

## Pengelompokan Penyebaran Covid-19 di Provinsi Jawa Barat menggunakan Metode *Clustering K-Medoids*

Ibtihal Qomariyyah Luthfiyyah<sup>1\*</sup>, Betha Nurina Sari<sup>2</sup>

<sup>1,2</sup>Jurusan Informatika, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang

\*Email: 2010631170079@student.unsika.ac.id

### Abstrak

Covid-19 merupakan suatu penyakit yang menjangkiti sistem pernapasan manusia dan memiliki kemampuan penularan yang sangat cepat. Provinsi Jawa Barat menjadi salah satu daerah yang terkena dampak pandemi Covid-19. Jumlah masyarakat terkonfirmasi virus Covid-19 di Jawa Barat yang masih bertambah hari demi hari. Oleh karena itu, diperlukan adanya pengelompokan tingkat kerawanan penyebaran Covid-19 khususnya di Provinsi Jawa Barat menggunakan data dari website resmi pemerintah Provinsi Jawa Barat dengan menggunakan 5 atribut, yaitu nama\_kab\_kota, konfirmasi\_total, konfirmasi\_sembuh, konfirmasi\_meninggal, dan konfirmasi\_aktif. Penelitian ini bertujuan untuk mengidentifikasi pola penyebaran Covid-19 untuk mendukung pengambilan keputusan yang lebih efektif di tingkat regional. Metode penelitian melibatkan proses data mining yaitu pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan penyebaran berdasarkan metodologi CRISP-DM. Proses modeling menggunakan algoritma K-Medoids dengan 3 cluster sesuai dengan zona warna pemerintah. Hasil penelitian ini menunjukkan 3 cluster, yaitu cluster hijau merupakan jumlah kasus yang minimal dengan 16 Kabupaten/Kota. Cluster kuning merupakan mulai waspada akan jumlah kasus dengan 6 Kabupaten/Kota. Cluster merah merupakan kasus sudah sangat parah dengan 5 Kabupaten/Kota. Hasil pengujian Silhouette Coefficient yang menguji  $n\_cluster = 2, 3, 4$ , dan 5 menunjukkan bahwa  $n\_cluster = 3$  merupakan cluster yang terbaik dengan nilai sebesar 0.77.

**Kata kunci:** Clustering, Covid-19, Data Mining, K-Medoids, Silhouette Coefficient

### Abstract

Covid-19 is a disease that infects the human respiratory system and has a high-speed transmission ability. West Java Province is one of the areas affected by the Covid-19 pandemic. The number of people confirmed with the Covid-19 virus in West Java is still increasing daily. Therefore, it is necessary to group the level of vulnerability to the spread of Covid-19, especially in West Java Province, using data from the official website of the West Java Provincial Government using 5 attributes, namely district\_city\_name, total\_confirmation, confirmed\_recovered, confirmed\_death, and confirmed\_active. This study aims to identify the pattern of the spread of Covid-19 to support more effective decision-making at the regional level. The research method involves a data mining process, namely business understanding, data understanding, data preparation, modeling, evaluation, and deployment based on the CRISP-DM methodology. The modeling process uses the K-Medoids algorithm with 3 clusters according to the government's color zone. The results of this study show 3 clusters, namely the green cluster is the minimum number of cases with 16 districts/cities. The yellow cluster is starting to be alert to the number of cases with 6 districts/cities. The red cluster is a very severe case with 5 districts/cities. The results of the Silhouette Coefficient test that tested  $n\_cluster = 2, 3, 4$ , and 5 showed that  $n\_cluster = 3$  is the best cluster with a value of 0.77.

**Keywords:** Clustering, Covid-19, Data Mining, K-Medoids, Silhouette Coefficient

## PENDAHULUAN

Pada akhir tahun 2019, di Wuhan, Cina, muncul suatu virus baru yang dikenal sebagai Corona Virus Disease 2019 atau Covid-19. Virus ini menyebar ke seluruh dunia pada tahun 2020, menjadi permasalahan kesehatan serius. *World Health Organization* (WHO) secara resmi menyatakan Covid-19 sebagai pandemi dalam waktu kurang dari empat bulan dan menjangkiti negara-negara, termasuk Indonesia.

Kasus pertama Covid-19 di Indonesia dikonfirmasi pada Senin, 2 Maret 2020, dan diumumkan secara resmi oleh pemerintah melalui situs web resmi Kementerian Kesehatan Republik Indonesia. Sejak saat itu, pemerintah

terus mengumumkan kasus yang dikonfirmasi, sembuh, aktif, dan meninggal setiap harinya.

Total kematian akibat Covid-19 di Indonesia menempatkannya di peringkat kedua tertinggi di Asia menurut Worldometer pada Agustus 2023, dengan jumlah mencapai 161,92 ribu orang. Indonesia juga memimpin dalam angka kematian di Asia Tenggara. Pandemi ini menjalar ke seluruh provinsi di Indonesia, menempatkan Indonesia di peringkat kedua tertinggi di Asia. (Darmawan, 2023).

