



ANALISIS KLASTER DENGAN METODE K-MEANS PADA PERSEBARAN KASUS COVID-19 BERDASARKAN DESA DI KECAMATAN KEMANG – BOGOR

Yuni Permata Sari^{*}, Choirul Basir² 

^{1,2} Program Studi Matematika,

Fakultas Matematika Dan Ilmu Pengetahuan Alam-Universitas Pamulang

ARTICLE INFO

Article history:

Received Nov 05, 2024

Revised Dec 12, 2024

Accepted Dec 24, 2024

Available online Jan 28, 2025

Kata Kunci :

Analisis Klaster, Covid-19,
K-means

Keywords:

Cluster analysis, Covid-19,
K-means



This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

Copyright ©2024 by Author. Published by LPPM Universitas Islam Syekh-Yusuf

ABSTRAK

Coronavirus (Covid-19) merupakan jenis virus baru yang ditemukan pada manusia di Provinsi Wuhan, China pada akhir tahun 2019. Indonesia merupakan salah satu negara yang terjangkit virus Covid-19. Tujuan penelitian ini adalah untuk memperoleh pengelompokan persebaran kasus Covid-19 berdasarkan desa di Kecamatan Kemang-Bogor dengan menerapkan metode K-Means. Data yang digunakan dalam penelitian ini merupakan data sekunder tentang variabel penyebaran Covid-19 pada bulan Januari hingga Agustus 2021. Pengelompokan dilakukan berdasarkan parameter jumlah kasus terkonfirmasi positif, sembuh, dan meninggal. Pada penelitian ini, untuk melakukan pengelompokan data digunakan metode K-Means dan metode pengukuran jarak Euclidean. Berdasarkan hasil penelitian menunjukkan bahwa metode K-Means Clustering dapat diterapkan untuk mengelompokkan desa-desa di Kecamatan Kemang-Bogor dengan klaster optimal terletak pada $k = 3$ yang berarti klaster yang direkomendasikan adalah 3 klaster dengan klaster 1 berisi 6 desa, klaster 2 berisi 1 desa, dan klaster 3 berisi 2 desa.

ABSTRACT

Coronavirus (Covid-19) is a new type of virus found in humans in Wuhan Province, China at the end of 2019. Indonesia is one of the countries infected with the Covid-19 virus. The purpose of this study is to group the distribution of Covid-19 cases by village in Kemang-Bogor District by applying the K-Means method. The data used in this study is secondary data on the variables of the spread of Covid-19 from January to August 2021. The grouping was carried out based on the parameters of the number of confirmed positive cases, recovered, and died. In this study, to group the data, the K-Means method and the Euclidean distance measurement method were used. Based on the results of the study, it is shown that the K-Means Clustering method can be applied to group villages in Kemang-Bogor District with the optimal cluster located at $k = 3$ which means that the recommended cluster is 3 clusters with cluster 1 containing 6 villages, cluster 2 containing 1 village, and cluster 3 containing 2 villages.

*Corresponding author.

E-mail addresses: yuniprmatasari@gmail.com (First Author)

PENDAHULUAN

Virus *Severe Acute Respiratory Syndrome Coronavirus – 2* (SARS – Cov – 2) atau orang-orang biasa menyebutnya dengan sebutan Covid-19, saat ini masih menjadi suatu masalah yang menggemparkan seluruh negara termasuk salah satunya adalah Indonesia. Virus tersebut pertama kali ditemukan pada kota Wuhan Hubei China diakhir tahun 2019 yang membentuk kepanikan karena telah memakan banyak korban jiwa. Berdasarkan data *World Health Organization* (WHO), hingga 10 Agustus 2021 terdapat total 203.295.170 kasus terkonfirmasi dari seluruh dunia dengan rata-rata tingkat kematian sebesar 2% (WHO, 2021). Sementara itu WHO pertama kali melaporkan adanya kasus Covid-19 di Indonesia yaitu bertepatan pada tanggal 2 Maret 2020 dengan jumlah 2 kasus terkonfirmasi.

Untuk mengendalikan pandemi Covid-19, pemerintah telah melaksanakan Pembatasan Sosial Berskala Besar (PSBB) yang berlangsung pada sejumlah wilayah di Indonesia. Karena dianggap kurang signifikan dalam mengurangi kasus pandemi Covid-19 di Indonesia, maka setelah PSBB pemerintah mengambil kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM). PPKM ialah kebijakan pemerintah Indonesia sejak awal tahun 2021 yang berlangsung di beberapa wilayah yang menjadi titik penyebaran Covid-19, yakni di Pulau Jawa dan Bali.

