

502

By Teknika Teknika

46
**PENGGUNAAN METODE K-MEANS DAN K-MEANS++
SEBAGAI CLUSTERING DATA COVID-19
DI PULAU JAWA**

2
(Naskah masuk: dd mmm yyyy, direvisi: dd mmm yyyy, diterima: dd mmm yyyy)

Abstrak

38
Virus Corona (Covid-19) merupakan penyakit menular yang dapat ditularkan antara hewan dan manusia. Pada akhir Desember 2019, virus itu teridentifikasi di Provinsi Wuhan, China. Saat ini, seluruh dunia sedang berjuang, mencegah, dan akhirnya menaklukkan penyebaran virus corona. Penelitian ini bertujuan untuk mengkluster data penyebaran Covid-19 di setiap kabupaten di Pulau Jawa sehingga menghasilkan kluster zona yang harus dilaksanakan PPKM berdasarkan kasus positif, vaksin dosis pertama, dan dosis kedua. Metode K-Means digunakan dengan cara menentukan jumlah cluster (K), mengatur pusat cluster secara acak, mengelompokkan data ke dalam cluster dengan jarak terpendek, menghitung pusat cluster, dan mengulangi langkah 2-4 sampai tidak ada data yang berpindah ke lokasi yang berbeda. K-Means++ digunakan dengan cara memilih secara acak nilai k pertama dari pusat cluster pertama titik data, mengelompokkan data berdasarkan jarak minimum ke centroid, memperbarui nilai titik centroid dengan menentukan rata-rata setiap cluster, dan ulangi langkah 2 dan 3 sampai tidak ada yang bergerak. Berdasarkan jumlah kasus positif, sembuh, dan meninggal, kasus tersebut dikategorikan. Setelah dilakukan pengelompokan dan mendapatkan kluster pada masing-masing kelompok, setiap kluster akan dievaluasi kualitasnya menggunakan koefisien silhouet untuk memilih yang terbaik. Hasil kajian tersebut diharapkan dapat mengungkap sejauh mana penyebaran virus Covid-19 di setiap kabupaten/kota di Pulau Jawa, serta cluster dengan skor Silhouette Coefficient tertinggi. Untuk hasil pengujian menggunakan Silhouette Coefficient, metode K-Means K=3 menghasilkan 0,825, K=4 menghasilkan 0,873, dan K=5 menghasilkan 0,862; untuk metode K-Means++, k=3 menghasilkan 0,822, K=4 menghasilkan 0,865, dan K = 5 menghasilkan 0,882. Hasil penelitian menunjukkan bahwa K-Means++ lebih unggul dalam memberikan informasi sejauh mana penyebaran virus covid-19, dan uji Silhouette Coefficient digunakan untuk menentukan kualitas cluster yang optimal.

Kata Kunci: K-Means, K-Means++, Clustering, Covid-19, Vaksin Pertama, Vaksin Kedua, Silhouette Coefficient.

**USING K-MEANS AND K-MEANS++ METHODS GROUP DATA
AS COVID-19 ON THE ISLAND OF JAVA**

Abstract

21
Corona virus (Covid-19) is an infectious disease that can be transmitted between animals and humans. At the end of December 2019, the virus was identified in Wuhan province, China. Right now, the whole world is fighting, preventing, and finally conquering the spread of the coronavirus. This study aims to cluster data on the spread of Covid-19 in every district on the island of Java so as to produce zones that must be implemented by PPKM based on positive cases, the first dose of vaccine, and the second dose vaccine. The K-Means method is used by determining the number of clusters (K), setting the cluster center arbitrarily, grouping data into clusters with the shortest distance, calculating the cluster center, and repeating steps 2-4 until no data moves to a different location. K-Means++ is used by randomly selecting the first k value from the center of the first cluster of data points, grouping the data based on the minimum distance to the centroid, updating the centroid point value by determining the average of each cluster, and repeating steps 2 and 3 until nothing moves. Based on the number of positive cases, recovered, and died, the cases are categorized. After grouping data into clusters in each group, each cluster will be evaluated for its quality using silhouette coefficients to choose the best. The results of the study are expected to reveal

26 the extent of the spread of the Covid-19 virus in every district/city on the island of Java, as well as the cluster with the highest Silhouette Coefficient score. For the test results using the Silhouette Coefficient, the K-Means method $K=3$ produces 0.825, $K=4$ produces 0.873, and $K=5$ produces 0.862; for the K-Means++ method, $k=3$ produces 0.822, $k=4$ produces 0.865, and $K=5$ produces 0.882. The results showed that K-Means++ was superior in providing information on the extent of the spread of the covid-19 virus, and the Silhouette Coefficient test was used to determine the optimal cluster quality.

Keywords: K-Means, K-Means++, Clustering, Covid-19, Vaksin Pertama, Vaksin Kedua, Silhouette Coefficient.

I. PENDAHULUAN

Virus Corona (Covid-19) adalah penyakit menular yang dapat menyerang manusia dan hewan. Virus ini ditemukan sekitar akhir Desember 2019 di Provinsi Wuhan, China. Saat ini, seluruh dunia sedang memerangi, menghentikan, dan mengalahkan penyebaran virus corona. Berdasarkan statistik World Health Organization hingga 27 Oktober 2021, jumlah kasus Covid-19 di Indonesia sebanyak 4.241.849 dengan 143.249 kasus kematian [1]. Akibat pandemi covid-19 ini mempengaruhi ekonomi dunia tidak hanya Indonesia saja.

Angka kasus positif COVID-19 sama dengan jumlah kasus yang dilaporkan di setiap kabupaten atau kota di Pulau Jawa. Pulau Jawa memiliki jumlah penduduk terbanyak, sehingga dapat mewakili setiap wilayah di Indonesia, dengan masing-masing kabupaten/kota memiliki ciri khasnya masing-masing. Akibatnya, kepadatan penduduk tiap kabupaten/kota berbeda-beda. Beberapa kabupaten/kota dengan skala kasus positif tertinggi akan membutuhkan perhatian khusus. Kota Jakarta Barat, Kota Jakarta Selatan, Kota Jakarta Utara, Kota Jakarta Pusat, Kab. Bogor, Kota Surabaya, dan Kota Jakarta Timur termasuk kabupaten/kota dengan prevalensi penularan virus corona yang tinggi. Hal ini menyulitkan pemerintah Indonesia untuk mendefinisikan contoh penanganan virus corona [2].