Pemerintah memfokuskan perhatiannya pada delapan provinsi dengan tingkat penularan tertinggi, termasuk Jawa Barat. Langkah-langkah pencegahan, seperti pembatasan sosial

dan pemeriksaan di pusat keramaian, diterapkan untuk mengendalikan penyebaran. Masyarakat diimbau untuk mematuhi protokol kesehatan, seperti menjaga jarak, menggunakan masker, mencuci tangan, menghindari kerumunan, dan menjalani gaya hidup sehat.

Pandemi Covid-19 menjadi tantangan serius bagi kesehatan global dan memengaruhi berbagai aspek kehidupan. Dalam upaya mengendalikan penyebaran Covid-19, pemerintah menetapkan target untuk menurunkan kematian, meningkatkan angka kesembuhan, dan mengendalikan pertumbuhan kasus positif.

Meskipun WHO telah mencabut status pandemi global untuk Covid-19, penelitian ini tetap relevan karena pandemi tersebut menjadi pengingat nyata bahwa dunia rentan terhadap wabah-wabah baru di masa depan. Juru Bicara Covid-19 Kementerian Kesehatan, dr. Mohammad Syahril, menekankan bahwa pencabutan status pandemi tidak berarti hilangnya ancaman dari Covid-19 (Rokom, 2023).

Menurut berita yang diterbitkan oleh DetikJabar (Bagaskara, 2023), Dinas Kesehatan Jawa Barat mengimbau masyarakat untuk tetap waspada karena adanya peningkatan kasus baru di berbagai wilayah di Jawa Barat jelang Natal dan Tahun Baru 2023. Oleh karena itu, analisis data menjadi sangat penting sebagai alat yang kritis untuk memahami pola penyebaran dan merumuskan strategi penanggulangan yang efektif. Dalam menghadapi tantangan ini, sinergi antara pemerintah, tenaga kesehatan, dan partisipasi aktif masyarakat merupakan kunci utama untuk memitigasi penyebaran lebih lanjut, serta menjaga keamanan bersama di tengah ketidakpastian yang masih mengintai.

*Data mining* sebagai pendekatan yang kuat untuk mengekstraksi pola dan pengetahuan dari *dataset*, menawarkan peluang yang besar dalam pemahaman lebih lanjut tentang dinamika penyebaran Covid-19. Dalam konteks penelitian ini menggunakan metode *clustering*, sebuah teknik dalam *data mining* yang fokus pada pengelompokan data menjadi kelompok yang homogen berdasarkan karakteristik tertentu.

Penelitian *clustering* penyebaran Covid-19 pernah diangkat sebelumnya, misalnya penelitian oleh (Hardiani, 2022) yang menggunakan algoritma *K-Means* dengan tools *Rapidminer* yang menghasilkan 3 *cluster*

optimal berdasarkan perhitungan *Davies Bouldin Index*. CRISP-DM digunakan sebagai metode analisis *data mining*. Data yang digunakan yaitu nama 34 provinsi di Indonesia dengan total kasus, total meninggal, total sembuh, dan total kasus aktif. Penelitian lain oleh (Wijaya dkk., 2022) yaitu klasterisasi data kawalcovid19.id berdasarkan 34 provinsi di Indonesia dilakukan dengan menerapkan *K-Means* menggunakan tools WEKA menghasilkan 3 *Cluster*. Penelitian lain oleh (Azhari dkk., 2023) yang menggunakan algoritma *K-Means* terhadap data positif Covid-19 di 34 provinsi Indonesia. Dengan tahapan proses KDD dan perhitungan manual pada *microsoft excel* yang dikombinasikan dengan perhitungan menggunakan tools *Rapidminer* menghasilkan 2 *cluster*. Penelitian lain oleh (Fira dkk., 2021), di mana *clustering* data Covid-19 di 34 Provinsi Indonesia menghasilkan 2 *cluster*. KDD digunakan sebagai metode proses analisis *data mining* dan pada proses pengolahan data menggunakan bantuan tools *Rapidminer*.

Penelitian sebelumnya telah berhasil mengidentifikasi pola umum penyebaran Covid-19 di 34 provinsi Indonesia menggunakan metode *clustering*. Namun, sebagian besar penelitian tersebut masih bersifat nasional dan tools yang digunakan *Rapidminer* dan WEKA. Penelitian ini akan menggunakan data wilayah yang lebih spesifik yaitu Provinsi Jawa Barat di Tingkat Kabupaten/Kota dan bahasa pemrograman *Python* dengan tools *Google Colab* yang menawarkan fleksibilitas dan kemudahan dalam analisis data.

Selain itu, berbeda dengan penelitian *clustering* penyebaran Covid-19 sebelumnya yang seringkali menggunakan algoritma *K-Means*, penelitian ini memilih *K-Medoids*. Pemilihan algoritma *K-Medoids* juga dikarenakan hasil penelitian (Fira dkk., 2021), yang menunjukkan hasil *Silhouette Coefficient* dengan metode *K-Means* sebesar 0.207, *K-Medoids* sebesar 0.347, sehingga dinyatakan bahwa *K-Medoids* memiliki akurasi lebih tinggi daripada *K-Means*.

