.

Klaster 2 terdiri dari provinsi dengan jumlah kasus relatif rendah, namun memiliki tingkat kesembuhan sangat rendah dan tingkat kematian yang tinggi. Provinsi dalam klaster ini adalah:

- Jawa Timur, Sumatera Utara, dan Sulawesi Selatan. Meskipun kasus tidak setinggi klaster 1, tetapi buruknya rasio kesembuhan dan tingginya tingkat kematian menunjukkan adanya tantangan besar dalam kualitas layanan kesehatan atau keterlambatan deteksi kasus.

Klaster 3 merupakan klaster terbesar, beranggotakan 26 provinsi, yang secara umum memiliki jumlah kasus lebih rendah, tingkat kesembuhan tinggi, dan tingkat kematian rendah. Contohnya antara lain:

- Aceh, Bali, Kalimantan Barat, Papua, Maluku, Riau, dan sebagainya. Provinsi dalam klaster ini menunjukkan kondisi yang lebih terkendali, baik dari segi jumlah

kasus maupun efektivitas penanganan, meskipun sebagian besar termasuk daerah dengan kepadatan penduduk yang lebih rendah.

Secara keseluruhan, klustering ini mengelompokkan provinsi-provinsi berdasarkan performa dan karakteristik penanganan COVID-19 yang dapat menjadi dasar bagi pengambilan kebijakan berbasis wilayah. Pendekatan berbeda dapat diterapkan untuk tiap klaster, sesuai dengan tingkat risiko dan kapasitas daerah. Selain itu, dibandingkan dengan metode lain seperti K-Medoids, algoritma K-Means telah terbukti memberikan hasil klaster yang lebih optimal dalam pemetaan penyebaran COVID-19 di Indonesia. (Utomo, 2021)

Analysis spasial

```
> neighbors
Neighbour list object:
Number of regions: 32
Number of nonzero links: 62
Percentage nonzero weights: 6.054688
Average number of links: 1.9375
4 regions with no links:
2, 3, 20, 21
10 disjoint connected subgraphs
> weights
Characteristics of weights list object:
Neighbour list object:
Number of regions: 32
Number of nonzero links: 62
Percentage nonzero weights: 6.054688
Average number of links: 1.9375
4 regions with no links:
2, 3, 20, 21
10 disjoint connected subgraphs

Weights style: W
Weights constants summary:
  n  nn S0      S1      S2
w 28 784 28 30.74722 120.2694
```

Sebelum melakukan analisis spasial, penting untuk memahami konsep autokorelasi spasial. Salah satu metode yang umum digunakan adalah indeks Moran's I, yang mengukur tingkat kemiripan suatu variabel antar lokasi secara global. Moran's I juga banyak digunakan dalam eksplorasi data spasial karena mampu mendeteksi pola distribusi spasial non-acak pada wilayah geografis. (Yamada, 2024). Berdasarkan Output tersebut menunjukkan hasil pembentukan spatial weights matrix berdasarkan kedekatan spasial antar provinsi di Indonesia. Terdapat 32 wilayah (provinsi) yang terdaftar dalam daftar tetangga (*neighbour list*), dari total seharusnya 34 provinsi. Ini berarti dua provinsi tidak terdeteksi atau tidak memiliki informasi spasial yang valid, kemungkinan karena data shapefile yang tidak lengkap atau kesalahan saat proses *merge*. Jumlah total hubungan spasial (non-zero links) adalah 62, dengan rata-rata setiap provinsi memiliki sekitar 1,94 tetangga.

Namun, terdapat 4 provinsi (berkode 2, 3, 20, dan 21) yang tidak memiliki tetangga (*no links*), kemungkinan besar merupakan provinsi kepulauan yang terpisah secara geografis dari daratan utama, seperti Bali, Bangka Belitung, Maluku Utara, atau Papua Barat. Selain itu, ditemukan 10 *disjoint connected subgraphs*, yaitu kelompok-kelompok wilayah yang saling terhubung tetapi tidak terkoneksi dengan kelompok lainnya, yang mencerminkan struktur spasial Indonesia sebagai negara kepulauan yang tersebar.