Selain itu, langkah pemerintah untuk membatasi penyebaran COVID-19 antara lain penetapan Program Pembatasan Kegiatan Masyarakat (PPKM). Jika jumlah kasus positif bertambah, PPKM akan dilaksanakan dengan menggunakan data variabel yang dikumpulkan dari kasus positif, dosis vaksin pertama, dan dosis vaksin kedua. Selain itu, lonjakan kasus baru ini disebabkan oleh masih banyaknya penduduk yang menolak menggunakan masker saat bepergian. Penggunaan masker secara langsung dapat menghambat penularan infeksi karena masker memenuhi peraturan kesehatan [3].

Virus Covid-19 telah menyerang berbagai daerah dalam jarak yang berdekatan. Warga di zona merah memiliki pilihan untuk berburu tanpa harus jauh-jauh atau keluar kota. Mereka yang dinyatakan positif virus Covid-19 dikarantina selama 14 hari dalam satu ruangan dengan pasien positif Covid-19 lainnya [4]. Penggunaan K-Means Cluster Analysis untuk mengorganisasikan data Covid-19 merupakan salah satu pendekatan analisis cluster non-hierarchical yaitu strategi cluster dimana jumlah cluster diatur secara manual. Tujuan dari cluster ini adalah untuk membagi item yang diamati

menjadi satu atau lebih kelompok bergantung pada karakteristiknya. Objek dengan kualitas yang sama akan ditempatkan dalam cluster yang sama, sedangkan objek dengan fitur yang berbeda akan dicampur dengan cluster lainnya [5]. Hasil investigasi oleh Haron, Sarjon Defit, dan Gunadi Widi Nurcahyo berjudul "Efisiensi Penggunaan Resep Dokter Spesialis Dengan Metode K-Means Clustering" menemukan bahwa resep dokter spesialis RSUD Kepulauan Meranti digunakan secara efisien. Hasil K-Means Clustering untuk tingkat efisiensi penggunaan resep dokter spesialis dari Apotek RSUD Kabupaten Kepulauan Meranti pada Desember 2019 menunjukkan terdapat tiga cluster, kategori menurut Fornas pertama (C0) dengan 18 data dari poliklinik anak dan bedah, 17 data dalam kategori anak sesuai Fornas [6].

Selain menggunakan Algoritma K-Means untuk mengelompokkan data Covid-19, penelitian ini juga menggunakan Algoritma K-Means++, yaitu mengorganisasikan data dengan menerapkan jumlah awal grup K dan mendasarkan setiap iterasi pada jarak terdekat. Iterasi adalah proses pengerjaan dari awal ke titik pusat baru dengan nilai yang sama dengan titik pusat sebelumnya; jika persyaratan ini belum terpenuhi, proses akan berlanjut sampai jumlah maksimum iterasi telah ditetapkan dari awal. Tujuan utama algoritma ini adalah untuk menempatkan titik-titik data ini sebagai pusat awal sejauh mungkin dari satu sama lain [7].

Clustering adalah teknik untuk mengkategorikan data berdasarkan kesamaan kualitas di seluruh kelompok data. Ada beberapa cara pengelompokan data, antara lain metode K-Means, metode K-Means++, pendekatan K-Mode, Hierarchical clustering, dan lain-lain. Penelitian ini menggunakan metode K-Means dan K-Means++ karena sebagian besar data dalam penelitian ini adalah numerik, sedangkan metode K-Modes lebih efektif dalam mengolah data kategorikal. Masing-masing metode memiliki kelebihan dan kekurangan, dan hasil clustering yang optimal dapat dipengaruhi oleh metode clustering yang digunakan, properti dari kumpulan data, struktur densitas data, jumlah data, dan jumlah cluster yang digunakan. Silhouette Coefficient merupakan teknik pengujian yang dapat digunakan untuk menilai kualitas kelompok yang dihasilkan dengan metode clustering jika hasilnya mendekati 1 maka kualitas cluster semakin baik, begitu juga sebaliknya. Jika hasilnya mendekati nol, cluster tersebut berkualitas buruk [8].

Penelitian ini bertujuan untuk mengelompokkan data sebaran Covid-19 di kabupaten/kota di Pulau Jawa sehingga

dapat membantu pembentukan kluster zona berdasarkan karakteristik positif, dosis vaksinasi pertama, dan dosis vaksin kedua. Tujuan dari metode *clustering* adalah untuk menempatkan data dalam set lengkap atau subset dan menetapkan ke grup atau subgroup [9].

Penelitian (Siti Azizatus Sholihah) menyarankan untuk memanfaatkan teknik K-Means serta menambahkan cara-cara ³³ bahan untuk mengembangkan metode K-Means, sehingga penelitian ini akan memanfaatkan metode K-Means dan K-Means++. Pendekatan yang lebih fleksibel digunakan dengan menggunakan teknik K-Means++, daripada yang lebih tradisional. Untuk mengidentifikasi perkembangan kasus Covid-19, penelitian ini akan mengelompokkan zona penyebaran virus Corona dan memanfaatkan temuan cluster yang digunakan untuk menawarkan informasi status zona berdasarkan perkembangan faktor positif, dosis vaksinasi pertama dan dosis vaksin kedua. Kami akan menggunakan pendekatan koefisien siluet untuk menganalisis temuan cluster untuk melihat mana yang memiliki kualitas terbaik [10].

II. METODOLOGI PENELITIAN

A. Subjek dan Objek Penelitian

Subjek pada penelitian ini adalah pengelompokan data pasien covid – 19. Sedangkan objek penelitiannya adalah data pasien covid – 19 dengan pembaruan data pada tanggal 09 Februari 2022 yang bersumber dari web <https://m.andrafarm.com/>.

B. Jenis Penelitian

Penelitian ini berdasarkan dari hasil riset dan data terbaru mengenai angkasipien yang sudah dinyatakan positif, jumlah vaksin dosis pertama dan vaksin dosis kedua pada setiap penduduk di pulau jawa. Oleh karena itu penelitian initermasuk jenis penelitian kuantitatif.