Banyak penelitian yang telah menggunakan *K-Medoids*, diantaranya digunakan untuk pengelompokan mahasiswa yang berpotensi *drop out* (Bahri & Marisa Midyanti, 2023), *clustering* indikator kemiskinan di Jawa Timur (Alfiah dkk., 2022), produksi perikanan tangkap Kabupaten Aceh

Utara (Fajriana, 2021), aplikasi pembelajaran di masa pandemi Covid-19 (Samudi dkk., 2020), dan *cluster* kejahatan di Yogyakarta (Atmaja, 2019).

Algoritma *K-Medoids* telah digunakan sebelumnya dalam penelitian *clustering* penyebaran Covid-19 di Jawa Barat oleh (Aria, 2021). Penelitian tersebut menggunakan data per 20 Mei 2021 yang berasal dari portal (<https://pikobar.jabarprov.go.id>). Dengan bantuan *tools Rapidminer*, penelitian tersebut menghasilkan 3 *cluster*, namun tidak disertai dengan pengujian kualitas *clustering* untuk memastikan keoptimalan *clustering*.

Penelitian ini berbeda dari (Aria, 2021) dalam beberapa aspek penting. Pertama, sumber data yang akan digunakan berbeda yaitu dari portal (<https://opendata.jabarprov.go.id/id>) dan data per 10 Januari 2023. Kedua, penelitian ini akan menggunakan variabel yang berbeda, mencakup nama kabupaten/kota, konfirmasi total, konfirmasi sembuh, konfirmasi meninggal, dan konfirmasi aktif. Ketiga, proses pengolahan data akan menggunakan bahasa pemrograman *Python* dengan dukungan *tools Google Colab*. Keempat, metode alur pengolahan data akan menggunakan CRISP-DM. Kelima, penelitian ini akan menggunakan metode parameter pengujian yaitu *Silhouette Coefficient* untuk memastikan hasil *cluster* yang optimal.

Pentingnya penelitian ini terletak pada kemampuannya untuk memberikan informasi yang lebih mendalam tentang pola penyebaran virus di tingkat regional. Informasi ini bisa menjadi dasar pengambilan keputusan yang lebih baik dalam menanggulangi dampak pandemi. Meskipun Covid-19 telah berkurang, pemahaman tentang pola penyebarannya dapat memberikan landasan ilmiah yang kuat untuk menghadapi penyakit menular lainnya.

Selain itu, penggunaan metode *clustering K-Medoids* dan evaluasi *Silhouette Coefficient* dengan memanfaatkan *Python* di *Google Colab*, diharapkan mampu menghasilkan analisis yang lebih akurat mengenai titik-titik fokus penyebaran penyakit. Dengan pemahaman ini, alokasi sumber daya yang terbatas dapat diarahkan secara lebih efektif.

Hasil dari penelitian ini diharapkan dapat bermanfaat untuk pengendalian penyakit menular saat ini dan memperkuat kesiapsiagaan sistem kesehatan masyarakat di era pascapandemi, sehingga respons terhadap

wabah di masa depan bisa lebih cepat dan tepat sasaran.

## TINJAUAN PUSTAKA

*Data mining* didefinisikan sebagai proses ekstraksi informasi yang bermanfaat dari sejumlah data yang tersimpan secara tersirat dalam suatu basis data (Mulaab, 2017). Dalam ranah *data mining*, terdapat beragam algoritma, metode, dan teknik untuk menggali atau mencari pengetahuan serta informasi. Berdasarkan fungsi dan tujuan terdapat pengelompokan *data mining* seperti, deskripsi, klasifikasi, prediksi, estimasi, pengklasteraan, dan asosiasi (Buulolo, 2020).

*Clustering* merupakan proses pengelompokan data yang mempunyai kesamaan nilai, menciptakan kelompok-kelompok homogen. Data yang dapat diorganisasikan dalam pengelompokan ini melibatkan hasil pengamatan, rekaman data, atau kelas dan objek yang memiliki kesamaan. Dalam konsep *clustering* ini, perbedaannya dengan klasifikasi adalah tidak adanya penggunaan variabel target sebagai acuan. (Buulolo, 2020). *Clustering* masuk dalam kategori metode pembelajaran tak terawasi (*unsupervised learning*). Dalam sistem *unsupervised*, suatu set input diberikan tanpa adanya set target yang diharapkan. Sistem ini akan memisahkan input ke dalam beberapa kelompok atau *cluster*, berdasarkan tingkat kesamaan pola yang dimiliki oleh kumpulan input tersebut (Muflikhah dkk., 2018). Beberapa algoritma yang digunakan dalam *clustering* antara lain *K-Means*, *K-Medoids*, DBSCAN, *Hierarchical Clustering*, *Fuzzy C-Means*, *Self-Organizing Map* dan berbagai algoritma lainnya.