Gaya pembobotan yang digunakan adalah *row-standardized weights (style W)*, di mana setiap nilai bobot dibagi dengan jumlah bobot dari satu wilayah, sehingga jumlah bobot per baris menjadi 1. Nilai-nilai seperti $S_0 = 28$, $S_1 = 30.74722$, dan $S_2 = 120.2694$ digunakan dalam perhitungan statistik spasial, termasuk Moran's I dan LISA. Secara keseluruhan, output ini memberikan gambaran mengenai

bagaimana hubungan spasial antar provinsi terstruktur, yang menjadi dasar penting dalam analisis autokorelasi spasial lebih lanjut.

```
attr(,"call")
localmoran(x = indo_covid_clean$Kasus, listw = weights, zero.policy = TRUE)
attr(,"class")
[1] "localmoran" "matrix"      "array"
attr(,"quadr")
      mean      median      pysal
1   Low-Low  Low-High  Low-High
2   High-Low  High-Low  High-Low
3   Low-Low  High-Low  Low-Low
4   High-High  High-High  High-High
5   Low-Low  Low-High  Low-Low
6   Low-Low  Low-Low  Low-Low
7   Low-Low  Low-High  Low-Low
8   High-High  High-High  High-High
9   High-High  High-High  High-High
10  High-High  High-High  High-High
11  Low-Low  Low-High  Low-High
12  Low-Low  High-High  Low-High
13  Low-Low  Low-High  Low-Low
14  High-Low  High-Low  High-Low
15  Low-High  Low-High  Low-High
16  Low-Low  High-Low  Low-Low
17  Low-Low  High-Low  Low-Low
18  Low-Low  Low-Low  Low-Low
19  Low-Low  Low-Low  Low-Low
20  Low-Low  Low-Low  Low-Low
21  Low-Low  High-Low  Low-Low
22  Low-Low  Low-Low  Low-Low
23  Low-Low  Low-Low  Low-Low
24  High-Low  High-High  High-Low
25  Low-Low  Low-High  Low-Low
26  High-Low  High-Low  High-Low
27  Low-Low  Low-Low  Low-Low
28  Low-Low  Low-High  Low-Low
29  Low-Low  Low-Low  Low-Low
30  Low-Low  High-High  Low-Low
31  Low-Low  High-Low  Low-Low
32  High-Low  High-High  High-Low
> |
```

Berdasarkan hasil kuadran dari analisis Local Indicators of Spatial Association (LISA) menggunakan fungsi `localmoran()` pada variabel jumlah kasus COVID-19, diperoleh klasifikasi spasial untuk masing-masing provinsi di Indonesia ke dalam empat kuadran utama: **High-High**, **Low-Low**, **High-Low**, dan **Low-High**. Klasifikasi ini menggambarkan pola spasial lokal antara suatu wilayah dengan wilayah sekitarnya.

Beberapa provinsi teridentifikasi sebagai kluster **High-High**, yaitu wilayah dengan jumlah kasus tinggi yang dikelilingi oleh wilayah dengan jumlah kasus tinggi pula. Provinsi-provinsi ini termasuk provinsi ke-4, ke-8, ke-9, ke-10, dan ke-26. Pola ini menunjukkan adanya konsentrasi atau kluster kasus COVID-19 yang signifikan secara spasial, dan wilayah-wilayah ini bisa menjadi prioritas utama dalam upaya pengendalian dan mitigasi.

Sebaliknya, sejumlah provinsi masuk dalam kategori **Low-Low**, seperti provinsi ke-1, ke-5, ke-6, dan ke-23. Ini mengindikasikan wilayah dengan jumlah kasus rendah yang berdekatan dengan wilayah lain yang juga memiliki jumlah kasus rendah. Pola ini bisa merepresentasikan daerah yang relatif aman atau berhasil dalam pengendalian penyebaran COVID-19.

Kategori **High-Low** dan **Low-High** menunjukkan adanya anomali atau outlier spasial. Misalnya, provinsi ke-2 dan ke-3 termasuk dalam kategori High-Low dan Low-Low, sementara provinsi ke-11 dan ke-30 termasuk dalam Low-High. Ini menandakan adanya provinsi dengan jumlah kasus

tinggi yang dikelilingi oleh provinsi dengan kasus rendah (High-Low), atau sebaliknya, wilayah dengan kasus rendah dikelilingi wilayah dengan kasus tinggi (Low-High). Pola ini perlu perhatian khusus karena bisa menunjukkan potensi penyebaran atau ketidaksesuaian antarwilayah.