C. Sumber Data

Jenis data yang digunakan dalam penelitian ini adalah data sekunder. Data bisa didapatkan dari laman web <https://m.andrafarm.com/> dengan variable yang dapat dilihat pada Tabel 1

Table 1 Variabel Penelitian

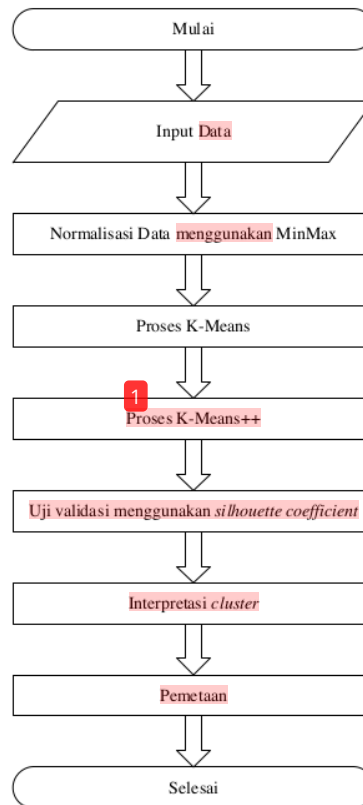
Variabel	Pengertian
X1	Total orang yang sudah dinyatakan positif Covid – 19
X2	Total orang yang sudah vaksin dosis pertama
X3	Total orang yang sudah vaksin dosis kedua

D. Desain Penelitian

Prosedur penelitian ini ditunjukkan dalam Gambar 1. Berikut penjelasan dari Gambar 1 :

1. Menginputkan data dalam bentuk table/matriks.

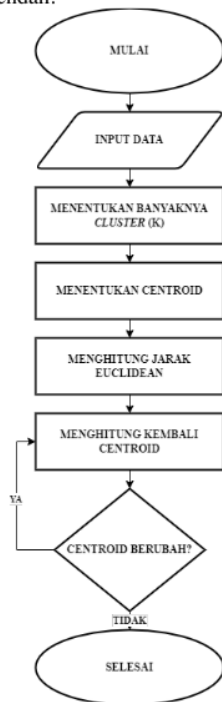
2. Normalisasi data dengan persamaan 2.1
 Dari data covid – 19 diatas akan di normalisasikan menggunakan Min- Max untuk mengurangi jarak antara nilai setiap variabel karena perbedaan skala.
3. Membuat cluster menggunakan metode non-herarki (K-Means dan K- Means++
 - a. Menentukan banyaknya nilai k yaitu banyaknya cluster. Nilai k akandiuji coba pada penelitian ini sebanyak 3 sampai 5
 Nilai k disini akan berdampak pada hasil yang akan dihitung nanti,semakin banyak nilai k akan mengurangi nilai error.
 - b. Memilih centroid (titik pusat) di setiap cluster
 - c. Menghitung jarak setiap objek dan setiap centroid dengan persamaan 2.2 dan 2.4



Gambar 1 Flowchart Penelitian

- d. Dihitung kembali nilai centroid untuk cluster yang baru terbeni dengan persamaan 2.2 dan 2.4 Mengulangi Langkah c dan d sampai tidak ada lagi perubahan posisi nilai centroid.
4. Hasil clustering akan divalidasi dengan silhouette coefficient dalam persamaan 2.5
 5. Tentukan kategori dari setiap cluster dengan melihat

- nilai centroidnya. Membuat kategori untuk mengimplementasikan hasil cluster yang telah dihitung.
6. Dengan memeriksa anggota cluster, Anda dapat mewarnai kabupaten/kota. Mewarnai peta sesuai dengan kategori yang sudah ditentukan, warna merah untuk zona tinggi, warna kuning untuk zona sedang, dan warna hijau zona rendah.



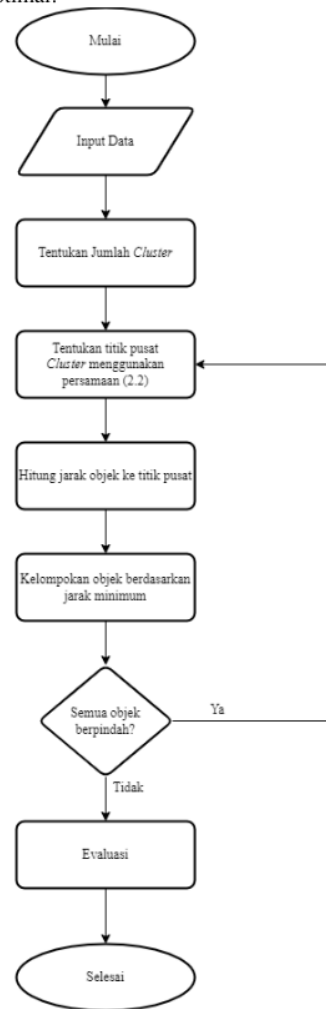
Gambar 3. 1 Flowchart K-Means

E. 3. Flowchart K-Means++

Adapun langkah – langkah pada algoritma K-Means++ adalah sebagai berikut:

1. Memilih nilai K untuk titik centroid dengan Persamaan (2.4) Setelah menentukan banyaknya nilai k dan dihitung menggunakan persamaan (2.4) akan mengeluarkan hasil yang akan digunakan untuk titik centroid (titik pusat).
2. Mengelompokkan berdasarkan jarak minimum terhadap centroid. Dikelompokkan hasil dari nilai k yang sudah dihitung sebelumnya dengan jarak minimum terhadap centroid (titik pusat).
3. Memperbarui titik centroid dengan mencari rata – rata setiap cluster Setelah mengelompokkan nilai centroid (titik pusat) dengan jarak minimum, step ini mencari rata – rata dari setiap cluster untuk dijadikan nilai centroid (titik pusat).
4. Mengulangi langkah 2 dan 3 sampai semua objek

- tidak berpindah.
5. Menguji validasi menggunakan silhouette coefficient di persamaan (2.5) Setelah nilai centroid semua tidak berpindah akan diuji menggunakan silhouette coefficient dan mendapatkan persentase nilai yang optimal.

