

ANALISIS KASUS COVID-19 DI KABUPATEN JEMBER DENGAN METODE K-MEANS DAN REGRESI

Ika Arofatul Hidayah¹

¹UIN Kiai Haji Achamd Siddiq Jember, Indonesia

Email : ikaarofatul@lecturer.uinkhas.ac.id

ABSTRAK

Pandemi Coronavirus Disease 2019 (COVID-19) telah menimbulkan dampak yang luas terhadap sektor kesehatan dan memerlukan analisis berbasis data untuk mendukung pengambilan kebijakan di tingkat daerah. Penelitian ini bertujuan untuk menganalisis pola persebaran COVID-19 serta mengidentifikasi pengaruh jumlah penduduk dan jumlah kasus suspek terhadap jumlah kasus positif COVID-19 di Kabupaten Jember. Penelitian menggunakan pendekatan kuantitatif dengan data sekunder dari 31 kecamatan di Kabupaten Jember yang diperoleh dari website resmi Siaga COVID-19 Pemerintah Kabupaten Jember. Analisis dilakukan menggunakan metode K-Means Clustering untuk mengelompokkan wilayah berdasarkan tingkat penyebaran COVID-19 dan Regresi Linier Berganda untuk menguji pengaruh jumlah penduduk dan jumlah kasus suspek terhadap jumlah kasus positif. Hasil penelitian menunjukkan bahwa terbentuk tiga kluster, yaitu 17 kecamatan pada kluster risiko rendah, 12 kecamatan pada kluster risiko sedang, dan 2 kecamatan (Kaliwates dan Sumbersari) pada kluster risiko tinggi. Hasil regresi linier berganda menghasilkan persamaan $Y = -1,81 + 8,27X_1 + 2,46X_2$, dengan variabel jumlah kasus suspek berpengaruh signifikan terhadap jumlah kasus positif ($p\text{-value} = 0,0001$), sedangkan jumlah penduduk tidak berpengaruh signifikan ($p\text{-value} = 0,34$). Nilai Adjusted R^2 sebesar 0,6024 menunjukkan bahwa model mampu menjelaskan 60,24% variasi jumlah kasus positif COVID-19. Integrasi metode K-Means dan Regresi Linier Berganda memberikan informasi spasial dan statistik yang komprehensif sehingga dapat dimanfaatkan sebagai dasar pengambilan kebijakan dalam pengendalian penyebaran COVID-19 di Kabupaten Jember

Kata Kunci : COVID-19, K-Means Clustering, Regresi Linier Berganda, Quantum GIS, Kabupaten Jember.

ABSTRACT

The Coronavirus Disease 2019 (COVID-19) pandemic has had a significant impact on the public health sector and requires data-driven analysis to support evidence-based policymaking at the regional level. This study aims to analyze the spatial distribution of COVID-19 and examine the influence of population size and the number of suspected COVID-19 cases on the number of confirmed positive cases in Jember Regency, Indonesia. A quantitative approach was employed using secondary data collected from 31 sub-districts in Jember Regency, obtained from the official COVID-19 Monitoring website of the Jember Regency Government. The analysis was conducted using the K-Means Clustering method to classify sub-districts based on the level of COVID-19 transmission and Multiple Linear Regression to examine the effects of population size and suspected COVID-19 cases on the number of confirmed positive cases. The results show that three clusters were identified, consisting of 17 sub-districts in the low-risk cluster, 12 sub-districts in the medium-risk cluster, and 2 sub-districts (Kaliwates and Sumbersari) in the high-risk cluster. The multiple linear regression analysis produced the equation $Y = -1.81 + 8.27X_1 + 2.46X_2$, where the number of suspected COVID-19 cases had a significant effect on the number of confirmed positive cases ($p\text{-value} = 0.0001$), while population size had no significant effect ($p\text{-value} = 0.34$). The Adjusted R^2 value of 0.6024 indicates that the model explains 60.24% of the variation in confirmed positive COVID-19 cases. The integration of K-Means Clustering and Multiple Linear Regression provides comprehensive spatial and statistical information that can support evidence-based decision-making for controlling the spread of COVID-19 in Jember Regency

Keywords: COVID-19, K-Means Clustering, Multiple Linear Regression, Quantum GIS, Jember Regency.

PENDAHULUAN

Kesehatan merupakan aspek fundamental dalam kehidupan manusia karena menentukan kemampuan individu untuk menjalankan aktivitas sosial, ekonomi, dan produktif secara optimal. Salah satu tantangan terbesar terhadap sektor kesehatan global dalam beberapa tahun terakhir adalah pandemi Coronavirus Disease 2019 (COVID-19) yang disebabkan oleh Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2). Virus ini pertama kali diidentifikasi di Kota Wuhan, Provinsi Hubei, Tiongkok, pada akhir Desember 2019 dan dengan cepat menyebar ke berbagai negara hingga Organisasi Kesehatan Dunia (WHO) menetapkan sebagai pandemi global pada 11 Maret 2020 [1]-[2]. COVID-19 menyerang sistem pernapasan dengan tingkat keparahan yang bervariasi, mulai dari tanpa gejala hingga kondisi berat yang dapat menyebabkan kematian, sehingga menimbulkan dampak yang luas terhadap sistem kesehatan, mobilitas masyarakat, serta aktivitas ekonomi [3].