Dalam buku karya (Saputra dan Kristiyanti, 2022), dijelaskan awal ditemukan *K-Medoids* dan tahapan algoritmanya. Permasalahan utama pada *K-Means* adalah nilai *outlier*, karena itulah harus dilakukan normalisasi data sebelum menggunakan algoritma tersebut. Selain itu, untuk mencari nilai *centroid* dibutuhkan waktu cukup lama karena langkah perhitungan yang harus dilakukan cukup lama karena langkah perhitungan yang harus dilakukan. Sehingga, muncul algoritma *K-Medoids* yang dapat mengatasi beberapa permasalahan tersebut. Cara kerjanya lebih sederhana dibandingkan dengan *K-Means*. Pada *K-Medoids*, *centroid* akan tetap dipilih secara acak namun dengan nilai yang terdapat di salah satu baris *dataset*, bukan

mencari atau menentukan nilai baru. Berikut adalah tahapan algoritma *K-Medoids* yang memberikan gambaran komprehensif tentang bagaimana algoritma ini diimplementasikan:

1. Menentukan jumlah *cluster* yang diinginkan.
2. Pilih secara acak medoid sebanyak  $k$  dari dataset yang tersedia;
3. Hitung masing-masing jarak terhadap medoid menggunakan pengukuran jarak, bisa menggunakan *euclidean*, *manhattan distance*, atau yang lainnya.
4. Lakukan iterasi medoid untuk mendapatkan *cluster*;
5. Hitung total simpangan (S). Jika *cluster* tidak berubah dan jumlah jarak terdekat pada iterasi pertama lebih kecil dari jumlah jarak terdekat pada iterasi kedua, maka hentikan iterasi;
6. Jika langkah 5 tidak terpenuhi, maka ulangi langkah 3 sampai 5 dan atur kembali medoid yang baru untuk dihitung kembali seperti sebelumnya hingga terpenuhi.

*Silhouette coefficient* adalah parameter yang mengukur seberapa cocok setiap titik data dengan kelompok yang ditentukan. Nilai *silhouette* mencerminkan sejauh mana suatu objek serupa dengan kelompoknya sendiri (kohesi) dibandingkan dengan kelompok lainnya (pemisahan). Rentang nilai koefisien ini berkisar antara -1 hingga 1, dimana nilai mendekati 1 menunjukkan bahwa titik data tersebut berhasil terkelompok dengan baik, nilai mendekati 0 mengindikasikan adanya tumpang tindih antar kelompok, dan nilai yang mendekati -1 menandakan bahwa titik data tersebut salah terkelompok (Yadav, 2023). Dapat dilihat rentang nilai dan interpretasi pada Tabel 1 (Sari dan Primajaya, 2020).

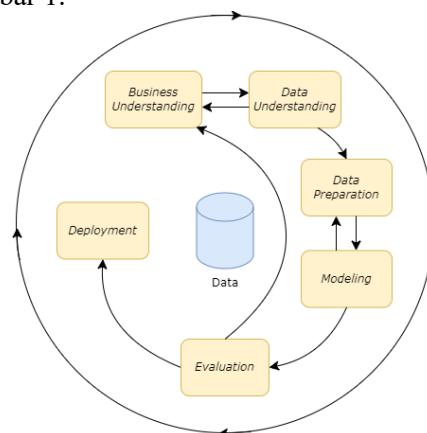
Tabel 1. Interpretasi Nilai *Silhouette*

Rentang Nilai	Interpretasi
0.71 – 1.0	Struktur yang kuat
0.51 – 0.70	Struktur yang beralasan
0.26 – 0.50	Struktur lemah
< 0.25	Tidak ditemukan struktur yang substansial

## METODE

Metode yang digunakan pada penelitian ini adalah (*Cross-Industry Standard Process for Data Mining*). Tahapan yang ditempuh yaitu

*Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*. Setiap langkah memiliki tujuan khusus dan koneksi ke tugas yang berbeda, tahapan metodologinya dapat dilihat pada Gambar 1.



Gambar 1. Alur CRISP-DM

### 3.1 Business Understanding

Dalam tahap ini, fokus utama adalah memahami tujuan penelitian dan kebutuhan informasi yang mendasari pengelompokan data Covid-19 di Jawa Barat. Penting untuk melakukan studi literatur dan pengumpulan informasi terkait guna memperdalam pemahaman terhadap konteks, yang melibatkan tinjauan mendalam terhadap kerangka teoritis dan temuan penelitian terkait.

### 3.2 Data Understanding

Tahap kedua melibatkan sumber data yang akan digunakan untuk *clustering*. Dalam hal ini, data berasal dari portal resmi data terbuka milik Pemerintah Daerah Provinsi Jawa Barat tahun 2023. Setelah data terkumpul kemudian data tersebut dipahami setiap atributnya. Pemahaman mendalam diperlukan untuk proses selanjutnya.

### 3.3 Data Preparation

Langkah berikutnya adalah mempersiapkan data dengan membersihkan data sebelum menggunakan algoritma *K-Medoids*. Tahap ini dilakukan mulai dengan memilih atribut apa saja yang akan digunakan untuk proses *clustering*, dan pembersihan data.

### 3.4 Modeling

Tahap ini menjelaskan pemilihan dan implementasi model. Dalam konteks penelitian

ini, algoritma yang dipilih adalah *K-Medoids* untuk *clustering* data penyebaran Covid-19 di Provinsi Jawa Barat tingkat kabupaten/kota.