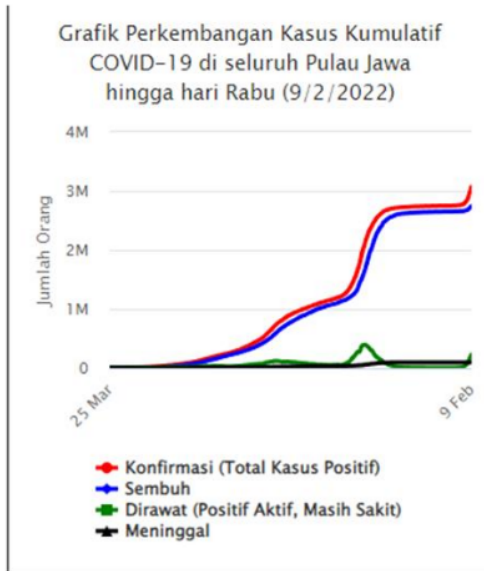


Gambar 3. 2 Flowchart K-Means++

a. Penjelasan Data

Data dari MAndrafarm merupakan 25 mpulan data sekunder dari berbagai website resmi setiap provinsi, kabupaten, dan kota yang ada di Indonesia, bisa diakses secara gratis pada laman web <https://m.andrafarm.com/>. Data tersebut

digunakan untuk analisis cluster dengan objek kabupaten dan kota pada Pulau Jawa berdasarkan 3 variabel, yaitu kasus positif, vaksin pertama, dan, vaksin kedua.



Gambar 2 Grafik Penyebaran Covid-19 Setiap Kabupaten dan Kota di Pulau Jawa

Gambar 4.1 menunjukkan epidemi virus Covid-19 di Indonesia. Jika ternyata semakin hari semakin meningkat, maka perlu dilakukan analisa daerah mana saja yang rawan dan aman dari Covid-19

b. Normalisasi Data

Sebelum pengelompokan, data dinormalisasi menggunakan perhitungan *Min – Max Normalization* yang ditunjukkan pada Persamaan 2.1. Di bagian ini menggambarkan proses normalisasi salah satu objek dari input data variabel X1 di Kota Tangerang Selatan dengan max = 203252 dan min = 1359. Di bawah ini adalah proses normalisasi untuk variabel X1 yang range-nya adalah Kota Tangerang Selatan.

Di bawah ini adalah proses normalisasi variabel X1 di Kota Tangerang Selatan dengan margin 0 hingga 1.

$$\begin{aligned}
 X^1 &= \frac{X - X_{min}}{X_{max} - X_{min}} \\
 &= \frac{47364 - 1359}{203252 - 1359} \\
 &= \frac{46005}{201893} \\
 &= 0.227868
 \end{aligned}$$

Dengan menggunakan rumus diatas akan mendapatkan data normalisasi yang bisa dilihat pada Tabel 2.

Table 2 Normalisasi Data

Kabupaten/kota	X1	X2	X3
Kota Tangerang Selatan	0.227868	0.291192	0.316020
Kota Tangerang	0.222732	0.438048	0.447272
Kab. Tangerang	0.168634	0.617048	0.625351
Kota Cilegon	0.061929	0.071801	0.081838
Kab. Serang	0.042334	0.136434	0.145711
Kab. Lebak	0.038857	0.204924	0.168744
Kota Serang	0.035083	0.136434	0.132965
Kab. Pandeglang	0.027841	0.202095	0.145877
Kota Jakarta Timur	0.114531	0.761681	0.849941
Kota Jakarta Selatan	1.000000	0.903166	1.000000

Table 2 adalah hasil data yang dinormalisasi menggunakan rumus *Min-Max Normalization*.

c. Proses K-Means

Algoritma *K-Means* berusaha untuk mengklasifikasikan data ke dalam kelompok berdasarkan sifat bersama. Data yang digunakan adalah data yang dinormalisasi. Prosedur *K-Means* dibagi menjadi beberapa tahapan sebagai berikut:

1. Menentukan nilai k

Pada tahap ini, peneliti melakukan tes 3 cluster sampai 5 cluster untuk mencari berapa jumlah cluster yang optimal. Mengenai perhitungan dibantu dengan program python, peneliti akan menguji 3 cluster dengan kelompok tinggi, sedang, dan rendah. Sedangkan 4 cluster maka akan dikelompokkan berdasarkan sangat tinggi, tinggi, sedang, dan rendah. Yang terakhir 5 cluster dikelompokkan berdasarkan sangat tinggi, tinggi, sedang, sangat aman, dan aman

2. Menentukan centroid awal

Pada tahap ini peneliti mendapatkan centroid awal secara acak yang bisa dilihat pada Tabel 3.

Table 3 Centroid awal

Cluster	C1	C2	C3
3	0.59342077	0.74905316	0.76830423
4	0.05000647	0.17249937	0.1717042
5	0.14397775	0.48489708	0.54648138

3. Menghitung jarak menggunakan *Euclidean Distance*

Langkah selanjutnya adalah menentukan jarak dari setiap titik data ke centroid setelah centroid ditentukan secara acak. Tabel 4 sampai Tabel 6 menampilkan

hasil perhitungan jarak Euclidean dari percobaan 3, 4, 5 cluster

6
Table 4 Data hasil perhitungan jarak euclidean pada percobaan 3 cluster

Kabupaten/Kota	C1	C2	C3
Kota Tangerang Selatan	0.740153	0.257972	0.312525
Kota Tangerang	0.580686	0.419867	0.135053
Kab. Tangerang	0.467231	0.646135	0.155859
Kota Cilegon	1.101085	0.135492	0.627115
Kab. Serang	1.032773	0.045113	0.540717
Kab. Lebak	0.981371	0.034416	0.481789
Kota Serang	1.044365	0.054992	0.551616
Kab. Pandeglang	1.003224	0.045103	0.503933
Kota Jakarta Timur	0.485963	0.900725	0.411782
Kota Jakarta Selatan	0.492687	1.456858	1.055179

Table 5 Data hasil perhitungan jarak euclidean pada percobaan 4 cluster

Kabupaten/Kota	C1	C2	C3	C4
Kota Tangerang Selatan	0.340098	0.345041	0.740153	0.179211
Kota Tangerang	0.511377	0.160983	0.580686	0.316684
Kab. Tangerang	0.739994	0.123033	0.467231	0.538278
Kota Cilegon	0.052065	0.664008	1.101085	0.241205
Kab. Serang	0.050340	0.577891	1.032773	0.152448
Kab. Lebak	0.115443	0.519256	0.981371	0.094669
Kota Serang	0.039545	0.588876	1.044365	0.162685
Kab. Pandeglang	0.101833	0.541446	1.003224	0.119022
Kota Jakarta Timur	0.994099	0.377057	0.485963	0.793478
Kota Jakarta Selatan	1.538339	1.026399	0.492687	1.365640