Kabupaten Bogor merupakan kabupaten yang berada di daerah Jawa Barat, dan termasuk salah satu daerah yang menjadi Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM). Pada kamis 08 Juli 2021, satgas penanganan Covid-19 Kabupaten Bogor telah merilis bahwa kasus baru Covid-19 telah mencapai 22.670 kasus, dengan 2.341 kasus yang terkonfirmasi aktif, 20.137 yang terkonfirmasi sembuh, dan 186 kasus yang terkonfirmasi meninggal termasuk didalamnya ada tenaga kesehatan.

Tingginya penyebaran Covid-19 di Kabupaten Bogor ini maka diperlukan adanya pengelompokan daerah-daerah yang terdampak Covid-19 untuk mengetahui tingkat kemiripan karakteristik atau kriteria dari setiap daerah yang berada di kabupaten Bogor. Daerah yang memiliki karakteristik yang sama atau hampir sama akan terkumpul dalam suatu kelompok. Hal ini bertujuan untuk memberikan informasi terkait Covid-19 kepada masing-masing daerah. Salah satu cara yang dapat dilakukan dalam pengelompokan ini yaitu analisis klaster.

Analisis klaster dalam penelitian ini dilakukan untuk mengetahui tingkat persebaran Covid-19 dengan tingkat kemiripan karakteristik pada masing-masing desa di Kecamatan Kemang-Bogor yang memiliki total sembilan desa dengan masing-masing desa memiliki jumlah penduduk dan luas wilayah yang berbeda-beda. Untuk mengetahui karakteristik tersebut ada beberapa aspek yang bisa diukur seperti terkonfirmasi aktif, sembuh, dan meninggal.

Penelitian ini menggunakan metode Analisis *K-Means Clustering*, dimana ada beberapa peneliti juga menggunakan metode yang sama untuk kasus Covid-19 ini. Penelitian yang dilakukan oleh Siti Azizah Sholihah yang menggunakan metode *K-Means* untuk mengelompokkan provinsi berdasarkan pada data persebaran kasus Covid-19 di Indonesia. Penelitian tersebut menghasilkan dua kelompok yaitu kelompok rawan dan kelompok aman, dengan klaster optimal berada pada nilai $k = 2$ dan *Silhouette Coefficient* (SC) = 0,74 yang berarti struktur klaster termasuk kuat (Sholihah, 2021).

METODOLOGI PENELITIAN

Jenis Penelitian

Metode pendekatan yang digunakan dalam penelitian ini yaitu metode pendekatan kuantitatif, dimana metode pendekatan kuantitatif adalah metode penelitian yang

berlandaskan pada filsafat positivisme, digunakan untuk meneliti pada populasi atau sampel tertentu, pengumpulan data menggunakan instrumen penelitian, analisis data bersifat kuantitatif/statistik, dengan tujuan untuk menguji hipotesis yang telah diterapkan (Sugiono, 2011).

Penelitian ini menggunakan metode pendekatan kuantitatif karena penelitian ini berdasarkan hasil riset dan data terbaru mengenai kasus Covid-19 yang terdiri dari konfirmasi positif, sembuh, dan meninggal. Serta berdasarkan jumlah penduduk dan kepadatan penduduk yang ada di Kecamatan Kemang-Bogor. Sehingga data yang terbentuk dalam penelitian ini berwujud dalam bentuk angka yang kemudian diolah dengan menggunakan analisis multivariat yaitu analisis *K-Means Clustering*.

Variabel Penelitian

Variabel penelitian merupakan suatu atribut atau sifat atau nilai dari orang, objek atau kegiatan yang mempunyai variasi tertentu yang ditetapkan oleh peneliti untuk dipelajari dan kemudian ditarik kesimpulan (Sugiono, 2011).