Di Indonesia, pandemi COVID-19 menunjukkan pola penyebaran yang berbeda pada setiap wilayah, dipengaruhi oleh karakteristik demografi, mobilitas penduduk, kepadatan permukiman, serta kapasitas layanan kesehatan. Kepadatan penduduk yang sering dikaitkan dengan tingkat penularan lebih tinggi di daerah perkotaan, apabila dikombinasikan dengan variasi distribusi tenaga kesehatan, akses terhadap sanitasi, serta kondisi lingkungan, dapat berperan penting dalam membentuk dinamika penyebaran dan angka kematian akibat COVID-19 di suatu wilayah [4]. Sejalan dengan hal tersebut, kajian multivariat mengenai hubungan dan perbedaan kasus COVID-19 antarwilayah turut menegaskan bahwa penyebaran virus ini dipengaruhi oleh kepadatan penduduk, mobilitas masyarakat, serta efektivitas kebijakan pemerintah dalam mengendalikan pandemi [5].

Salah satu daerah yang mengalami peningkatan kasus secara signifikan adalah Kabupaten Jember, Provinsi Jawa Timur. Kasus konfirmasi positif pertama di Kabupaten Jember tercatat pada 9 April 2020 di Kecamatan Kaliwates, kemudian berkembang dengan cepat ke berbagai kecamatan lain sehingga memerlukan kebijakan penanganan yang terarah dan berbasis data [6]-[7]. Kondisi tersebut menunjukkan bahwa pemerintah daerah memerlukan informasi yang akurat mengenai pola persebaran kasus dan faktor-faktor yang memengaruhi peningkatan jumlah kasus positif, agar strategi mitigasi, alokasi sumber daya kesehatan, dan penentuan prioritas penanganan dapat dilakukan secara lebih efektif.

Dalam konteks tersebut, analisis data berbasis data mining dan statistik menjadi pendekatan yang penting untuk mendukung pengambilan keputusan. Salah satu metode yang banyak digunakan adalah K-Means Clustering [8], yaitu algoritma pengelompokan nonhierarkis yang mampu mengelompokkan objek berdasarkan tingkat kemiripan karakteristiknya [9]. Berbagai penelitian menunjukkan bahwa metode K-Means efektif digunakan untuk mengelompokkan wilayah berdasarkan tingkat risiko penyebaran COVID-19, sehingga menghasilkan kategori risiko rendah, sedang, dan tinggi yang dapat dimanfaatkan dalam proses pemetaan wilayah terdampak [3]. Selain itu, analisis regresi linier berganda juga telah digunakan secara luas untuk memodelkan hubungan antara jumlah kasus COVID-19 dengan berbagai variabel prediktor, seperti kepadatan penduduk, jumlah penduduk, maupun jumlah kasus suspek, dengan kemampuan yang baik dalam menjelaskan variasi data berdasarkan nilai koefisien determinasi (R^2) [10].

Meskipun penelitian mengenai K-Means dan regresi linier pada data COVID-19 telah banyak dilakukan, sebagian besar penelitian terdahulu masih menerapkan kedua metode tersebut secara terpisah. Penelitian clustering umumnya hanya berfokus pada pengelompokan wilayah berdasarkan jumlah kasus, sedangkan penelitian regresi lebih menitikberatkan pada pemodelan hubungan antarvariabel tanpa mengintegrasikan informasi spasial hasil pengelompokan wilayah. Selain itu, penelitian yang mengombinasikan visualisasi spasial menggunakan Quantum GIS dengan implementasi analisis pada perangkat lunak R pada tingkat kecamatan di Kabupaten Jember masih relatif terbatas. Kondisi ini menunjukkan adanya kesenjangan penelitian (research gap) dalam penyediaan model analisis yang mampu menggabungkan pemetaan persebaran wilayah dan identifikasi faktor-faktor yang memengaruhi peningkatan kasus secara simultan.

Berdasarkan kesenjangan tersebut, penelitian ini menawarkan kebaruan berupa integrasi metode K-Means Clustering dan Regresi Linier Berganda untuk menganalisis data COVID-19 pada 31 kecamatan di Kabupaten Jember. Integrasi kedua metode tersebut tidak hanya menghasilkan pengelompokan wilayah berdasarkan tingkat risiko penyebaran, tetapi juga mengidentifikasi pengaruh jumlah penduduk dan jumlah kasus suspek terhadap peningkatan jumlah kasus positif COVID-19. Proses analisis dilakukan menggunakan perangkat lunak R, sedangkan hasil pengelompokan divisualisasikan dalam bentuk peta menggunakan Quantum GIS sehingga menghasilkan informasi yang lebih komprehensif dan mudah diinterpretasikan.

Tujuan penelitian ini adalah: (1) mengelompokkan kecamatan di Kabupaten Jember berdasarkan karakteristik penyebaran COVID-19 menggunakan metode K-Means Clustering; (2) menganalisis pengaruh jumlah penduduk dan jumlah kasus suspek terhadap jumlah kasus positif COVID-19 menggunakan Regresi Linier Berganda; dan (3) menghasilkan visualisasi spasial yang dapat mendukung proses pemetaan wilayah terdampak. Kontribusi penelitian ini diharapkan dapat memberikan model analisis terpadu berbasis data mining, statistik, dan sistem informasi geografis yang dapat dimanfaatkan sebagai dasar pengambilan kebijakan dalam upaya penanganan dan pengendalian penyebaran COVID-19 di tingkat daerah.