1
Table 6 Data hasil perhitungan jarak euclidean pada percobaan 5 cluster

Kabupaten/Kota	C1	C2	C3	C4	C5
Kota Tangerang Selatan	0.340098	0.345041	0.740153	0.179211	
Kota Tangerang	0.511377	0.160983	0.580686	0.316684	
Kab. Tangerang	0.739994	0.123033	0.467231	0.538278	
Kota Cilegon	0.052065	0.664008	1.101085	0.241205	
Kab. Serang	0.050340	0.577891	1.032773	0.152448	
Kab. Lebak	0.115443	0.519256	0.981371	0.094669	
Kota Serang	0.039545	0.588876	1.044365	0.162685	
Kab. Pandeglang	0.101833	0.541446	1.003224	0.119022	
Kota Jakarta Timur	0.994099	0.377057	0.485963	0.793478	
Kota Jakarta Selatan	1.538339	1.026399	0.492687	1.365640	

Kota Jakarta Timur	0.994099	0.158236	0.789072	0.384109	0.721601
Kota Jakarta Selatan	1.538339	0.774423	1.362299	1.005530	0.375401

6
Hasil perhitungan jarak Euclidean pada uji coba tiga cluster, empat cluster, dan lima cluster yang memiliki hasil pengelompokan masing-masing data yang memiliki fitur yang sama ditunjukkan pada Tabel 4.3 sampai dengan Tabel 4.5. Tabel tersebut sampai pada kesimpulan bahwa data dikelompokkan atau dimasukkan ke dalam cluster 1 jika paling dekat dengan centroid 1, dan seterusnya.

4. Menentukan centroid baru

Langkah selanjutnya adalah menghitung centroid sekali lagi dengan mencari rata-rata dari setiap cluster yang identik. Hasil centroid baru ditunjukkan pada Tabel 7. Untuk menentukan centroid sekali lagi, yaitu menghasilkan centroid baru dengan rata-rata setiap cluster.

32
Table 7 Centroid baru

Uji Coba	Pusat	X ₁	X ₂	X ₃
3	0	0.593421	0.749053	0.768304
	1	0.050006	0.172499	0.171704
	2	0.143978	0.484897	0.546481
4	0	0.033637	0.109380	0.104160
	1	0.154025	0.505072	0.576512
	2	0.593421	0.749053	0.768304
5	0	0.033637	0.109380	0.104160
	1	0.239609	0.858044	0.860356
	2	0.068200	0.247846	0.252898
3	3	0.181742	0.504977	0.572227
	4	0.824970	0.686680	0.748159

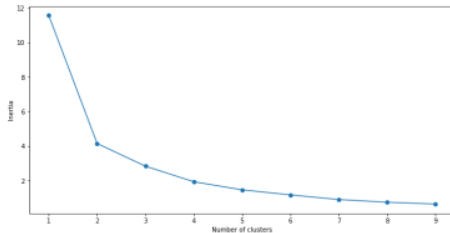
Operasi clustering telah berakhir karena centroid baru pada Tabel 7 sama dengan centroid lama pada Tabel 3.

d. Proses K-Means++

1. Menentukan nilai k

Pada tahap ini, peneliti melakukan tes 3 cluster sampai 5 cluster untuk mencari berapa jumlah cluster yang optimal. Mengenai perhitungan dibantu dengan program python, peneliti akan menguji 3 cluster dengan kelompok tinggi, sedang, dan rendah.

Sedangkan 4 cluster maka akan dikelompokkan berdasarkan sangat tinggi, tinggi, sedang, dan rendah. Yang terakhir 5 cluster dikelompokkan berdasarkan sangat tinggi, tinggi, sedang, sangat aman, dan aman. Berikut gambar 4.2 percobaan cluster 1-10 menggunakan *Elbow Method*.



Gambar 3 Elbow Method

1

- Menentukan centroid awal
Pada tahap ini peneliti mendapatkan centroid awal yang bisa dilihat pada Tabel 4.3

Table 8 Centroid awal

Uji Coba	C1	C2	C3
3	0.05000647	0.17249937	0.1717042
4	0.06819959	0.24784561	0.25289828
5	0.068009	0.24536864	0.24922218

- Menghitung jarak menggunakan *Euclidean Distance*
Langkah selanjutnya adalah menentukan jarak dari setiap titik data ke centroid setelah nilai centroid yang sudah didapatkan. Tabel 4.3 menampilkan hasil perhitungan jarak Euclidean dari percobaan 3, 4, 5 cluster.

Table 9 Data hasil perhitungan jarak euclidean pada percobaan 3 cluster

Kabupaten/Kota	C1	C2	C3
Kota Tangerang Selatan	0.257972	0.291335	0.717517
Kota Tangerang	0.419867	0.118030	0.547636
Kab. Tangerang	0.646135	0.176224	0.410869
Kota Cilegon	0.135492	0.605904	1.078398
Kab. Serang	0.045113	0.519813	1.006365
Kab. Lebak	0.034416	0.461186	0.952879
Kota Serang	0.054992	0.530779	1.017916
Kab. Pandeglang	0.045103	0.483469	0.974877
Kota Jakarta Timur	0.900725	0.433455	0.416539
Kota Jakarta Selatan	1.456858	0.66720	0.545188

Table 10 Data hasil perhitungan jarak euclidean pada percobaan 4 cluster

Kabupaten/Kota	C1	C2	C3	C4
Kota Tangerang Selatan	0.177080	0.356396	0.340098	0.740153
Kota Tangerang	0.312792	0.170222	0.511377	0.580686
Kab. Tangerang	0.533964	0.111107	0.739994	0.467231
Kota Cilegon	0.245545	0.677173	0.052065	1.101085

4

Kab. Serang	0.156750	0.591194	0.050340	1.032773
Kab. Lebak	0.098920	0.532545	0.115443	0.981371
Kota Serang	0.167013	0.602188	0.039545	1.044365
Kab. Pandeglang	0.123189	0.554722	0.101833	1.003224
Kota Jakarta Timur	0.789072	0.365056	0.994099	0.485963
Kota Jakarta Selatan	1.362299	1.015330	1.538339	0.492687