Secara keseluruhan, hasil kuadran LISA ini memperkuat temuan sebelumnya terkait adanya kluster spasial COVID-19 di Indonesia dan membantu mengidentifikasi wilayah yang perlu diintervensi secara spesifik berdasarkan karakteristik lingkungan sekitarnya.

```
> lisa
```

	Ii	E.Ii	Var.Ii	Z.Ii	Pr(z != E(Ii))
1	-0.013224716	-5.745390e-03	0.1827961834	-0.0174935868	9.860428e-01
2	0.000000000	0.000000e+00	0.0000000000	NaN	NaN
3	0.000000000	0.000000e+00	0.0000000000	NaN	NaN
4	0.867079669	-1.457268e-03	0.0465646142	4.0249481317	5.698787e-05
5	0.152544645	-8.776947e-03	0.0626393656	0.6445681228	5.192071e-01
6	0.242149722	-1.153626e-02	0.1763691572	0.6040670626	5.457991e-01
7	0.095087262	-7.353230e-03	0.0404860098	0.5091188386	6.106689e-01
8	5.707192010	-5.368528e-01	3.8456609116	3.1840546912	1.452276e-03
9	7.858634448	-2.156276e-01	2.6159133941	4.9921906068	5.969830e-07
10	5.169964544	-1.289874e-01	3.5951897180	2.7946606347	5.195422e-03
11	-0.002505793	-5.342079e-03	0.0821827679	0.0098937198	9.921061e-01
12	-0.001277653	-1.388820e-03	0.0214505874	0.0007590239	9.993944e-01
13	0.028547665	-4.457711e-03	0.0441811645	0.1570240277	8.752259e-01
14	-0.136977976	-4.757959e-03	0.0340943079	-0.7160713059	4.739473e-01
15	-0.169167906	-6.258812e-03	0.1990284502	-0.3651638268	7.149891e-01
16	0.150105015	-3.188518e-03	0.1017072361	0.4806709849	6.307504e-01
17	0.137355772	-3.846729e-03	0.0592672040	0.5800095811	5.619082e-01
18	0.345467383	-1.081935e-02	0.3424733288	0.6088163031	5.426462e-01
19	0.345467383	-1.147862e-02	0.3630994538	0.5923654520	5.536059e-01
20	0.000000000	0.000000e+00	0.0000000000	NaN	NaN
21	0.000000000	0.000000e+00	0.0000000000	NaN	NaN
22	0.234586051	-6.503743e-03	0.2067662319	0.5301991408	5.959739e-01
23	0.234586051	-8.804761e-03	0.2792715955	0.4605648571	6.451108e-01
24	-0.032404079	-1.130778e-03	0.0112447952	-0.2949157402	7.680582e-01
25	0.091107769	-1.142579e-02	0.1746997621	0.2453126166	8.062144e-01
26	-0.028808598	-1.069788e-04	0.0010649191	-0.8795247938	3.791168e-01
27	0.152545149	-4.268356e-03	0.0306009848	0.8964287128	3.700238e-01
28	0.082951188	-9.471542e-03	0.1451056738	0.2426255874	8.082955e-01
29	0.266766729	-6.419109e-03	0.2040929374	0.6047058838	5.453744e-01
30	0.014266461	-1.724447e-04	0.0012413877	0.4098081213	6.819467e-01
31	0.122742405	-2.420048e-03	0.0240346149	0.8073381603	4.194717e-01
32	-0.003216289	-3.167596e-05	0.0003153418	-0.1793353311	8.576744e-01