METODE PENELITIAN

1. Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan desain analisis deskriptif dan eksplanatori. Pendekatan kuantitatif digunakan karena penelitian bertujuan mengolah data numerik untuk mengidentifikasi pola persebaran COVID-19 dan mengukur hubungan antara variabel-variabel yang memengaruhi peningkatan jumlah kasus positif. Analisis dilakukan dengan mengombinasikan metode K-Means Clustering untuk pengelompokan wilayah dan Regresi Linier Berganda untuk menguji pengaruh variabel jumlah penduduk dan jumlah kasus suspek terhadap jumlah kasus positif COVID-19 [11]-[12].

2. Sumber Data dan Karakteristik Data

Data yang digunakan merupakan data sekunder yang diperoleh dari website resmi Siaga COVID-19 Pemerintah Kabupaten Jember. Data merepresentasikan kondisi kumulatif kasus COVID-19 pada 31 kecamatan di Kabupaten Jember [13]. Dengan demikian, jumlah observasi yang dianalisis adalah 31 observasi, yaitu satu observasi untuk setiap kecamatan.

Variabel yang digunakan dalam penelitian ini terdiri atas tiga variabel, yaitu:

- Y (Jumlah Kasus Positif COVID-19): jumlah kumulatif kasus positif COVID-19 pada setiap kecamatan.
- X1 (Jumlah Penduduk): jumlah penduduk pada masing-masing kecamatan.
- X2 (Jumlah Kasus Suspek COVID-19): jumlah kasus suspek COVID-19 pada setiap kecamatan.

Seluruh variabel berbentuk data numerik sehingga dapat digunakan dalam analisis clustering maupun regresi linier berganda.

Tabel 2. Variabel Penelitian

Variabel	Jenis	Satuan
Jumlah Kasus Positif Covid-19 (Y)	Dependen	Orang
Jumlah Penduduk (X1)	Independen	Orang
Jumlah Kasus Suspek COVID-19 (X2)	Independen	Orang

3. Tahap Preprocessing Data

Sebelum dilakukan analisis, data penelitian terlebih dahulu melalui tahap preprocessing untuk memastikan kualitas dan kesiapan data sebelum diterapkan pada metode K-Means Clustering dan Regresi Linier Berganda. Tahap preprocessing meliputi pemeriksaan kelengkapan data untuk memastikan tidak terdapat missing value pada seluruh variabel, pemeriksaan konsistensi data guna memverifikasi kesesuaian nilai antarvariabel pada setiap kecamatan, serta proses normalisasi data menggunakan fungsi `scale()` pada perangkat lunak R. Normalisasi dilakukan karena variabel jumlah penduduk memiliki rentang nilai yang jauh lebih besar dibandingkan variabel jumlah kasus suspek dan jumlah kasus positif COVID-19, sehingga penyamaan skala diperlukan untuk mencegah dominasi variabel tertentu dalam proses pembentukan klaster dan menghasilkan pengelompokan yang lebih akurat serta representative.

4. Metode K-Means Clustering

K-means clustering merupakan sebuah metode clustering analisis non hirarki yang berguna untuk mempartisi objek yang ada kedalam satu atau lebih klaster atau kelompok objek berdasarkan karakteristiknya, sehingga objek yang mempunyai karakteristik yang sama dikelompokkan dalam satu klaster yang sama dan objek yang mempunyai karakteristik yang berbeda dikelompokkan kedalam klaster yang lain [14].

Langkah-langkah algoritma K-Means adalah sebagai berikut:

- Tentukan jumlah cluster
- Alokasikan data kedalam cluster secara random
- Hitung centroid/rata-rata dari data yang ada di masing-masing cluster dan Alokasikan masing-masing data ke centroid/rata-rata terdekat berikut:

$$D(a, b) = \sqrt{\sum_{i=1}^n (b_i - a_i)^2}$$

d. Kembali ke langkah (c), apabila masih ada data yang berpindah cluster atau apabila ada perubahan nilai centroid.

5. Regresi Linier Berganda

Regresi linear ganda (multiple linear regression) adalah model regresi linear dengan 1 variabel dependen kontinu beserta k (dua atau lebih) variabel independen kontinu dan kategorik [15].

Model untuk regresi linear ganda yaitu:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i$$

Perintah Stata untuk *fitting* regresi linear ganda dengan prediktor kontinu adalah

regress var_dep vars_indep [if] [in] [, options]

var_dep : Variabel dependen

vars_indep : Himpunan variabel independen kontinu

$B_0, B_1, B_2, \dots, B_k$ adalah nilai – nilai parameter yang akan diestimasi dengan perintah Stata tersebut. Sebagai keluaran akan diperoleh estimasi persamaan garis regresi linier ganda :

$$\hat{Y}_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \dots + b_k X_{ki} ;$$

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

HASIL DAN PEMBAHASAN

1. HASIL

Analisis data dilakukan menggunakan perangkat lunak R untuk proses K-Means Clustering dan Regresi Linier Berganda, sedangkan visualisasi spasial hasil clustering dilakukan menggunakan Quantum GIS (QGIS). Sebelum proses clustering dilakukan, data dinormalisasi menggunakan fungsi `scale()` agar seluruh variabel memiliki skala yang sebanding. Selanjutnya, algoritma K-Means diterapkan dengan jumlah klaster ($k = 3$) dan parameter `nstart = 19` untuk memperoleh hasil pengelompokan yang lebih stabil.

Implementasi K-Means pada perangkat lunak R dilakukan menggunakan kode program berikut.