### 3.5 Evaluation

Tahap evaluasi melibatkan penggunaan *Sillhouette Coefficient* sebagai metrik evaluasi utama. Penelitian membandingkan hasil metrik evaluasi dari beberapa *k* yang sudah ditentukan.

### 3.6 Deployment

Pada tahap terakhir, fokus utama diberikan pada penyusunan dokumentasi mengenai langkah-langkah dari proses *clustering* melalui pendekatan *data mining*. Hasil *clustering* dapat divisualisasikan dalam bentuk *scatter plot*.

## HASIL DAN PEMBAHASAN

### 4.1 Business Understanding

Pada tahap ini, peneliti telah memahami bahwa pandemi Covid-19 memiliki dampak serius di Provinsi Jawa Barat, dan diperlukan analisis yang mendalam untuk memahami pola penyebarannya. Fokus pada pengelompokan penyebaran dengan metode *clustering K-Medoids* untuk memberikan wawasan yang lebih tajam terkait distribusi kasus di tingkat Kabupaten/Kota.

### 4.2 Data Understanding

Data terkait Covid-19 di Provinsi Jawa Barat telah dikumpulkan untuk digunakan dalam analisis. Data diperoleh dari portal resmi Open Data Jabar (<https://opendata.jabarprov.go.id/id>) mengenai Perkembangan Harian Kasus Terkonfirmasi Positif Covid-19 Berdasarkan Kabupaten/Kota. Satuan Tugas Penanganan Covid-19 mengeluarkan data dalam periode 1 hari sekali.

Tabel 2. Data Covid-19 Jawa Barat

Atribut	Tipe Data
tanggal	object
kode_kab_kota	integer
nama_kab_kota	object
konfirmasi_total	integer
konfirmasi_sembuh	integer
konfirmasi_meninggal	integer
konfirmasi_aktif	integer
konfirmasi_total_daily_growth	integer
konfirmasi_sembuh_daily_growth	integer
konfirmasi_meninggal_daily_growth	integer
konfirmasi_aktif_daily_growth	integer
kota_kab_belum_teridentifikasi	float
sembuh_unidentified	float
meninggal_unidentified	float
id	integer

Penelitian ini menggunakan data per 10 Januari 2023. *Dataset* ini berisi data perkembangan harian kasus terkonfirmasi Covid-19 dari tahun ke tahun yang memiliki beberapa atribut terlihat pada Tabel 2.

### 4.3 Data Preparation

Dalam tahap persiapan data, langkah-langkah dilakukan untuk menghilangkan data dan *noise* yang tidak konsisten, serta melakukan koreksi pada kesalahan data dengan menggunakan bahasa pemrograman *python*. Beberapa tindakan yang dilakukan dalam persiapan data melibatkan hal-hal berikut:

#### 1. Penentuan atribut

Dalam *dataset* terdapat banyak atribut, sehingga diperlukan penentuan atribut yang penting agar mendapatkan hasil *cluster* yang sesuai. Adapun atribut yang dipilih untuk penelitian ini sejumlah 5 yaitu nama kab/kota, konfirmasi total, konfirmasi sembuh, konfirmasi meninggal, dan konfirmasi aktif.

#### 2. Data Reduction

Data yang digunakan yaitu jumlah kasus Covid-19 pada tanggal 10 Januari 2023 dan pada kolom nama\_kab\_kota terdapat kabupaten/kota 'jawa barat' itu dihilangkan sehingga jumlah data kabupaten/kota ada 27. Selain itu, atribut yang tidak dipilih pun harus dihilangkan.

Selanjutnya, perlu ditekankan bahwa seleksi atribut dan reduksi data ini bertujuan untuk menyederhanakan kompleksitas *dataset*, memastikan fokus pada variabel yang paling relevan, dan meningkatkan efisiensi *clustering*. Lima atribut yang dipilih sesuai dengan penelitian (Hardiani, 2022).

Berikut merupakan hasil *data preparation* terlihat pada Tabel 3.

Tabel 3. Hasil *Data Preparation*

nama_kab_kota	konfirmasi_total	konfirmasi_sembuh	konfirmasi_meninggal	konfirmasi_aktif
Kabupaten Bandung	59259	58449	638	172
Kabupaten Bandung Barat	30588	30133	311	144
Kabupaten Bekasi	92984	92307	571	106
Kabupaten Bogor	104904	104576	206	122
Kabupaten Ciamis	16790	16448	331	11
Kabupaten Cianjur	16076	15841	228	7
Kabupaten Cirebon	32847	31849	986	12
Kabupaten Garut	35183	33842	1307	34

nama_kab_kota	konfirmasi_total	konfirmasi_sembuh	konfirmasi_meninggal	konfirmasi_aktif
Kabupaten Indramayu	22392	21599	766	27
Kabupaten Karawang	57893	55758	2052	83
Kabupaten Kuningan	17248	16954	290	4
Kabupaten Majalengka	14282	14005	268	9
Kabupaten Pangandaran	6100	5943	157	0
Kabupaten Purwakarta	17768	17117	622	29
Kabupaten Subang	15019	14686	299	34
Kabupaten Sukabumi	16858	16449	358	51
Kabupaten Sumedang	14138	13907	192	39
Kabupaten Tasikmalaya	10771	10424	325	22
Kota Bandung	112773	112022	455	296
Kota Banjar	6228	6076	149	3
Kota Bekasi	200729	199545	1076	108
Kota Bogor	68054	67176	560	318
Kota Cimahi	24065	23757	204	104
Kota Cirebon	17471	16823	574	74
Kota Depok	188542	186093	2278	171
Kota Sukabumi	12869	12564	291	14
Kota Tasikmalaya	19554	18875	599	80

Data pada Tabel 3 berisi lima atribut yaitu nama kabupaten kota, jumlah konfirmasi total, jumlah konfirmasi sembuh, jumlah konfirmasi meninggal, dan jumlah konfirmasi aktif.