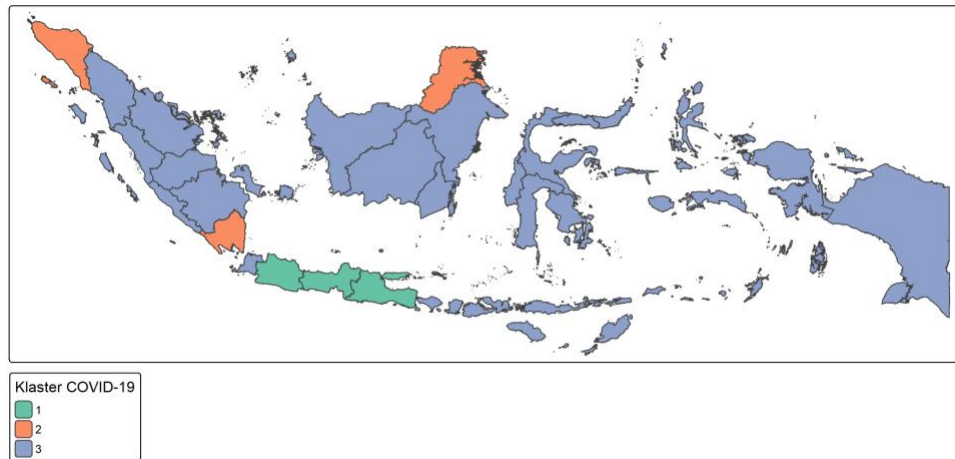
Dalam konteks data spasial, penggunaan Indeks Moran sangat penting untuk menentukan apakah pola distribusi nilai menunjukkan kecenderungan mengelompok, tersebar, atau acak. Moran's I dapat memberikan gambaran menyeluruh terhadap struktur spasial yang mendasari data multivariat, menjadikannya alat yang relevan dalam eksplorasi spasial epidemiologi. Berdasarkan hasil analisis Local Moran's I pada data kasus COVID-19 antarprovinsi di Indonesia, diperoleh nilai indeks I_i untuk masing-masing wilayah yang merepresentasikan tingkat autokorelasi spasial lokal. Nilai I_i yang positif menunjukkan adanya hubungan spasial positif, di mana provinsi dengan jumlah kasus tinggi cenderung berdekatan dengan provinsi dengan jumlah kasus tinggi pula (High-High), dan sebaliknya untuk kasus rendah (Low-Low). Sebaliknya, nilai I_i negatif menunjukkan adanya pola *outlier*, seperti provinsi dengan kasus tinggi dikelilingi wilayah dengan kasus rendah (High-Low) atau sebaliknya (Low-High).

Dari hasil tersebut, terlihat bahwa beberapa provinsi memiliki nilai I_i dan $Z.I_i$ yang signifikan secara statistik ($p\text{-value} < 0.05$), seperti provinsi ke-4, ke-8, dan ke-10. Misalnya, provinsi ke-4 memiliki nilai I_i sebesar 0.867, $Z.I_i$ sebesar 4.02, dan $p\text{-value}$ sebesar $5.69e-05$ yang menunjukkan hubungan spasial positif yang sangat signifikan, mengindikasikan adanya kluster kasus tinggi di wilayah tersebut. Provinsi ke-8 dan ke-10 juga menunjukkan pola High-High yang signifikan. Sebaliknya, sebagian besar provinsi lainnya memiliki nilai $Z.I_i$ yang tidak signifikan, menandakan tidak

adanya pola spasial yang kuat secara lokal atau hubungan spasial yang lemah antara provinsi dan tetangganya.

Selain itu, ada beberapa nilai I_i dan $Z.I_i$ yang menghasilkan NaN , seperti pada provinsi ke-2, ke-3, dan ke-20. Hal ini terjadi karena wilayah tersebut tidak memiliki tetangga spasial (neighbor) dalam definisi struktur bobot spasial yang digunakan, sehingga tidak dapat dihitung nilai autokorelasinya. Secara keseluruhan, hasil ini mengindikasikan bahwa meskipun terdapat beberapa klaster spasial signifikan, distribusi spasial kasus COVID-19 di Indonesia cenderung tidak sepenuhnya homogen, dan beberapa wilayah berperan sebagai klaster signifikan yang perlu mendapatkan perhatian khusus dalam upaya penanggulangan pandemi.