Berikut coding clustering K-Means pada R Studio;

```
#membuat clustering
```

```
Data2= Data [2:4]
```

```
Data2
```

```
datafix<- scale (Data2)
```

```
datafix
```

```
clustering <- kmeans (datafix, centers = 3, nstart = 19)
```

```
clustering
```

```
View(hasil)
```

Hasil proses clustering menghasilkan pengelompokan 31 kecamatan di Kabupaten Jember ke dalam tiga klaster berdasarkan karakteristik jumlah penduduk, jumlah kasus suspek, dan jumlah kasus positif COVID-19. Hasil pengelompokan ditunjukkan pada Tabel 2.

Tabel 2. Hasil Clustering K-Means Kabupaten Jember

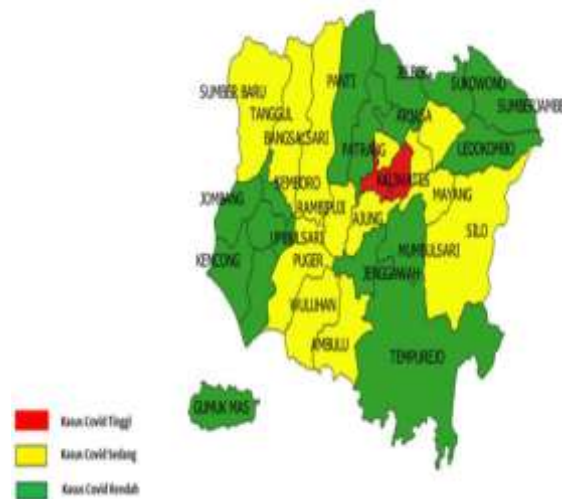
Kecamatan	Penduduk	Suspek	Positif	Clustering-cluster
Kencong	65.173	26	31	1
Gumungkmass	79.224	29	13	1
Puger	114.506	46	43	2
Wuluhan	114.695	69	43	2
Ambulu	105.103	55	72	2
Tempurejo	70.663	52	27	1
Silo	103.850	76	12	2

Mayang	48.362	27	3	1
Mumbulsari	62.339	26	13	1
Jenggawah	81.318	24	31	1
Ajung	74.416	55	51	2
Rambipuji	78.934	50	26	2
Balung	77.005	69	60	2
Umbulsari	69.539	26	28	1
Semboro	43.475	11	25	1
Jombang	50.003	11	38	1
Sumber Baru	99.416	26	28	2
Tanggul	82.760	55	51	2
Bangsalsari	113.905	34	27	2
Panti	59.399	39	20	1
Sukorambi	37.950	21	19	1
Arjasa	38.055	15	24	1
Pakusari	41.713	14	12	1
Kalisat	74.962	111	35	2
Ledokombo	62.528	70	12	1
Sumber Jambe	60.126	30	4	1
Sukowono	58.734	32	6	1
Jelbuk	31.962	8	9	1
Kaliwates	111.861	124	181	3
Sumbersari	126.279	169	181	3
Patrang	94.471	74	104	2

Berdasarkan hasil clustering, Cluster 1 terdiri atas 17 kecamatan, Cluster 2 terdiri atas 12 kecamatan, dan Cluster 3 terdiri atas 2 kecamatan, yaitu Kaliwates dan Summersari. Selanjutnya hasil clustering divisualisasikan dalam bentuk peta wilayah Kabupaten Jember menggunakan Quantum GIS dengan kategori warna hijau (Cluster 1), kuning (Cluster 2), dan merah (Cluster 3).

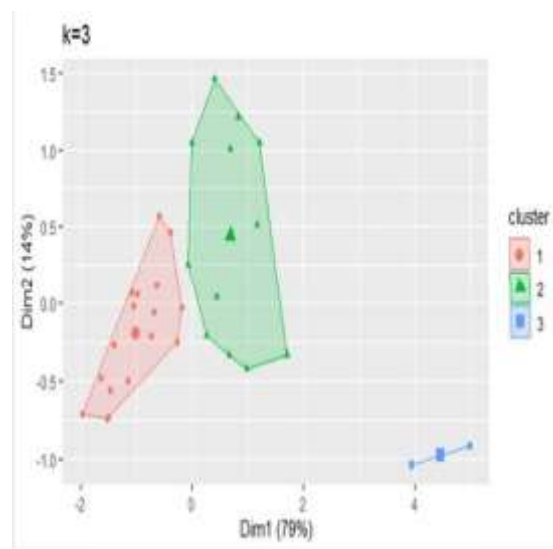
Setelah mendapatkan hasil pengklasteran pada masing-masing kecamatan, kemudian melakukan visualisasi dalam bentuk map atau peta wilayah untuk seluruh kecamatan pada kabupaten jember dengan menggunakan Quantum GIS dengan ketentuan sebagai berikut:

Klaster 3 adalah merah jumlah positif covid-19 dengan ketentuan tinggi; Klaster 2 adalah kuning jumlah positif covid-19 dengan ketentuan sedang; Klaster 1 adalah hijau jumlah positif covid-19 dengan ketentuan rendah:



Gambar 3. Flowchart regresi linier

Grafik plot menunjukkan pemisahan antar kluster yang terbentuk berdasarkan karakteristik masing-masing kecamatan. Visualisasi ini memperlihatkan bahwa Kecamatan Kaliwates dan Sumberasari berada pada kelompok dengan nilai variabel yang relatif paling tinggi dibandingkan kecamatan lainnya.