#### 4.4 Modeling

Dalam menerapkan *clustering* data menggunakan *K-Medoids*, diperlukan penentuan jumlah *cluster* yang sesuai. Data kasus Covid-19 tersebut dikelompokkan menjadi tiga *cluster*, dengan memperhatikan tiga zona warna sebagai perbedaannya. Penentuan zona warna didasarkan pada penentuan zona warna yang dilakukan pemerintah di mana warnanya ada hijau, kuning, dan merah. Hijau menunjukkan jumlah kasus yang minimal dan masih ditangani, kuning menunjukkan bahwa kita harus waspada akan jumlah kasus, dan merah menunjukkan bahwa kasus sudah mulai sangat parah.

Sebuah set awal dari titik pusat klaster (medoids) dinisialisasi dalam analisis dengan menggunakan fungsi-fungsi dan *library* yang tersedia dalam bahasa pemrograman proses *python*. Inisialisasi ini memainkan peran krusial dalam membentuk klaster yang optimal. Melalui serangkaian iterasi, medoids diperbarui untuk meminimalkan jarak antara objek dan medoid di dalam klaster. Proses ini berulang hingga konvergensi atau batas iterasi yang ditentukan.

Tahapan ini memberikan gambaran *K-Medoids* mengelompokkan dengan mempertimbangkan zona warna yang sudah ditentukan sebelumnya.

Berikut adalah hasil *clustering* yang terlihat pada Tabel 4, yang memberikan informasi tentang distribusi kasus Covid-19 di berbagai *cluster* setelah proses pengelompokan.

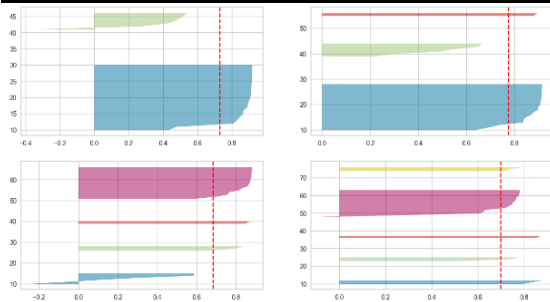
Tabel 4. Hasil *Clustering K-Medoids*

Cluster	Jumlah	Nama Kabupaten/Kota
Hijau	16	Kab. Ciamis, Kab. Cianjur, Kab. Indramayu, Kab. Kuningan, Kab. Majalengka, Kab. Pangandaran, Kab. Purwakarta, Kab. Subang, Kab. Sukabumi, Kab. Sumedang, Kab. Tasikmalaya, Kota Banjar, Kota Cimahi, Kota Cirebon, Kota Sukabumi, Kota Tasikmalaya
Kuning	6	Kab. Bandung, Kab. Bandung Barat, Kab. Cirebon, Kab. Garut, Kab. Karawang, Kota Bogor
Merah	5	Kab. Bekasi, Kab. Bogor, Kota Bandung, Kota Bekasi, Kota Depok

Dari hasil pengelompokan pada Tabel 4, menunjukkan bahwa jumlah terkonfirmasi Covid-19 dengan zona hijau atau *cluster* 0 ada 16 Kabupaten/Kota. Sebanyak 6 Kabupaten/Kota tergabung dalam zona kuning atau *cluster* 1, menandakan tingkat kekhawatiran yang perlu diperhatikan. Sementara itu, terdapat 5 Kabupaten/Kota lainnya yang masuk dalam zona merah atau *cluster* 2, menunjukkan bahwa kasus di wilayah tersebut sudah mencapai tingkat yang cukup serius.

#### 4.5 Evaluation

Pengujian bermanfaat untuk menilai kualitas *clustering* yang dihasilkan. Salah satu teknik yang diterapkan adalah menggunakan *Silhouette Coefficient*. Proses Pengujian membuat plot *Silhouette* untuk *cluster K-Medoids* dengan *n\_cluster* sebagai 2, 3, 4, 5. Eksekusi program *python* menghasilkan plot *Silhouette* berikut untuk 2, 3, 4, 5 *cluster* terlihat pada Gambar 2.

Gambar 2. Visualisasi *Silhouette Score*

Berikut adalah interpretasi dari visualisasi *Silhouette Score* yang dilakukan pada Gambar 2 dengan tujuan untuk mengetahui nilai optimal pada  $n\_cluster$ .