Dalam penelitian ini menggunakan metode analisis kluster, dimana sifat dari analisis kluster adalah interdependensi yang berarti tidak adanya variabel dependen maupun variabel independen, tetapi hanya ada variabel-variabel yang akan dicari kemiripannya. Oleh karena itu, variabel-variabel yang digunakan dalam penelitian ini yaitu :

Tabel 3.1 Variabel Penelitian

Variabel	Keterangan
X_1	Kasus Positif
X_2	Kasus Sembuh
X_3	Kasus Meninggal

HASIL DAN PEMBAHASAN

Hasil Data Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder mengenai kasus Covid-

19 yang berada di Kecamatan Kemang-Bogor, peneliti mendapatkan data dari puskesmas-puskesmas yang menangani kasus Covid-19 tersebut. Puskesmas Jampang menangani tiga desa yaitu desa Jampang, desa Pondok Udik, dan desa Tegal. Kemudian puskesmas Kemang menangani 6 desa yaitu desa Kemang, Parakan Jaya, Semplak Barat, Pabuaran, Bojong, dan Atang Sanjaya. Data yang digunakan adalah data yang diakses pada bulan Januari hingga Agustus 2021 dengan jumlah data yaitu 9 dari total 9 desa, dan dengan 3 variabel yaitu variabel kasus positif, sembuh, dan meninggal.

Tabel 4.1 Total data kasus Covid-19 Kec.Kemang Januari – Agustus 2021

Desa	X_1	X_2	X_3
Kemang	207	207	7
Jampang	371	290	6
Parakan Jaya	217	210	9
Semplak Barat	186	184	7
Tegal	647	563	5
Desa	X_1	X_2	X_3
Pondok Udik	194	201	1
Pabuaran	105	110	2
Bojong	195	172	8
Atang sanjaya	319	292	2

Adapun statistik deskriptif dari data tersebut dapat ditunjukkan seperti pada tabel 4.2 berikut.

Tabel 4.2 Descriptive Statistics

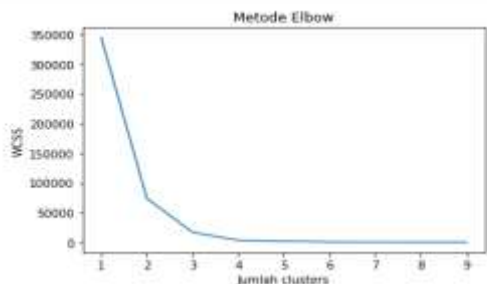
Variabel	Min	Mean	Std.Dev	Maks
Positif	105	271.22	160.926	647
Sembuh	110	247.67	130.946	563
Meninggal	1	5.22	2.906	9

Berdasarkan tabel 4.2 di atas diperoleh bahwa desa dengan kasus positif terendah adalah desa Pabuaran dengan total kasus positif sebesar 105 kasus, sedangkan untuk kasus positif tertinggi adalah desa Tegal dengan total 647 kasus. Kemudian untuk desa dengan tingkat kesembuhan terendah adalah desa Pabuaran

dengan total kasus sembuh adalah 110 dan untuk desa dengan tingkat tertinggi dari kasus sembuh ini adalah desa Tegal dengan total 563 kasus sembuh. Begitupun dengan kasus yang terkonfirmasi meninggal, desa dengan tingkat terendah dalam kasus meninggal ini adalah desa Pondok Udik yaitu 1 kasus, dan kasus meninggal terbesar adalah desa Parakan Jaya dengan total 9 kasus yang meninggal.

Proses Analisis Klaster Menggunakan Metode K-Means

Langkah pertama dalam analisis *K-Means Clustering* adalah menentukan jumlah klaster yang akan dibentuk. Dalam penelitian ini digunakan metode *elbow* untuk menentukan jumlah dan mendapatkan hasil klaster yang terbaik dan optimum seperti yang ditunjukkan pada gambar 4.1 berikut yang didapatkan dari hasil running menggunakan Google Colab.



Grafik 4.1 Penentuan Jumlah Cluster Terbaik

Berdasarkan metode *elbow* diperoleh klaster optimum yaitu sebanyak 3 klaster, yang dapat diketahui dari nilai *K* yang turun secara drastis dan membentuk siku yaitu terdapat pada *K*=3.

Langkah kedua yaitu menentukan titik pusat *centroid*. Karena klaster yang terbentuk adalah 3 klaster, maka untuk titik pusat yang akan digunakan yaitu sebanyak 3 *centroid* yang dipilih secara random oleh peneliti. Dalam metode *K-Means* ini peneliti menentukan sendiri

centroid yang akan digunakan, maka dipilih data ke-1, ke-5, dan data ke-9 sebagai titik pusat *centroid*. Untuk lebih jelasnya dapat dilihat pada tabel 4.3 berikut.