Gambar 4. Grafik plot Clustering sebaran covid-19 kabupaten Jember

Analisis Regresi Linier Berganda menggunakan perangkat lunak R dengan variabel dependen jumlah kasus positif COVID-19 serta variabel independen jumlah penduduk dan jumlah kasus suspek.

```
modelreg <- lm(Positif ~ Suspek + Penduduk, data = Data) summary(modelreg)
```

Hasil analisis regresi menghasilkan persamaan sebagai berikut:

$$Y = -1,81 + 8,27X_1 + 2,46X_2$$

dengan:

- Y = jumlah kasus positif COVID-19,
- X₁ = jumlah kasus suspek COVID-19,
- X₂ = jumlah penduduk.

Hasil uji signifikansi menunjukkan bahwa variabel jumlah kasus suspek memiliki nilai p-value = 0,0001 (< 0,05) sehingga berpengaruh signifikan terhadap jumlah kasus positif COVID-19. Sementara itu, variabel jumlah penduduk memiliki nilai p-value = 0,34 (> 0,05) sehingga tidak berpengaruh signifikan terhadap jumlah kasus positif COVID-19.

Nilai Adjusted R Square sebesar 0,6024 menunjukkan bahwa 60,24% variasi jumlah kasus positif COVID-19 dapat dijelaskan oleh jumlah kasus suspek dan jumlah penduduk, sedangkan 39,76% sisanya dipengaruhi oleh faktor-faktor lain di luar model penelitian.

2. PEMBAHASAN

a. Interpretasi Hasil K-Means Clustering

Hasil K-Means Clustering menunjukkan bahwa penyebaran COVID-19 di Kabupaten Jember tidak terjadi secara merata, melainkan membentuk tiga kelompok wilayah dengan karakteristik risiko yang berbeda. Berdasarkan hasil pengelompokan, 17 kecamatan termasuk ke dalam Cluster 1 (risiko rendah), 12 kecamatan termasuk Cluster 2 (risiko sedang), dan 2 kecamatan, yaitu Kaliwates dan Sumbersari, termasuk Cluster 3 (risiko tinggi). Pengelompokan ini menunjukkan bahwa sebagian besar wilayah Kabupaten Jember berada pada kategori penyebaran yang relatif rendah, sedangkan pusat penyebaran tertinggi terkonsentrasi pada wilayah perkotaan yang memiliki jumlah penduduk dan jumlah kasus suspek lebih besar.

Temuan tersebut mengindikasikan bahwa metode K-Means Clustering mampu mengidentifikasi wilayah dengan karakteristik epidemiologi yang homogen berdasarkan kemiripan nilai variabel yang dianalisis. Hal ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa K-Means efektif digunakan untuk mengelompokkan wilayah administratif berdasarkan tingkat penyebaran COVID-19 sehingga menghasilkan kategori risiko rendah, sedang, dan tinggi yang dapat dimanfaatkan sebagai dasar pemetaan wilayah terdampak [3]-[16]. Kemampuan K-Means dalam mempartisi data ke dalam kelompok yang memiliki tingkat kemiripan tinggi menjadikan metode ini relevan untuk analisis spasial berbasis data epidemiologi [14].

Proses normalisasi data sebelum clustering juga berkontribusi terhadap kualitas hasil pengelompokan. Variabel jumlah penduduk memiliki rentang nilai yang jauh lebih besar dibandingkan jumlah kasus suspek dan jumlah kasus positif sehingga tanpa normalisasi dapat mendominasi proses pembentukan centroid. Penggunaan fungsi `scale()` pada perangkat lunak R menghasilkan penyamaan skala antarvariabel sehingga proses clustering menjadi lebih representatif dan stabil [17]-[18]. Selain itu, penggunaan parameter `nstart = 19` meningkatkan peluang diperolehnya solusi global yang lebih baik karena algoritma melakukan inisialisasi centroid secara berulang [18]-[19].

b. Analisis Persebaran Spasial COVID-19 Kabupaten Jember

Kesehatan merupakan aspek fundamental dalam kehidupan manusia karena menentukan kemampuan individu untuk menjalankan aktivitas sosial, ekonomi, dan produktif secara optimal. Salah satu tantangan terbesar terhadap sektor kesehatan global dalam beberapa tahun terakhir adalah pandemi Coronavirus Disease 2019 (COVID-19) yang disebabkan oleh Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2). Virus ini pertama kali diidentifikasi di Kota Wuhan, Provinsi Hubei, Tiongkok, pada akhir Desember 2019 dan dengan cepat menyebar ke berbagai negara hingga Organisasi Kesehatan Dunia (WHO) menetapkan sebagai pandemi global pada 11 Maret 2020 [1]-[2]. COVID-19 menyerang sistem pernapasan dengan tingkat keparahan yang bervariasi, mulai dari tanpa gejala hingga kondisi berat yang dapat menyebabkan kematian, sehingga menimbulkan dampak yang luas terhadap sistem kesehatan, mobilitas masyarakat, serta aktivitas ekonomi [3].

Visualisasi hasil clustering dalam bentuk peta menggunakan Quantum GIS memperlihatkan bahwa Kaliwates dan Sumbersari berada pada kategori penyebaran tertinggi (Cluster 3), sedangkan wilayah lain tersebar pada kategori sedang dan rendah. Kedua kecamatan tersebut merupakan kawasan perkotaan yang berfungsi sebagai pusat pemerintahan, perdagangan, pendidikan, dan aktivitas ekonomi di Kabupaten Jember sehingga memiliki intensitas mobilitas penduduk yang relatif lebih tinggi dibandingkan kecamatan lainnya.