1

Table 11 Data hasil perhitungan jarak euclidean pada percobaan 5 cluster

Kabupaten/Kota	C1	C2	C3	C4	C5
Kota Tangerang Selatan	0.179211	0.753725	0.340098	0.785977	0.321684
Kota Tangerang	0.316684	0.628342	0.511377	0.589337	0.139874
Kab. Tangerang	0.538278	0.580632	0.739994	0.344011	0.144504
Kota Cilegon	0.241205	1.106228	0.052065	1.120641	0.641180
Kab. Serang	0.152448	1.048832	0.050340	1.034579	0.555443
Kab. Lebak	0.094669	1.005254	0.115443	0.972210	0.497178
Kota Serang	0.162685	1.060660	0.039545	1.044820	0.566510
Kab. Pandeglang	0.119022	1.026862	0.101833	0.992771	0.519524
Kota Jakarta Timur	0.793478	0.644455	0.994099	0.158236	0.400623
Kota Jakarta Selatan	1.365640	0.448132	1.538339	0.774423	1.037842

Hasil perhitungan jarak Euclidean pada uji coba dua cluster, tiga cluster, dan empat cluster yang memiliki hasil pengelompokan masing-masing data yang memiliki fitur yang sama ditunjukkan pada Tabel 4.8 sampai dengan Tabel 4.10. Tabel tersebut sampai pada kesimpulan bahwa data dikelompokkan atau dimasukkan ke dalam cluster 1 jika paling dekat dengan centroid 1, dan seterusnya.

- Menentukan centroid baru

Langkah selanjutnya adalah menghitung centroid sekali lagi dengan mencari rata-rata dari setiap cluster yang identik. Hasil dari centroid baru ditunjukkan pada Tabel 12. Untuk menentukan centroid baru, yaitu untuk mendapatkan centroid baru dengan rata-rata setiap cluster.

Table 12 Centroid baru

Uji Coba	Pusat	X ₁	X ₂	X ₃
3	C1	0.593421	0.749053	0.768304
	C2	0.050006	0.172499	0.171704
	C3	0.143978	0.484897	0.546481
4	C1	0.158818	0.513718	0.585709
	C2	0.033637	0.109380	0.104160
	C3	0.593421	0.749053	0.768304
	C4	0.068200	0.247846	0.252898
5	C1	0.068009	0.245369	0.249222
	C2	0.739057	0.670467	0.719675
	C3	0.033637	0.109380	0.104160
	C4	0.239609	0.858044	0.860356
	C5	0.156494	0.489034	0.559422

e. *Silhouette Coefficient*

Setelah menyelesaikan tahapan pengerjaan K-Means dan K-Means++ pengujian selanjutnya akan menguji hasil dari pengerjaan kedua metode tersebut dan melihat hasil cluster terbaik score nya saat diuji menggunakan Silhouette Coefficient berikut hasil dari kedua metode bisa di lihat pada Tabel 13 dan 14.

Table 13 Hasil pengujian setiap cluster dari K-Means

K	3	4	5
SC	0.825	0.873	0.862

Table 14 Hasil pengujian setiap cluster dari K-Means++

K	3	4	5
SC	0.822	0.865	0.882

Table 13 dan 14 merupakan hasil pengujian antara metode *K-Means* dan *K-Means++* diantara kedua metode tersebut yang mendapatkan hasil pengujian tertinggi adalah *K-Means++* dengan score 0.882 merupakan Struktur yang dihasilkan kuat.

f. Hasil dan Penjelasan

1. Penjelasan

Centroid akhir cluster yang ideal dapat diidentifikasi dalam analisis clustering menggunakan K-Means dengan membandingkan pola pembentukan masing-

masing centroid pada Tabel 3 dan 7 untuk K-Means++ pembentukan masing – masing centroid pada Tabel 8 dan 12 dengan centroid akhir yang dipilih.

Table 15 Pola setiap centroid K-Means++

Pusat	X ₁	X ₂	X ₃	Kategori Cluster
C1	Rendah	Rendah	Rendah	Aman
C2	Sangat Tinggi	Sangat Tinggi	Sangat Tinggi	Sangat Rawan
C3	Sangat Rendah	Sangat Rendah	Sangat Rendah	Sangat Aman
C4	Tinggi	Tinggi	Tinggi	Rawan
C5	Cukup Rendah	Cukup Rendah	Cukup Rendah	Cukup Aman

Table 15 menyimpulkan bahwa C3 memiliki pola centroid Sangat Rendah, sehingga sangat aman untuk kasus positif Covid-19. Karena C1 memiliki pola centroid yang rendah, maka dianggap aman untuk pasien positif Covid-19. Karena pola centroid rendah C5, itu dianggap cukup aman untuk kasus positif Covid-19. Karena pola centroid C2 yang sangat tinggi, diklasifikasikan sebagai sangat sensitif terhadap kasus positif Covid-19. Karena pola centroid C4 yang tinggi, maka tergolong lebih berpeluang menemui kasus positif Covid-19.

Data nama – nama kabupaten dan kota yang telah dikelompokkan berdasarkan cluster yang sudah dikerjakan menggunakan K-Means++ dengan score pengujian tertinggi dapat dilihat pada Tabel 16.

Table 16 Hasil cluster kabupaten dan kota di Pulau Jawa

Kelompok	Jumlah Anggota	Kabupaten/Kota
Sangat Aman	46	Kab. Bondowoso, Kota Blitar, Kab. Pandeglang, Kota Serang, Kab. Lumajang, Kota Madiun, Kab. Kota Madiun, Kab. Magetan, Kota Kediri, Kab. Banjarnegara, Kota Mojokerto, Kab. Pacitan, Kota Cilegon, Kab. Ngawi, Kab. Pamekasan, Kota Pasuruhan, Kab. Ponorogo, Kota Probolinggo, Kab. Sampang, Kab. Situbondo, Kab.