Tabel 4.3 Centroid Iterasi 1

Cluster	X ₁	X ₂	X ₃
C1	207	207	7
C2	647	563	5
C3	319	292	2

Langkah ketiga dalam analisis *K-Means cluster* yaitu menentukan jarak terdekat ke *centroid*. Untuk menentukan jarak terdekat ke *centroid* akan dihitung kemiripan antara dua hingga tiga objek yaitu (X₁, X₂), (X₁, X₃), (X₂, X₃), dan (X₁, X₂, X₃) dengan menggunakan jarak *euclidean* dalam menghitung kemiripan tersebut dapat dihitung dengan rumus *euclidean* berikut.

$$d_{ij} = \sqrt{\sum_{k=1}^p (X_{jk} - X_{ik})^2} \tag{4.2}$$

Berikut adalah perhitungan menggunakan rumus jarak *euclidean* tersebut dengan menggunakan data pada tabel 4.1.

1. Menghitung jarak *euclidean* antara variabel positif dan sembuh (X₁, X₂).

Misalkan akan dihitung jarak antara data ke-1 (Desa Kemang) dengan ketiga *centroid* yang telah ditentukan.

$$d_{1,1} = \sqrt{(207 - 207)^2 + (207 - 207)^2} = 0$$

$$d_{1,2} = \sqrt{(207 - 647)^2 + (207 - 563)^2} = 565,98$$

$$d_{1,3} = \sqrt{(207 - 319)^2 + (207 - 292)^2} = 140,602$$

Dari perhitungan tersebut dapat diperoleh hasil seperti pada tabel 4.4 berikut.

Tabel 4.4 Iterasi 1 Jarak Euclidean X1,X2

Desa	C1	C2	C3	Cluster
------	----	----	----	---------

Kemang	0	565.9823	140.60228	1
Jampang	183.807	388.2074	52.038447	3
Parakan Jaya	10.44031	556.3353	130.87399	1
Semplak Barat	31.14482	596.7931	171.32717	1
Tegal	565.9823	0	425.47033	2
Pondok Udik	14.31782	579.8733	154.61565	1
Pabuaran	140.7587	706.3802	280.92704	1
Bojong	37	597.6496	172.55724	1
Atang sanjaya	140.6023	425.4703	0	3

Dari tabel di atas dapat diketahui bahwa masing-masing desa telah terbentuk menjadi tiga kluster. Kluster pertama dengan enam desa, kluster kedua dengan satu desa, dan kluster ketiga dengan dua desa.

Kemudian langkah selanjutnya adalah akan memastikan apakah kluster yang telah terbentuk mengalami perpindahan atau tidak. Untuk mengetahui hal tersebut kita dapat menghitung *centroid* baru yaitu dengan menghitung rata-rata dari masing-masing kluster. Misal akan dihitung *centroid* baru pada kluster satu, yaitu

dengan langkah perhitungan sebagai berikut.

$$C_{1,1} = \frac{207 + 217 + 186 + 194 + 105 + 195}{6} = \frac{1.104}{6} = 184$$

$$C_{1,2} = \frac{647}{1} = 647$$

$$C_{1,3} = \frac{371 + 319}{2} = \frac{690}{2} = 345$$

Dari perhitungan tersebut dapat diperoleh *centroid* baru yang dapat ditunjukkan pada tabel 4.5 berikut.

Tabel 4.5 Centroid baru

Penentuan Cluster baru	X ₁	X ₂	X ₃
Centroid baru ke-1	184	180.67	5.67
Centroid baru ke-2	647	563	5
Centroid baru ke-3	345	291	4

Setelah mendapatkan *centroid* baru, langkah selanjutnya yaitu akan menghitung kembali jarak *euclidean* pada iterasi ke-2 dengan cara yang sama seperti pada iterasi

pertama hanya titik pusat *centroid* nya diganti dengan nilai *centroid* yang baru. Maka diperoleh hasil jarak *euclidean* yang ditunjukkan pada tabel 4.6 berikut.

Tabel 4.6 Iterasi 2 Jarak Euclidean X₁, X₂

Desa	X ₁	X ₂	C1	C2	C3	Cluster
Kemang	207	207	34.963	565.98	161.55	1
Jampang	371	290	216.62	388.21	26.019	3
Parakan Jaya	217	210	44.153	556.34	151.48	1
Semplak Barat	186	184	3.8873	596.79	191.65	1

Tegal	647	563	600.46	0	406.43	2
Pondok Udik	194	201	22.659	579.87	175.79	1
Pabuaran	105	110	105.99	706.38	300.6	1
Bojong	195	172	14.004	597.65	191.47	1
Atang sanjaya	319	292	174.99	425.47	26.019	3

Dari tabel di atas dapat dilihat bahwa pada iterasi kedua hasil kluster pada masing-masing desa sama dengan hasil kluster pada iterasi pertama, yang berarti kluster yang terbentuk tidak mengalami perubahan. Sehingga pengklasteran sudah mencapai final.