Pola ini memperkuat dugaan bahwa wilayah dengan tingkat interaksi sosial dan aktivitas ekonomi yang tinggi cenderung memiliki risiko penyebaran COVID-19 yang lebih besar. Hasil penelitian ini konsisten dengan berbagai studi epidemiologi yang menunjukkan bahwa visualisasi spasial berbasis clustering mampu mengidentifikasi pola geografis penyebaran COVID-19 dan membantu pemerintah dalam menentukan prioritas wilayah penanganan [20]-[21]. Dengan demikian, peta hasil clustering tidak hanya berfungsi sebagai representasi visual, tetapi juga sebagai alat bantu pengambilan keputusan untuk alokasi sumber daya kesehatan, penentuan wilayah prioritas, dan penyusunan strategi mitigasi pandemi.

Grafik plot hasil clustering juga menunjukkan pemisahan yang cukup jelas antar kelompok data, terutama pada Kecamatan Kaliwates dan Sumbersari yang memiliki nilai variabel jauh lebih tinggi dibandingkan kecamatan lainnya. Visualisasi menggunakan perangkat lunak R memberikan representasi statistik yang lebih mudah dipahami sehingga mempermudah interpretasi karakteristik masing-masing klaster [22]-[23].

c. Pengaruh Jumlah Penduduk dan Kasus Suspek terhadap Kasus Positif COVID-19

Hasil analisis Regresi Linier Berganda menunjukkan bahwa jumlah kasus suspek memiliki pengaruh yang signifikan terhadap jumlah kasus positif COVID-19 dengan nilai $p\text{-value} = 0,0001 (< 0,05)$. Persamaan regresi yang diperoleh, yaitu:

$$Y = -1,81 + 8,27X_1 + 2,46X_2$$

Menunjukkan bahwa setiap peningkatan satu kasus suspek diikuti oleh peningkatan sekitar 8,27 kasus positif, dengan asumsi variabel lain tetap. Temuan ini mengindikasikan bahwa jumlah kasus suspek merupakan indikator yang kuat dalam menjelaskan perkembangan kasus positif COVID-19 di Kabupaten Jember.

Sebaliknya, variabel jumlah penduduk tidak menunjukkan pengaruh yang signifikan terhadap jumlah kasus positif COVID-19 karena memiliki nilai $p\text{-value} = 0,34 (> 0,05)$. Hasil ini menunjukkan bahwa besarnya jumlah penduduk tidak

selalu berbanding lurus dengan peningkatan jumlah kasus positif. Penyebaran COVID-19 lebih dipengaruhi oleh faktor-faktor lain seperti mobilitas masyarakat, kepadatan permukiman, intensitas interaksi sosial, kepatuhan terhadap protokol kesehatan, dan efektivitas sistem deteksi kasus.

Nilai Adjusted R Square sebesar 0,6024 menunjukkan bahwa 60,24% variasi jumlah kasus positif COVID-19 dapat dijelaskan oleh jumlah kasus suspek dan jumlah penduduk, sedangkan 39,76% sisanya dipengaruhi oleh faktor-faktor lain di luar model penelitian. Nilai tersebut menunjukkan bahwa model regresi memiliki kemampuan penjelasan yang cukup baik dalam memodelkan hubungan antarvariabel [24]. Temuan ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa regresi linier berganda mampu menjelaskan hubungan antara variabel epidemiologi dan jumlah kasus COVID-19 dengan tingkat akurasi yang cukup tinggi berdasarkan nilai koefisien determinasi (R^2) [15]-[25].

d. Integrasi K-Means Clustering dan Regresi Linier Berganda

Kontribusi utama penelitian ini terletak pada integrasi metode K-Means Clustering dan Regresi Linier Berganda dalam satu kerangka analisis. K-Means digunakan untuk mengidentifikasi pola spasial penyebaran COVID-19, sedangkan regresi linier berganda digunakan untuk mengidentifikasi faktor-faktor yang memengaruhi peningkatan jumlah kasus positif. Integrasi kedua metode tersebut menghasilkan informasi yang lebih komprehensif dibandingkan penggunaan masing-masing metode secara terpisah.

Hasil penelitian menunjukkan bahwa wilayah yang termasuk Cluster 3 juga merupakan wilayah dengan jumlah kasus suspek dan jumlah kasus positif yang relatif tinggi, sehingga terdapat keterkaitan antara hasil pengelompokan spasial dan hasil analisis regresi. Dengan demikian, pendekatan integratif ini mampu memberikan dua jenis informasi sekaligus, yaitu pemetaan wilayah berisiko dan identifikasi faktor penyebab peningkatan kasus. Temuan tersebut menjawab kesenjangan penelitian yang telah diidentifikasi pada bagian pendahuluan, yaitu masih terbatasnya penelitian yang mengombinasikan analisis clustering, regresi linier, dan visualisasi spasial menggunakan R dan Quantum GIS pada tingkat kecamatan di Kabupaten Jember.

Secara praktis, model analisis yang dihasilkan dapat digunakan oleh Pemerintah Kabupaten Jember sebagai dasar dalam menentukan prioritas wilayah pengendalian COVID-19, distribusi sumber daya kesehatan, pelaksanaan surveilans epidemiologi, serta evaluasi efektivitas kebijakan penanganan pandemi. Oleh karena itu, penelitian ini memberikan kontribusi pada pengembangan pendekatan data-driven decision making dalam bidang kesehatan masyarakat, khususnya untuk analisis penyebaran penyakit menular berbasis data mining, statistik, dan sistem informasi geografis.