Visualisasi hasil



Visualisasi peta tematik memungkinkan representasi yang efektif dari pola spasial dan klaster, membantu pembaca untuk memahami distribusi geografis dari fenomena yang diamati. Gambar peta yang ditampilkan merupakan hasil visualisasi klaster COVID-19 di Indonesia berdasarkan metode K-Means Clustering terhadap data kasus konfirmasi, kesembuhan, dan kematian di setiap provinsi. Terdapat tiga klaster yang terbentuk, ditandai dengan warna berbeda: klaster 1 (hijau), klaster 2 (oranye), dan klaster 3 (biru).

Klaster 1 (hijau) hanya mencakup wilayah **Pulau Jawa bagian selatan**, yang terdiri dari provinsi dengan jumlah kasus yang sangat tinggi namun juga didukung oleh angka kesembuhan yang relatif tinggi. Wilayah ini, seperti Jawa Timur, Jawa Tengah, dan DIY, menunjukkan karakteristik pandemi yang berat namun dengan sistem pelayanan kesehatan yang mampu merespons.

Klaster 2 (oranye) mencakup beberapa provinsi seperti **DKI Jakarta, Sumatera Utara, Sumatera Barat, dan Papua Barat**, yang umumnya merupakan pusat-pusat aktivitas ekonomi atau wilayah dengan mobilitas tinggi. Klaster ini cenderung menunjukkan provinsi dengan kasus konfirmasi yang tinggi dan tingkat kematian yang cukup signifikan, namun angka kesembuhannya tidak setinggi klaster 1.

Sementara itu, Klaster 3 (biru) merupakan klaster terbesar yang mencakup sebagian besar provinsi lainnya di Indonesia. Provinsi-provinsi dalam klaster ini umumnya memiliki jumlah kasus konfirmasi yang lebih rendah dibanding dua klaster lainnya, yang mungkin mencerminkan keterbatasan dalam pelaporan atau testing, ataupun penyebaran kasus yang memang lebih terkendali.

Secara keseluruhan, peta ini menunjukkan adanya heterogenitas spasial dalam penyebaran COVID-19 di Indonesia, yang kemungkinan dipengaruhi oleh berbagai faktor seperti kepadatan penduduk, kapasitas sistem kesehatan, mobilitas masyarakat, serta kebijakan lokal yang diterapkan selama pandemi. Hasil ini dapat digunakan sebagai dasar pertimbangan dalam penentuan prioritas wilayah dalam penanggulangan pandemi COVID-19.

SIMPULAN

Penelitian ini berhasil menganalisis persebaran COVID-19 di Indonesia dengan menggabungkan metode K-Means Clustering dan Exploratory Spatial Data Analysis (ESDA). Hasil klastering menunjukkan bahwa provinsi-provinsi di Indonesia dapat dikelompokkan menjadi tiga

klaster berdasarkan karakteristik jumlah kasus, tingkat kesembuhan, dan tingkat kematian. Klaster 1 terdiri dari provinsi dengan kasus sangat tinggi dan tingkat kematian tinggi (misalnya DKI Jakarta, Jawa Barat, dan Jawa Tengah), Klaster 2 mencakup provinsi dengan kasus rendah tetapi tingkat kesembuhan sangat rendah dan kematian tinggi (seperti Jawa Timur dan Sulawesi Selatan), sedangkan Klaster 3 didominasi oleh provinsi dengan kasus rendah, kesembuhan tinggi, dan kematian rendah (sebagian besar provinsi di luar Jawa). Analisis spasial dengan ESDA mengungkapkan adanya autokorelasi spasial yang signifikan, terutama di wilayah-wilayah tertentu seperti Jawa dan Sumatera, yang membentuk klaster High-High (kasus tinggi dikelilingi kasus tinggi). Temuan ini menunjukkan bahwa penyebaran COVID-19 di Indonesia tidak merata dan dipengaruhi oleh faktor geografis, mobilitas penduduk, serta kapasitas layanan kesehatan. Hasil penelitian dapat menjadi dasar bagi pemerintah dalam merancang kebijakan penanganan pandemi yang berbasis wilayah, seperti intervensi khusus untuk daerah dengan risiko tinggi dan peningkatan kapasitas kesehatan di daerah yang rentan. Kombinasi pendekatan klastering dan analisis spasial ini juga membuktikan pentingnya integrasi aspek geografis dalam studi epidemiologi untuk pemahaman yang lebih holistik.

REFERENSI