- Menghitung jarak *euclidean* antara variabel positif dan meninggal (X_1, X_3).

Dari perhitungan tersebut diperoleh hasil seperti pada tabel 4.7 berikut.

Tabel 4.7 Iterasi 1 Jarak Euclidean X_1, X_3

Desa	C1	C2	C3	Cluster
Kemang	0	440.005	112.112	1
Jampang	164.003	276.002	52.1536	3
Parakan Jaya	10.198	430.019	102.24	1
Semplak Barat	21	461.004	133.094	1
Tegal	440.005	0	328.014	2
Pondok Udik	14.3178	453.018	125.004	1
Pabuaran	102.122	542.008	214	1
Bojong	12.0416	452.01	124.145	1
Atang sanjaya	112.112	328.014	0	3

Dari tabel di atas dapat dilihat ternyata kluster yang terbentuk antara variabel positif dan meninggal hasilnya sama dengan kluster pada variabel positif dan sembuh.

Langkah selanjutnya sama seperti mencari jarak *euclidean* variabel X_1 dan X_2 di atas yaitu menghitung centroid baru. Karena kluster yang terbentuk pada variabel X_1 dan X_3 sama dengan variabel X_1 dan X_2 maka titik pusat *centroid* akan sama dengan *centroid* baru yang telah terbentuk pada

Misalkan akan dihitung jarak antara data ke-2 (Desa Jampang) dengan ketiga *centroid* yang telah ditentukan pada tabel 4.3 di atas.

$$d_{2,1} = \sqrt{(371 - 207)^2 + (6 - 7)^2} = 164,003$$

$$d_{2,2} = \sqrt{(371 - 647)^2 + (6 - 5)^2} = 276,002$$

$$d_{2,3} = \sqrt{(371 - 319)^2 + (6 - 2)^2} = 52,1536$$

perhitungan di atas yaitu yang ditunjukkan pada tabel 4.5.

Selanjutnya akan dihitung kembali apakah kluster mengalami perubahan atau tidak, maka didapatkanlah hasil pada tabel 4.8 berikut.

Tabel 4.8 Iterasi 2 Jarak Euclide X_1, X_3

Desa	X_1	X_3	C1	C2	C3	Cluster
Kemang	207	7	23.039	440	138.03	1
Jampang	371	6	187	276	26.077	3
Parakan Jaya	217	9	33.168	430.02	128.1	1
Semplak Barat	186	7	2.4037	461	159.03	1
Tegal	647	5	463	0	302	2
Pondok Udik	194	1	11.035	453.02	151.03	1
Pabuaran	105	2	79.085	542.01	240.01	1
Bojong	195	8	11.245	452.01	150.05	1
Atang sanjaya	319	2	135.05	328.01	26.077	3

Dari tabel di atas dapat dilihat ternyata hasil kluster setelah menggunakan *centroid* baru hasilnya sama dengan kluster pada iterasi pertama. Karena kluster tidak mengalami perubahan maka perhitungan dicukupkan pada iterasi kedua.

- Menghitung jarak *euclidean* antara variabel sembuah dan meninggal (X_2, X_3)

Langkah-langkah serta proses perhitungan untuk mencari jarak *euclidean* pada variabel sembuah dan meninggal ini sama seperti proses yang telah dilakukan di atas pada saat mencari jarak *euclidean* variabel (X_1, X_2) dan (X_1, X_3). Maka didapatkan hasil jarak *euclidean* pada iterasi 1 yaitu sebagai berikut.

Tabel 4.9 Iterasi 1 Jarak Euclide X_2, X_3

Desa	C1	C2	C3	Cluster
Kemang	0	356.01	85.147	1
Jampang	83.006	273	4.4721	3
Parakan Jaya	3.6056	353.02	82.298	1
Semplak Barat	23	379.01	108.12	1
Tegal	356.01	0	271.02	2
Pondok Udik	8.4853	362.02	91.005	1
Pabuaran	97.129	453.01	182	1
Bojong	35.014	391.01	120.15	1
Atang sanjaya	85.147	271.02	0	3

Pada tabel di atas terlihat hasil kluster pada masing-masing desa. Kluster 1 terdiri dari 6 desa, kluster 2 terdiri dari 1 desa, dan kluster 3 terdiri dari 2 desa. Hasil tersebut masih sama dengan hasil pada iterasi 1

variabel (X_1, X_2) dan (X_1, X_3), yang berarti nilai titik pusat *centroid* baru masih menggunakan *centroid* yang terdapat pada tabel 4.5 sebelumnya.