1. Nilai  $n\_cluster$  4 (kiri bawah) dan 5 (kanan bawah) terlihat suboptimal untuk data yang diberikan karena skor siluet di bawah rata-rata dan adanya fluktuasi luas dalam ukuran plot siluet.
2. Nilai  $n\_cluster$  2 (kiri atas) dan 3 (kanan atas) terlihat optimal karena skor berada mendekati rata-rata skor dan fluktuasi ukurannya serupa. Untuk plot dengan ( $n\_cluster=3$ ) itu ketebalannya lebih seragam daripada plot dengan ( $n\_cluster=2$ ) dengan ketebalan plot lebih banyak dari yang lain. Garis merah ( $n\_cluster=3$ ) lebih mendekati nilai 1 daripada ( $n\_cluster=2$ ).

Berdasarkan evaluasi *Silhouette Coefficient* pada Tabel 5, dapat disimpulkan jumlah *cluster* ( $n\_cluster$ ) sebanyak 3 menunjukkan kualitas *cluster* yang paling optimal. Hal ini dikarenakan nilai hasil pengujian pada  $n\_cluster=3$  mencapai 0.77, mendekati nilai 1 dan memiliki struktur yang kuat berada dalam rentang nilai 0.71 – 1.00 sesuai dengan interpretasi pada Tabel 1.

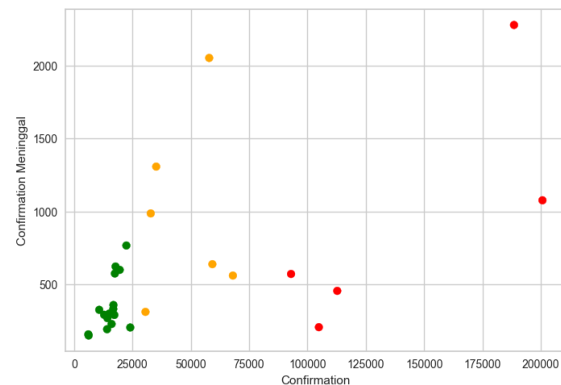
Tabel 5. Hasil Pengujian *Silhouette Coefficient*

Jumlah Cluster	<i>Silhouette Score</i>
2	0.72
3	<b>0.77</b>
4	0.68
5	0.69

#### 4.6 Deployment

Hasil pemodelan telah berhasil dievaluasi. Selanjutnya, tahap penyusunan dokumentasi mengenai langkah-langkah dari proses *clustering* dalam bentuk laporan penelitian berupa publikasi jurnal. Adapun hasil *clustering*

divisualisasikan dalam bentuk *scatter plot* dapat dilihat pada Gambar 3.

Gambar 3. Visualisasi *Clustering*

Dari visualisasi pada Gambar 3 tersebut dapat terlihat plot mana saja yang termasuk dalam tingkat penyebaran Covid-19 hijau, kuning, dan merah di Provinsi Jawa Barat. Zona warna dari hasil *clustering* dapat membantu merumuskan strategi penganggulangan yang lebih spesifik.

#### SIMPULAN

Berdasarkan penelitian pada pembahasan sebelumnya, dapat disimpulkan bahwa hasil *cluster* menunjukkan bahwa pada *cluster* hijau merupakan kasus yang minimal dengan 16 Kabupaten/Kota. Pada *cluster* kuning merupakan mulai waspada akan jumlah kasus dengan 6 Kabupaten/Kota. Pada *cluster* merah merupakan kasus sudah sangat parah dengan 5 Kabupaten/Kota. Pembentukan *cluster* zona penyebaran Covid-19 diharapkan dapat mempermudah Satuan Tugas Percepatan Penanganan Covid-19 Jawa Barat dalam mengatasi penyebaran Covid-19 berdasarkan tingkat *cluster*. Evaluasi menggunakan *Silhouette Coefficient* menyimpulkan bahwa jumlah *cluster* sebanyak 3 ( $n\_cluster=3$ ) memberikan kualitas *cluster* terbaik, karena mendapatkan nilai tertinggi, yaitu 0.77, yang mendekati nilai optimal 1 dan memiliki struktur yang kuat.

Untuk penelitian selanjutnya, topik *clustering* dapat dikembangkan lebih luas dengan fokus pada isu-isu yang relevan di era pascapandemi. Misalnya, *clustering* dapat diterapkan untuk menganalisis tingkat kerentanan kesehatan masyarakat, penyebaran penyakit endemik seperti demam berdarah.

Selain di bidang kesehatan, *clustering* juga dapat diaplikasikan pada bidang lain seperti pendidikan, ekonomi, sosial, dan pariwisata. Dari sisi algoritma, dapat mengeksplorasi algoritma *clustering* selain *K-Medoids* yang mungkin lebih efektif. Selain itu, untuk memastikan kualitas hasil *clustering*, pengujian keakuratan dapat dilakukan menggunakan berbagai metode evaluasi seperti *Silhouette Coefficient* dan *Davies-Bouldin Index*. Pengembangan ini diharapkan dapat memperkaya hasil analisis dan memberikan wawasan yang lebih mendalam untuk pengambilan keputusan di berbagai bidang.