SIMPULAN

Penelitian ini berhasil menerapkan metode K-Means Clustering dan Regresi Linier Berganda untuk menganalisis persebaran serta faktor-faktor yang memengaruhi peningkatan kasus positif COVID-19 di Kabupaten Jember. Hasil K-Means Clustering menunjukkan bahwa 31 kecamatan di Kabupaten Jember dapat dikelompokkan ke dalam tiga kluster tingkat risiko, yaitu 17 kecamatan pada Cluster 1 (risiko rendah), 12 kecamatan pada Cluster 2 (risiko sedang), dan 2 kecamatan, yaitu Kaliwates dan Sumbersari, pada Cluster 3 (risiko tinggi). Hasil visualisasi spasial menggunakan Quantum GIS memperlihatkan bahwa penyebaran kasus COVID-19 terkonsentrasi pada kawasan perkotaan yang memiliki tingkat aktivitas dan mobilitas masyarakat yang relatif lebih tinggi.

Analisis Regresi Linier Berganda menunjukkan bahwa jumlah kasus suspek COVID-19 berpengaruh signifikan terhadap jumlah kasus positif COVID-19, sedangkan jumlah penduduk tidak berpengaruh signifikan. Model regresi menghasilkan Adjusted R^2 sebesar 0,6024, yang berarti 60,24% variasi jumlah kasus positif COVID-19 dapat dijelaskan oleh variabel jumlah kasus suspek dan jumlah penduduk, sedangkan 39,76% sisanya dipengaruhi oleh faktor-faktor lain di luar model penelitian. Dengan demikian, integrasi metode K-Means Clustering dan Regresi Linier Berganda mampu menghasilkan informasi spasial dan statistik yang saling melengkapi, sehingga dapat dimanfaatkan sebagai dasar dalam penentuan prioritas wilayah penanganan, distribusi sumber daya kesehatan, serta penyusunan strategi pengendalian penyebaran COVID-19 di Kabupaten Jember.

UCAPAN TERIMA KASIH

Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas rahmat dan karunia-Nya sehingga penelitian berjudul "Analisis Kasus COVID-19 di Kabupaten Jember dengan Metode K-Means dan Regresi" ini dapat terselesaikan dan dipublikasikan pada DIDAKTIKA: Jurnal Pengabdian Masyarakat. Penulis menyampaikan ucapan terima kasih dan penghargaan yang setinggi-tingginya kepada:

1. Pimpinan UIN Kiai Haji Achmad Siddiq Jember, atas dukungan fasilitas dan iklim akademik yang sangat kondusif selama pelaksanaan penelitian ini.
2. Pemerintah Kabupaten Jember dan Tim Satuan Tugas Siaga COVID-19 Kabupaten Jember, atas ketersediaan data sekunder secara terbuka pada situs resminya sehingga penelitian berbasis data 31 kecamatan di Kabupaten Jember ini dapat dilaksanakan secara akurat.
3. Redaksi dan Dewan Reviewer DIDAKTIKA: Jurnal Pengabdian Masyarakat, atas arahan, koreksi, dan masukan konstruktif demi kesempurnaan naskah artikel ini.

4. Rekan-rekan Dosen dan Civitas Academica UIN KHAS Jember, atas masukan ilmiah serta diskusi kritis selama proses analisis statistik menggunakan R Studio dan pemetaan visualisasi spasial menggunakan Quantum GIS (QGIS).
5. Keluarga dan Seluruh Pihak yang Terlibat, yang memberikan dukungan moral dan doa hingga penelitian ini dapat diselesaikan dengan baik.