		Sumenep, Kab. Trenggalek, Kab. Kepulauan Sel., Kab. Kab. Bangkalan, Kota Pekalongan, Kab. Pengandaran, Kab. Batang, Kab. Blora. Kota Banjar, Kab. Purwakarta, Kota Sukabumi, Kota Cirebon, Kab. Kudus, Kota Cimahi, Kota Magelang, Kab. Pekalongan, Kab. Rembang, Kota Salatiga, Kab. Purworejo, Kab. Kulon Progo, Kab. Gunung Kidul, Kota Tegal, Kab. Wonosobo, Kab. Temanggung
Cukup Aman	16	Kota Bekasi, Kab. Sidoarjo, Kab. Cilacap, Kota Tangerang, Kota Jember, Kab. Garut, Kota Semarang, Kab. Cianjur, Kab. Bekasi, Kab. Karawang, Kota Depok, Kab. Malang, Kab. Bandung, Kab. Cirebon, Kab. Sukabumi, Kab. Tangerang
Aman	48	Kota Tangerang Purbalingga, Kab. Pemalang, Kab. Tuban, Kab. Pati, Kab. Magelang, Kab. Kendal, Kab. Kebumen, Kab. Karanganyar, Kab. Jepara, Kab. Grobogan, Kab. Demak, Kab. Brebes, Kab. Sukoharjo, Kota Surakarta, Kab. Tegal, Kab. Wonogiri, Kab. Probolinggo, Kab. Pasuruhan, Kab. Nganjuk, Kab. Mojokerto, Kota Malang, Kab. Lamongan, Kab. Kediri, Kab. Boyolali, Kab. Jombang, Kab. Gresik, Kab. Bojonegoro, Kab. Blitar, Kab.

		Banyuwangi, Kota Yogyakarta, Kab. Sleman, Kab. Bantul, Kab. Jember, Kab. Banyumas, Kab. Tulungagung, Kota Tasikmalaya, Kab. Indramayu, Kab. Kuningan, Kab. Majalengka, Kab. Ciamis, Kab. Subang, Kab. Lebak, Kab. Tasikmalaya, Kab. Sumedang
Sangat Rawan	4	Kota Jakarta Barat, Kota Jakarta Selatan, Kota Jakarta Utara, Kota Jakarta Pusat
Rawan	3	Kab. Bogor, Kota Surabaya, Kota Jakarta Timur

Tabel 16 menyimpulkan bahwa Kota Jakarta Barat, Kota Jakarta Selatan, Kota Jakarta Utara, Kota Jakarta Pusat merupakan kota yang sangat rawan akan Covid – 19 dan Kab. Bogor, Kota Surabaya, Kota Jakarta Timur merupakan kabupaten dan kota yang rawan akan Covid – 19 dikarenakan pada kabupaten/kota tersebut akan cepat dalam penyebarannya, sedangkan 110 kabupaten/kota lainnya dikategorikan aman akan Covid – 19 sehingga tingkat penularan dari kabupaten/kota tersebut akan kecil.

Ada tindakan yang dapat dilakukan untuk meningkatkan kesadaran masyarakat, seperti 5M, di kabupaten/kota yang aman terhadap COVID-19 (Menjaga Jarak, Cuci Tangan.). Selain itu, setiap orang yang masuk dari transportasi yang sama harus diperiksa dan karantina 14 hari harus diberlakukan bagi mereka yang menimbulkan bahaya.

Provinsi rawan Covid-19 dapat mengambil tindakan pencegahan yang sama dengan daerah aman, selain meminta masyarakat menahan diri untuk tidak menghadiri pertemuan yang tidak penting dan hanya bepergian untuk alasan yang sangat diperlukan.

IV. KESIMPULAN

1. Pendekatan *K-Means++*, sebagaimana ditunjukkan dari validasi cluster menggunakan *Silhouette*, menghasilkan hasil cluster terbaik berdasarkan metode *K-means* dan *K-Means++*, dengan rata-rata *Silhouette Coefficient* (SC) sebesar 0,882 pada k = 5. Struktur kuat terdiri dari nilai SC sebesar 0,882. Hal ini disebabkan fakta bahwa objek lebih dekat dengan anggota cluster mereka sendiri rata-rata daripada anggota cluster lain
2. Berdasarkan metode *K-Means++*, mendapatkan hasil untuk Kota Jakarta Barat, Kota Jakarta Selatan, Kota Jakarta Utara, Kota Jakarta Pusat tergolong sangat rawan oleh penyebaran Covid – 19 untuk Kab. Bogor, Kota Surabaya, Kota Jakarta Timur tergolong rawan oleh penyebaran Covid – 19. Untuk 110 kabupaten/kota lainnya dikategorikan aman dari kasus penyebaran Covid – 19

REFERENSI

-
- [1] A. Solichin and K. Khairunnisa, "Klasterisasi Persebaran Virus Corona (Covid-19) Di DKI Jakarta Menggunakan Metode K-Means," *Fountain Informatics J.*, vol. 5, no. 2, p. 52, 2020, doi: 10.21111/fij.v5i2.4905.
- [2] D. D. Darmansah, "Analisis Penyebaran Penularan Virus Covid-19 di Provinsi Jawa Barat Menggunakan Algoritma K-Means Clustering," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 3, pp. 1188–1199, 2021, doi: 10.35957/jatisi.v8i3.1034.
- [3] D. T. Utari, "Analisis Karakteristik Wilayah Transmisi Covid-19 dengan Menggunakan Metode K-Means Clustering," *J. Media Tek. dan Sist. Ind.*, vol. 5, no. 1, p. 25, 2021, doi: 10.35194/jmtsi.v5i1.1220.
- [4] S. A. Sholihah, "Analisis Cluster untuk Pemetaan Data Kasus Covid-19 di Indonesia Menggunakan K-Means," 2021.
- [5] D. N. P. Sari and Y. L. Sukestiyarno, "Analisis Cluster dengan Metode K-Means pada Persebaran Kasus Covid-19 Berdasarkan Provinsi di Indonesia," *Prism. Pros. Semin. Nas. Mat.*, vol. 4, pp. 602–610, 2021, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>
- [6] Sharon, S. Defit, and G. W. Nurcahyo, "Tingkat Efisiensi Penggunaan Resep Dokter Spesialis Menggunakan Metode K-Means Clustering," *J. Inf. dan Teknol.*, vol. 3, pp. 121–127, 2021, doi: 10.37034/jidt.v3i3.118.
- [7] D. A. I. C. Dewi and D. A. K. Pramita, "Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali," *Matrix J. Manaj. Teknol. dan Inform.*, vol. 9, no. 3, pp. 102–109, 2019, doi: 10.31940/matrix.v9i3.1662.
- [8] Santo Faskafri, "Bab 1 pendahuluan," *Pelayanan Kesehat.*, no. 2015, pp. 3–13, 2020, [Online]. Available: <http://repository.usu.ac.id/bitstream/123456789/23790/4/Chapter1.pdf>
- [9] R. A. Indraputra and R. Fitriana, "K-Means Clustering Data COVID-19," *J. Tek. Ind.*, vol. 10, no. 3, p. 3, 2020.
- [10] F. M. FALAHI, "Penerapan Metode Clustering Untuk Pengelompokan Mahasiswa Potensial Drop Out Menggunakan Algoritma K-Means ++," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, p. 64, 2019.