#### DAFTAR PUSTAKA

- Alfiah, F., Almadayani, A., Al Farizi, D., & Widodo, E. (2022). Analisis Clustering K-Medoids Berdasarkan Indikator Kemiskinan di Jawa Timur Tahun 2020. *Jurnal Ilmiah Sains*, 22(1), 1–7. <https://doi.org/10.35799/jis.v22i1.35911>
- Aria, R. R. (2021). Implementation of the K-Medoids Algorithm for Data Clustering of Covid 19 Cases in West Java. *International Journal of Information System & Technology Akreditasi*, 5(1), 11–16. <https://doi.org/10.30645/ijistech.v5i1.109>
- Atmaja, E. H. S. (2019). Implementation of k-Medoids Clustering Algorithm to Cluster Crime Patterns in Yogyakarta. *International Journal of Applied Sciences and Smart Technologies*, 1(1), 33–44. <https://doi.org/10.24071/ijasst.v1i1.1859>
- Azhari, R., Hartama, D., Lubis, M. R., Nasution, D. F., & Windarto, A. P. (2023). Analisis Penerapan Data Mining Terhadap Kasus Positif Covid-19 Menggunakan Metode K-Means Clustering. *Journal of Informatics, Electrical and Electronics Engineering*, 3(2), 221–235. <https://doi.org/10.47065/jieec.v3i2.1760>
- Bagaskara, B. (2023, Desember 22). *Waspada! COVID-19 di Jabar Meningkatkan Jelang Nataru*. detikJabar. <https://www.detik.com/jabar/berita/d-7103313/waspada-covid-19-di-jabar-meningkat-jelang-nataru>
- Bahri, S., & Marisa Midyanti, D. (2023). Penerapan Metode K-Medoids Untuk Pengelompokan Mahasiswa Berpotensi Drop Out. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 10(1), 165–172. <https://doi.org/10.25126/jtiik.2023106643>
- Buulolo, E. (2020). *Data Mining Untuk Perguruan Tinggi*. Deepublish.
- Darmawan, A. D. (2023, Agustus 14). *Total Kematian Covid-19 Indonesia Urutan Ke-2 di Asia*. databoks. <https://databoks.katadata.co.id/datapublish/2023/08/14/total-kematian-covid-19-indonesia-urutan-ke-2-di-asia>
- Fajriana. (2021). Analisis Algoritma K-Medoids pada Sistem Klasterisasi Produksi Perikanan Tangkap Kabupaten Aceh Utara. *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 7(2), 263–269. <https://doi.org/10.26418/jp.v7i2.47795>
- Fira, A., Rozikin, C., & Garno. (2021). Komparasi Algoritma K-Means dan K-Medoids Untuk Pengelompokan Penyebaran Covid-19 di Indonesia. *Journal of Applied Informatics and Computing (JAIC)*, 5(2), 133–138. <https://doi.org/10.30871/jaic.v5i2.3286>
- Hardiani, T. (2022). Analisis Clustering Kasus Covid 19 di Indonesia Menggunakan Algoritma K-Means. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, 11(2), 156–165. <https://doi.org/10.23887/janapati.v11i2.45376>
- Muflikhah, L., Ratnawati, D. E., & Putri, R. R. M. (2018). *Data Mining*. UB Press.
- Mulaab. (2017). *Data Mining : Konsep dan Aplikasi (I)*. Media Nusa Creative.
- Rokom. (2023, Mei 9). *Darurat Kesehatan Global Dicabut, Kemenkes : Bukan Berarti Pandemi COVID-19 Berakhir*. Kementerian Kesehatan RI. <https://www.tribunnews.com/seleb/2023/05/12/darurat-kesehatan-global-dicabut-kemenkes-covid-19-belum-hilang>
- Samudi, S., Widodo, S., & Brawijaya, H. (2020). The K-Medoids Clustering Method for Learning Applications during the COVID-19 Pandemic. *Sinkron*, 5(1), 116. <https://doi.org/10.33395/sinkron.v5i1.10649>



Saputra, I., & Kristiyanti, D. A. (2022). *Machine Learning Untuk Pemula*. INFORMATIKA.

Sari, B. N., & Primajaya, A. (2020). Penerapan Clustering Dbscan Untuk Pertanian Padi di Kabupaten Karawang. *Jurnal Informatika dan Komputer*, 4(1), 51–58. <https://doi.org/10.30865/mib.v4i1.1784>

Wijaya, A. P., Premana, A., & Ramdhan, N. A. (2022). Penerapan Algoritma K-Means pada Klasterisasi Data kawalcovid19.id. *Prosiding Sains Nasional dan Teknologi*, 12(1), 479. <https://doi.org/10.36499/psnst.v12i1.7294>

Yadav, S. (2023, Juni 14). *Silhouette Coefficient Explained with a Practical Example: Assessing Cluster Fit*. Medium. [https://medium.com/@Suraj\\_Yadav/silhouette-coefficient-explained-with-a-practical-example-assessing-cluster-fit-c0bb3fdef719](https://medium.com/@Suraj_Yadav/silhouette-coefficient-explained-with-a-practical-example-assessing-cluster-fit-c0bb3fdef719)