

**ANALISIS CLUSTER SEBARAN COVID-19 MENGGUNAKAN
ALGORITMA K-MEANS CLUSTERING
(STUDI KASUS: PROVINSI JAWA BARAT)**

**CLUSTER ANALYSIS OF COVID-19 DISTRIBUTION
USING K-MEANS CLUSTERING ALGORITHM
(CASE STUDY: WEST JAVA PROVINCE)**

Ato Sugiharto¹, Betha Nurina Sari², Tesa Nur Padilah³

Universitas Singaperbangsa Karawang^{1,2,3}

ato.sugiharto17010@student.unsika.ac.id

ABSTRACT

Coronavirus disease (covid-19) has become a global concern after on January 20, 2020, three people were killed in the city of Wuhan, Hubei province, China. Covid-19 was first reported to have entered Indonesia on March 2, 2020, with two cases. This study aims to conduct a cluster analysis of the distribution of COVID-19 cases in West Java province as of April 1, 2021 with the variables of isolation, recovery, and death. By using the elbow method, the difference in SSE in each cluster, the silhouette graph, and the factoextra diagram, the optimum number of clusters is 3, the evaluation results show the Dunn index value = 0.4776, connectivity = 9.4738, and silhouette = 0.5839 (data structure reasoned). The clustering results show a good variance of 75.8%. Cluster 1 consists of 1 city/district, cluster 2 consists of 6 cities/districts, and cluster 3 consists of 20 cities/districts.

Keywords: Data Mining, Clustering, Covid-19, K-Means, KDD.

ABSTRAK

Coronavirus disease (covid-19) mulai menjadi perhatian masyarakat dunia setelah pada 20 Januari 2020 menewaskan tiga orang di kota Wuhan, provinsi Hubei, Tiongkok. Covid-19 pertama kali dilaporkan masuk ke Indonesia pada tanggal 2 Maret 2020 sejumlah dua kasus. Penelitian ini bertujuan untuk melakukan analisis cluster terhadap sebaran kasus covid-19 di provinsi Jawa Barat per tanggal 1 April 2021 dengan variabel isolasi, sembuh, dan meninggal. Dengan menggunakan elbow method, selisih SSE pada setiap cluster, grafik silhouette, serta diagram factoextra didapatkan jumlah cluster optimum yaitu 3, hasil evaluasi menunjukkan nilai indeks dunn = 0,4776, connectivity = 9,4738, dan silhouette = 0,5839 (struktur data beralasan). Hasil clustering menunjukkan variance yang baik yaitu sebesar 75,8%. Cluster 1 berjumlah 1 kota/kabupaten, cluster 2 berjumlah 6 kota/kabupaten, dan cluster 3 berjumlah 20 kota/kabupaten.

Kata Kunci: Data Mining, Clustering, Covid-19, K-Means, KDD.

PENDAHULUAN

Coronavirus disease (Covid-19) termasuk kedalam virus RNA dengan ukuran partikel sebesar 120-160 nm. Terdapat enam jenis *coronavirus* yang mampu menginfeksi manusia, yaitu *alphacoronavirus* NL63, *alphacoronavirus* 229E, *betacoronavirus* HKU1, *betacoronavirus* OC43, *Middle East Respiratory Syndrome Coronavirus* (MERS-CoV), dan *Severe Acute Respiratory Illness Coronavirus* (SARS-CoV) (Riedel, S., Morse, S., Mietzner,

T., Miler, S. J., Melnick., & Adelberg's Medical Microbiology., 2019).

Kemudahan penularan covid-19 membuat virus ini tersebar ke beberapa daerah di Indonesia dengan cepat. Penyebaran kasus covid

d-19 yang kian merata di daerah-daerah Indonesia membuat kepanikan serta kerugian bagi masyarakat, daerah, serta negara. Jawa Barat merupakan salah satu daerah yang terkena imbas dari penularan covid-19. Dalam penelitian Dwitri, Tampubolon, Prayoga, Zer, & Hartama (2020) menghasilkan bahwa Jawa Barat merupakan *cluster*

provinsi dengan tingkat kasus positif covid-19 dan kasus meninggal paling besar kedua setelah DKI Jakarta.