21 %SIMILARITY INDEX

PRIMARY SOURCES

1	digilib.uinsby.ac.id Internet	249 words — 5%
2	www.journals.ukitoraja.ac.id Internet	82 words — 2%
3	123dok.com Internet	67 words — 1%
4	media.neliti.com Internet	65 words — 1%
5	ejournal.ikado.ac.id Internet	44 words — 1%
6	repository.ub.ac.id Internet	38 words — 1%
7	repositori.kemdikbud.go.id Internet	36 words — 1%
8	jurnal.wicida.ac.id Internet	34 words — 1%
9	ppid.distanbun.jatengprov.go.id Internet	33 words — 1%

10	Susana Limanto, Hendra Dinata. "Aplikasi Multimedia Untuk Media Pembelajaran Menggambar Grafik Fungsi Dengan Operasi Penskalaan, Pergeseran, dan Pencerminan", Teknika, 2018 Crossref	32 words — 1%
11	text-id.123dok.com Internet	19 words — < 1%
12	id.123dok.com Internet	18 words — < 1%
13	dinkesjatengprov.go.id Internet	14 words — < 1%
14	qdoc.tips Internet	14 words — < 1%
15	Lestari Sinaga, Abdullah Ahmad, Muhammad Safii. "PENERAPAN DATA MINING PADA JUMLAH PELANGGAN PERUSAHAAN AIR BERSIH MENURUT PROVINSI MENGGUNAKAN METODE K-MEANS CLUSTERING", Jurnal RESISTOR (Rekayasa Sistem Komputer), 2019 Crossref	13 words — < 1%
16	blkbekasi.kemnaker.go.id Internet	13 words — < 1%
17	doku.pub Internet	12 words — < 1%
18	moam.info Internet	12 words — < 1%
19	ppid.jakarta.go.id Internet	12 words — < 1%

20	docplayer.info Internet	11 words — < 1%
21	info.rdi.ku.ac.th Internet	11 words — < 1%
22	ojs.palcomtech.ac.id Internet	10 words — < 1%
23	Vincentius Riandaru Prasetyo. "Searching Cheapest Product on Three Different E-Commerce Using K-Means Algorithm", 2018 International Seminar on Intelligent Technology and Its Applications (ISITIA), 2018 Crossref	9 words — < 1%
24	ojs.unm.ac.id Internet	9 words — < 1%
25	repository.upi.edu Internet	9 words — < 1%
26	www.asjp.cerist.dz Internet	9 words — < 1%
27	www.bbc.com Internet	9 words — < 1%
28	www.ibn.ac.id Internet	9 words — < 1%
29	www.semanticscholar.org Internet	9 words — < 1%
30	Dian Juliarti Bantam, Dewi Tri Resky Yanti. "Perguruan Tinggi Berbasis Islam Sebagai Learning	8 words — < 1%

Organization: Survei Berdasarkan Persepsi Pegawai", Journal of Psychological Perspective, 2021

Crossref

31 Muhammad Rizki Wardhana, Agung Triayudi, Nur Hayati. "Analisis Faktor Calon Nasabah PT. Bank Central Asia dalam Pembuatan Rekening Online menggunakan Metode K-Means Clustering Studi Kasus Wisma Asia BCA", Jurnal JTik (Jurnal Teknologi Informasi dan Komunikasi), 2022
8 words — < 1%
Crossref

32 Ogi Suhendra, Muhammad Arif Tiro, Ruliana. "Ordinary Least Square Method in Multiple Regression Analysis to Estimating Coefficients of Factors Affecting Human Development Index", ARRUS Journal of Mathematics and Applied Science, 2022
8 words — < 1%
Crossref

33 Suhardi Rustam, Haditsah Annur. "AKADEMIK DATA MINING (ADM) K-MEANS DAN K-MEANS K-NN UNTUK MENGELOMPOKAN KELAS MATA KULIAH KOSENTRASI MAHASISWA SEMESTER AKHIR", ILKOM Jurnal Ilmiah, 2019
8 words — < 1%
Crossref

34 apiv2.tonixcomputer.co.id
Internet
8 words — < 1%

35 core.ac.uk
Internet
8 words — < 1%

36 ejournal.upbatam.ac.id
Internet
8 words — < 1%

37 jagowebdev.com
Internet
8 words — < 1%

38	journal.perbanas.id Internet	8 words — < 1%
39	journal.stmikjayakarta.ac.id Internet	8 words — < 1%
40	journal.um-surabaya.ac.id Internet	8 words — < 1%
41	muthebogara.blog Internet	8 words — < 1%
42	nasional.tempo.co Internet	8 words — < 1%
43	silemlit21.unila.ac.id Internet	8 words — < 1%
44	sismatik.nusaputra.ac.id Internet	8 words — < 1%
45	Nurul Hidayati. "APPLICATION OF CLUSTERING ANALYSIS TO DATA DISTRIBUTION OF COVID-19 IN BENGKULU PROVINCE", BAREKENG: Jurnal Ilmu Matematika dan Terapan, 2022 Crossref	7 words — < 1%
46	jutif.if.unsoed.ac.id Internet	7 words — < 1%
47	Raden Muhammad Ali Satria, Resty Varia Tutupoho, Djazuly Chalidyanto. "Analisis Faktor Risiko Kematian dengan Penyakit Komorbid Covid-19", Jurnal Keperawatan Silampari, 2020 Crossref	6 words — < 1%

48 U. Maulik, S. Bandyopadhyay. "Performance evaluation of some clustering algorithms and validity indices", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002
Crossref 6 words — < 1%

49 infolayananmasyarakat.blogspot.com
Internet 6 words — < 1%

50 jurnal.fikom.umi.ac.id
Internet 6 words — < 1%

EXCLUDE QUOTES OFF
EXCLUDE BIBLIOGRAPHY ON

EXCLUDE SOURCES OFF
EXCLUDE MATCHES OFF