Tabel 4.10 Iterasi 2 Jarak Euclide X2,X3

Desa	X2	X3	C1	C2	C3	Cluster
Kemang	207	7	26.367	356.01	84.054	1
Jampang	290	6	109.33	273	2.2361	3
Parakan Jaya	210	9	29.522	353.02	81.154	1
Semplak Barat	184	7	3.5901	379.01	107.04	1
Tegal	563	5	382.33	0	272	2
Pondok Udik	201	1	20.862	362.02	90.05	1
Pabuaran	110	2	70.762	453.01	181.01	1
Bojong	172	8	8.9753	391.01	119.07	1
Atang sanjaya	292	2	111.39	271.02	2.2361	3

4. Menghitung jarak *euclide* antara variabel positif, sembuh dan meninggal (X_1, X_2, X_3)
 Misalkan akan dihitung jarak *euclide* antara data ke-3 (Desa Parakan Jaya) dengan masing-masing *centroid* yang telah ditentukan sebelumnya pada tabel 4.3.

$$d_{3,1} = \sqrt{(217 - 207)^2 + (210 - 207)^2 + (9 - 7)^2} = 10,63$$

$$d_{3,2} = \sqrt{(217 - 647)^2 + (210 - 563)^2 + (9 - 5)^2} = 556,35$$

$$d_{3,3} = \sqrt{(217 - 319)^2 + (210 - 292)^2 + (9 - 2)^2} = 131,06$$

Dari perhitungan tersebut diperoleh hasil seperti pada tabel 4.9 berikut.

Tabel 4.11 Iterasi 1 Jarak Euclide X1,X2,X3

Desa	C1	C2	C3	Cluster
Kemang	0	565.99	140.69	1
Jampang	183.81	388.21	52.192	3
Parakan Jaya	10.63	556.35	131.06	1
Semplak Barat	31.145	596.8	171.4	1
Tegal	565.99	0	425.48	2

Pondok Udik	15.524	579.89	154.62	1
Pabuaran	140.85	706.39	280.93	1
Bojong	37.014	597.66	172.66	1
Atang sanjaya	140.69	425.48	0	3

Hasil pada tabel di samping menunjukkan hasil yang sama pada pengklasteran sebelumnya, yaitu pada saat menghitung jarak *euclide* antara variabel (X_1, X_2), (X_1, X_3) dan (X_2, X_3). Maka langkah selanjutnya yaitu menghitung *centroid* baru seperti pada perhitungan sebelumnya, karena cluster menunjukkan hasil yang sama maka *centroid* baru pun tetap menggunakan hasil yang terdapat pada tabel 4.5 di atas.

Langkah selanjutnya akan dipastikan apakah klaster yang telah terbentuk mengalami perpindahan atau tidak, akan dipastikan dengan cara menghitung jarak *euclide* dengan pusat *centroid* yang baru. Maka didapatkan lah hasil seperti pada tabel 4.12 berikut.

Tabel 4.12 Iterasi 2 Jarak Euclide X_1, X_2, X_3

Desa	X_1	X_2	X_3	C1	C2	C3	<i>Distance</i>	Cluster
Kemang	207	207	7	34.989	565.99	161.58	34.989	1
Jampang	371	290	6	216.62	388.21	26.096	26.096	3
Parakan Jaya	217	210	9	44.278	556.35	151.56	44.278	1
Semplak Barat	186	184	7	4.1096	596.8	191.67	4.1096	1
Tegal	647	563	5	600.46	0	406.43	0	2
Pondok Udik	194	201	1	23.135	579.89	175.81	23.135	1
Pabuaran	105	110	2	106.06	706.39	300.61	106.06	1
Bojong	195	172	8	14.197	597.66	191.51	14.197	1
Atang sanjaya	319	292	2	175.02	425.48	26.096	26.096	3

Hasil iterasi kedua pada tabel di atas menunjukkan bahwa ternyata kluster yang terbentuk tidak mengalami perubahan dan masih sama dengan hasil pada iterasi pertama. Karena kluster tidak berubah pada iterasi kedua ini, maka hasil pengelompokan sudah mencapai *final cluster* dan proses dicukupkan sampai pada iterasi kedua.