Daftar Pustaka

- [1] N. Chen *et al.*, “Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study,” *Lancet*, pp. 507–513, 2020, doi: 10.1016/S0140-6736(20)30211-7.
- [2] H. A. Rothan and S. N. Byrareddy, “The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak,” *J. Autoimmun.*, vol. 109, no. February, p. 102433, 2020, doi: 10.1016/j.jaut.2020.102433.
- [3] D. Abdullah, S. S. Ansari, S. Ahmar, and R. R. Rahmat, “The application of K - means clustering for province clustering in Indonesia of the risk of the COVID - 19 pandemic based,” *Qual. Quant.*, vol. 56, no. 3, pp. 1283–1291, 2022, doi: 10.1007/s11135-021-01176-w.
- [4] I. Ardiansyah, A. Subagiyo, M. R. Maulana, S. Abdullah, R. Ihwani, and T. Nova, “Sosial Demografi , Lingkungan , dan Pelayanan Kesehatan Terhadap Case Fatality Rate COVID-19 : Analisis Ekologi di Jawa Tengah,” *Banua J. Kesehat. Lingkung.*, vol. 5, no. 1, pp. 1–17, 2025, doi: 10.33860/bjkl.v5i1.4119.
- [5] I. N. Wahid, B. Aprilian, and A. D. Handoyo, “Analisis Multivariat Hubungan dan Perbedaan Kasus COVID-19 antar Wilayah di DKI Jakarta,” *Kohesi J. Sains dan Teknol.*, vol. 8, no. 6, pp. 1–9, 2025, doi: 10.8734/Kohesi.v1i2.365.
- [6] B. Audina *et al.*, “Analisis Survival pada Data Pasien Covid 19 di Kabupaten Jember (Survival Analysis on Covid 19 Patients Data in Jember Regency),” *Berk. Sainstek*, vol. 8, no. 4, pp. 118–121, 2020.
- [7] S. Komarayanti, W. Suharso, and E. Herrianto, “Local Fruits and Vegetables of Jember District That Can Increase Immunity during the Covid-19 Pandemic,” *Budapest Int. Res. Exact Sci. J.*, vol. 2, no. 4, pp. 492–508, 2020.
- [8] I. Karim, H. B. Daud, N. Zainuddin, and R. Sokkalingam, “Addressing limitations of the K-means clustering algorithm : outliers , non-spherical data , and optimal cluster selection,” *AIMS Math.*, vol. 9, no. 9, pp. 25070–25097, 2024, doi: 10.3934/math.20241222.
- [9] A. Iqbal, A. Shil, M. J. M. Chowdhury, M. A. Moni, and I. H. Sarker, “An Improved K-means Clustering Algorithm Towards an,” *Ann. Data Sci.*, vol. 11, pp. 1525–1544, 2024, doi: 10.1007/s40745-022-00428-2.
- [10] S. Rath, A. Tripathy, and A. Ranjan, “Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model,” *Diabetes Metab. Syndr. Clin. Res. Rev.*, vol. 14, no. 5, pp. 1467–1474, 2020, doi: 10.1016/j.dsx.2020.07.045.
- [11] P. T. A. Clustering-based *et al.*, “Subject Category : Subject Areas : Author for correspondence : Clustering-based methodology for comparing epidemiological dynamics with application to COVID-19 epidemiology in Europe,” *R. Soc. Open Sci.*, vol. 12, no. 9, p. 250440, 2025, doi: 10.1098/rsos.250440.
- [12] J. Sun, J. Qi, Z. Yan, Y. Li, and J. Liang, “Quantitative Study on American COVID-19 Epidemic Predictions and Scenario Simulations,” *ISPRS Int. J. Geo-Information*, vol. 13, no. 1, p. 31, 2024, doi: 10.3390/ijgi13010031.
- [13] Kompas.com, “UPDATE 31 Desember 2021: 4.292 Kasus Aktif Covid-19 di Indonesia,” 2021. [Online]. Available: <https://nasional.kompas.com/read/2021/12/31/20055991/update-31-desember-2021-4292-kasus-aktif-covid-19-di-indonesia>
- [14] B. M. Metisen and H. L. Sari, “ANALISIS CLUSTERING MENGGUNAKAN METODE K-MEANS DALAM PENGELOMPOKAN PENJUALAN PRODUK PADA SWALAYAN FADHILA,” *J. Media Infotama*, vol. 11, no. 2, pp. 110–118, 2015, doi: 10.37676/jmi.v11i2.258.
- [15] F. O. Lusiana, I. Fatma, and A. P. Windarto, “Estimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Simalungun,” *J. Informatics Manag. Inf. Technol.*, vol. 1, no. 2, pp. 79–84, 2021, doi:

10.47065/jimat.v1i2.104.

- [16] C. N. Id, L. Beattie, M. B. Id, T. Razzaghi, and S. Chen, "A machine learning and clustering-based approach for county-level COVID-19 analysis," *PLoS One*, vol. 17, no. 4, pp. 1–24, 2022, doi: 10.1371/journal.pone.0267558.
- [17] M. Z. El Khattabi, M. El Jai, and Y. Lahmadi, "Understanding the interplay between metrics, normalization forms, and data distribution in K-means clustering: A comparative simulation study," *Arab. J. Sci. Eng.*, vol. 49, pp. 2987–3007, 2023, doi: 10.1007/s13369-023-07741-9.
- [18] D. Steinley, "K-means clustering: A half-century synthesis," *Br. J. Math. Stat. Psychol.*, vol. 59, no. 1, pp. 1–34, 2006, doi: 10.1348/000711005X48266.
- [19] A. K. Jain, "Data clustering : 50 years beyond K-means q," *PATTERN Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, 2009, doi: 10.1016/j.patrec.2009.09.011.
- [20] S. K. Shukla, S. Shabnam, B. Boro, A. Saha, and B. Debbarma, "District level correlates of COVID-19 pandemic in India during March-October 2020," *PLoS One*, vol. 16, no. 9, pp. 1–17, 2021, doi: 10.1371/journal.pone.0257533.
- [21] C. Eduardo, R. Id, M. C. Oliveira, T. D. A. Eleuterio, M. Gonc, and E. Regina, "Spatial analysis of COVID-19 incidence and the sociodemographic context in Brazil," *PLoS One*, vol. 16, no. 3, pp. 1–16, 2021, doi: 10.1371/journal.pone.0247794.
- [22] N. W. Wardani and K. Selatan, "Analisis Pesebaran Penularan Virus Corona Di Provinsi Jawa Tengah Menggunakan Metode K-Means Clustering," *JATISI*, vol. 8, no. 1, pp. 105–117, 2021, doi: 10.35957/jatisi.v8i1.590.
- [23] J. Luo, Y. Zhang, and Y. Song, "Design for Pandemic Information : Examining the Effect of Graphs on Anxiety and Social Distancing Intentions in the COVID-19," *Front. Public Heal.*, vol. 10, no. May, pp. 1–9, 2022, doi: 10.3389/fpubh.2022.800789.
- [24] L. Berganda, "Penerapan Metode Suksesif Interval pada Analsis Regresi," *Jambura J. Math.*, vol. 1, no. 1, pp. 43–53, 2019, doi: 10.34312/jjom.v1i1.1742.
- [25] R. Diana, "PEMODELAN KASUS COVID-19 MENGGUNAKAN MODEL REGRESI NONPARAMETRIK Studi Kasus di Jakarta A CASE STUDY IN JAKARTA)," *Semin. Nas. Off. Stat. 2020*, vol. 2, no. 1, pp. 108–115, 2020, doi: 10.34123/semnasoffstat.v2020i1.588.