Dari keempat proses *clusterisasi* setiap variabel yaitu (X_1, X_2) , (X_1, X_3) , (X_2, X_3) dan (X_1, X_2, X_3) didapatkan hasil kluster yang sama dengan kluster 1 terdiri dari 6 desa, kluster 2 terdiri dari 1 desa, dan kluster 3 terdiri dari 2 desa. Maka dapat disimpulkan bahwa metode *K-Means Clustering* ini dapat diterapkan untuk mengelompokkan desa-desa berdasarkan kasus Covid-19 yang terjadi di Kecamatan Kemang-Bogor.

Hasil kluster yang terbentuk dari perhitungan manual yang telah dilakukan di atas ternyata berbanding lurus dengan hasil yang didapatkan menggunakan bantuan Google Colab, yaitu seperti dibawah ini.

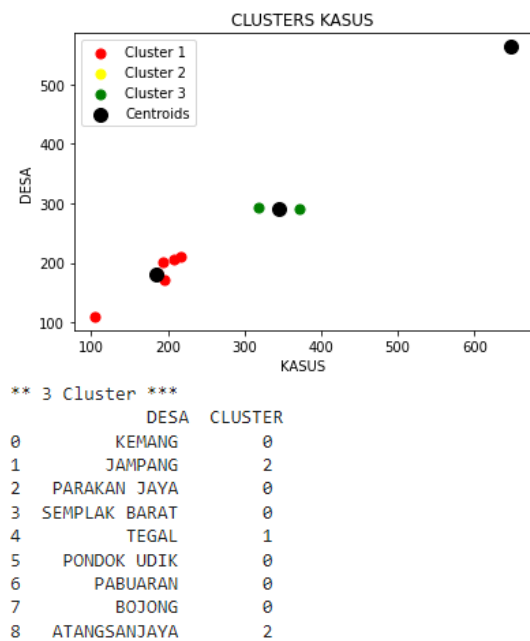
```
***** 3 Cluster Model *****
Cluster centers:
[[184.          180.66666667  5.66666666
 [647.          563.          5.
 [345.          291.          4.
Inertia (WCSS): 16548.666666666668
No. of iterations: 2
```

Gambar 4.1 Cluster Center

Gambar di atas menunjukkan bahwa pusat kluster atau titik pusat *centroid* sama dengan hasil yang ditunjukkan pada tabel 4.5 sebelumnya. Kemudian hasil iterasi yang didapatkan menunjukkan angka 2 yang berarti iterasi sama seperti perhitungan manual yaitu berhenti pada iterasi kedua.

Setelah dibuktikan ternyata hasil titik pusat *centroid* antara perhitungan manual

menggunakan Ms. Excel dengan bantuan Google Colab hasilnya sama, maka selanjutnya akan dibuktikan apakah hasil kluster pada masing-masing desa yang telah didapatkan akan sama dengan hasil running menggunakan Google Colab. Berikut hasil menggunakan Google Colab ditunjukkan pada gambar 4.3 berikut.



Gambar 4.2 Plot Cluster menggunakan Python

Dari gambar plot di atas menunjukkan hasil kluster yang sama pada masing-masing desa dengan hasil yang didapatkan pada perhitungan manual sebelumnya. Dengan kluster 1 ditunjukkan dengan warna merah, kluster 2 ditunjukkan dengan warna kuning, dan kluster 3 ditunjukkan dengan warna hijau. Angka 0 pada visualisasi di atas yang berarti angka 1 atau masuk ke kluster 1, kemudian angka 1 masuk ke kluster 2, dan angka 2 masuk ke kluster 3.

Pembahasan

Berdasarkan hasil running menggunakan Google Colab sebagaimana yang ditunjukkan pada gambar plot di atas, serta hasil yang didapatkan secara manual

*Corresponding author.

E-mail addresses: yuniprmatasari@gmail.com (First Author)

menggunakan Ms. Excel, kedua proses tersebut menunjukkan hasil yang sama pada setiap klaster. Dari proses analisis yang telah dilakukan ternyata penerapan metode *K-Means Clustering* dapat digunakan untuk mengelompokkan desa-desa di Kecamatan Kemang-Bogor dengan hasil sebagai berikut :

1. Klaster pertama beranggotakan enam desa yaitu Desa Kemang, Parakan Jaya, Semplak Barat, Pondok Udik, Pabuaran, dan Bojong. Dengan nilai *centroid* pada masing-masing variabel adalah 184 untuk variabel kasus positif (X_1); 180,67 untuk variabel kasus sembuh (X_2), dan 5,67 untuk nilai *centroid* pada variabel kasus meninggal (X_3).
2. Klaster kedua hanya beranggotakan satu desa yaitu Desa Tegal dengan nilai *centroid* pada masing-masing variabel adalah 647; 563; dan 5.
3. Klaster ketiga beranggotakan dua desa yaitu Desa Jampang dan Desa Atang sanjaya dengan nilai *centroid* pada masing-masing variabel adalah 345; 291; dan 4.

KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan menggunakan metode *K-Means Clustering*, maka dapat disimpulkan bahwa metode *K-Means Clustering* dapat diterapkan untuk mengelompokkan persebaran kasus Covid-19 berdasarkan desa di Kecamatan Kemang-Bogor, dengan hasil pengelompokkan yaitu klaster pertama terdiri dari 6 desa yaitu Desa Kemang, Parakan Jaya, Semplak Barat, Pondok Udik, Pabuaran, dan Bojong. Klaster kedua terdiri dari satu desa yaitu Desa Tegal. Sedangkan klaster ketiga terdiri dari dua desa yaitu Desa Jampang dan Desa Atang sanjaya.

DAFTAR PUSTAKA

- Bastian, N., & Lardo, C. (2018). Multiple stellar populations in globular clusters. *Annual review of Astronomy and Astrophysics*, 56, 83-136.
- Blair, R. A., Morse, B. S., & Tsai, L. L. (2017). *Public health and public trust: Survey evidence from the Ebola Virus Disease epidemic in Liberia*. *Social Science & Medicine*, 89-97.
- Fadhli. (2011). Analisis Cluster Untuk Pemetaan Mutu Pendidikan di Aceh. *Tesis*. PPs-UGM.
- Han, J., Kamber, M., Pei, J.: *Data Mining Concept and Techniques*, 3rd ed. Morgan Kaufmann-Elsevier, Amsterdam (2012)
- Kemenkes, R. I. (2019). Keputusan menteri kesehatan Republik Indonesia Nomor HK. 01.07/menkes/413/2020 tentang pedoman pencegahan dan pengendalian corona virus disease 2019 (Covid-19). *MenKes/413/2020*, 207.
- Kementerian Kesehatan, RI. (2020). *Frequently Asked Questions (FAQ) COVID-19*. Jakarta
- MacQueen, J.B. (1967). Some methods for classification and analysis of multivariate observations. In L.M. Le Cam & J. Neyman (Eds.), *Proceedings of the fofty Berkeley symposium on mathematical statistics and probability* (Vol. 1, pp. 281-297). California: University of California Press.
- Prof. Dr. Sugiono, (2011). *Metode Penelitian Pendidikan (Pendekatan Kuantitatif, Kualitatif, dan R&D)*; Alfabeta. Bandung.
- Randolph, H. E., & Barreiro, L. B. (2020). *Herd Immunity: Understanding COVID-19*. *Immunity*, 737-741.
- Sari, D. N. P., & Sukestiyarno, Y. L. (2021) Analisis Cluster Dengan Metode K-Means Pada Persebaran Kasus COVID-19 Berdasarkan Provinsi di Indonesia. In *PRISMA, Prosiding*

- Seminar Nasional Matematika* (Vol. 4, pp. 602-610).
- Satyahadewi, N., & Ediyanto, M. N. M. (2013). Pengklasifikasian karakteristik dengan metode K-Means Cluster analysis.
- Sholihah, S. A. (2021). *Analisis cluster untuk pemetaan data kasus Covid-19 di Indonesia menggunakan K-Means* (Doctoral dissertation, UIN Sunan Ampel Surabaya).
- Solichin, A., & Khairunnisa, K. (2020). Klasterisasi persebaran virus Corona (Covid-19) di DKI Jakarta menggunakan metode K-Means. *Fountain of Informatics Journal*, 5(2), 52-59.
- Suryahadi, A., Izzati, R. A., & Suryadarma, D. (2020). *The Impact of COVID-19 Outbreak on Poverty; An Estimation for Indonesia*. Jakarta: The SMERU Institute.
- Tan, P.N., Steinbach, M., Kumar, V. (2006) *Introduction to Data Mining*. Boston:Pearson Education.
- World Health Organization. (2021). Data Covid-19 di Indonesia. Website : <https://covid19.go.id/id>