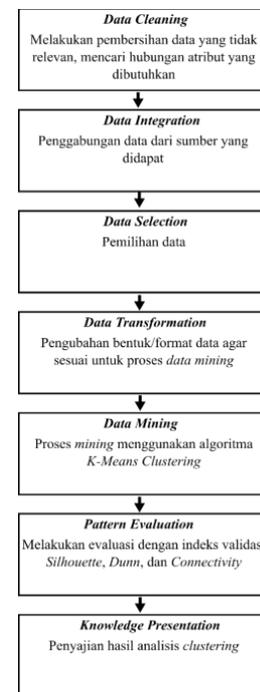
Clustering merupakan kegiatan untuk mengelompokkan data yang memiliki tingkat kemiripan (*similarity*) dengan data lainnya (Aditya, Jovian, & Sari, 2020; Yogi, dkk., 2020; Hertina, dkk., 2021). Penelitian Sindi, Ningse, Sihombing, Ilmi, & Hartama (2020) menghasilkan pengelompokan 34 record wilayah di Indonesia yang terinfeksi covid-19 dengan *clustering* terbaik sebanyak 3 *cluster* yang terdiri dari cluster pertama sebanyak 1 wilayah, *cluster* kedua sebanyak 2 wilayah, dan *cluster* ketiga sebanyak 31 wilayah. Penelitian dilakukan menggunakan algoritma *k-medoids* dengan tools rapidminer, variabel yang digunakan adalah kasus positif, sembuh, dan meninggal. Penelitian *clustering* kasus covid-19 di Indonesia selanjutnya dilakukan oleh Khotimah & Darsin (2020) menggunakan algoritma *Self Organizing Map* (SOM) menghasilkan 4 *cluster* yang terdiri dari *cluster* pertama sebanyak 1 provinsi, *cluster* kedua sebanyak 11 provinsi, *cluster* ketiga sebanyak 1 provinsi, dan *cluster* keempat sebanyak 21 provinsi (Sinaga & Yang, 2020).

METODE

Metode yang digunakan pada penelitian ini adalah KDD (*Knowledge Discovery in Database*) dengan menerapkan teknik *clustering* menggunakan algoritma *k-means*. Metode KDD dapat dilihat pada Gambar 1 (Liu, dkk., 2020).

KDD adalah salah satu metode pada *data mining* yang berfungsi untuk memperoleh pengetahuan dari *database*. Hasil pengetahuan yang diperoleh dari proses KDD kemudian dapat dimanfaatkan untuk keperluan pengambilan keputusan yang

menguntungkan atau membuat suatu kebijakan (Mardi, 2016; Plotnikova, dkk., 2020).



Gambar 1. Tahapan KDD

1. Data Cleaning

Pada fase pertama dataset dari penelitian dicari keterkaitan atau hubungan antara atribut-atribut yang dibutuhkan untuk proses analisis, di tahap ini juga dilakukan pembersihan data yaitu dengan membuang data yang tidak konsisten, *noise*, dan data yang tidak relevan. Proses *data cleaning* akan berpengaruh kepada proses *mining* nantinya karena jumlah banyaknya data dan kompleksitasnya berpengaruh terhadap performa teknik *data mining*.

2. Data Integration

Data integration atau integrasi data merupakan aktivitas melakukan penggabungan data dari sumber yang didapat, kemudian diolah menjadi tabel-tabel yang nantinya akan diproses pada tahap selanjutnya.

3. Data Selection

Pada tahap data selection adalah melakukan pemilihan data berdasarkan data yang diperoleh dari sumber data.

4. Data Transformation

Pada tahap data transformation dilakukan perubahan data ke dalam format yang sesuai untuk proses data mining.

5. Data Mining

Pada tahap data mining dilakukan pengaplikasian algoritma k-means menggunakan tools Rstudio.

6. Pattern Evaluation

Pattern evaluation adalah melakukan evaluasi terhadap hasil dari proses data mining, evaluasi pada penelitian ini yaitu menggunakan indeks validitas silhouette, dunn, dan connectivity.

7. Knowledge Presentation

Knowledge presentation merupakan penyajian pengetahuan atau visualisasi hasil dari penelitian dalam bentuk yang mudah dimengerti.

HASIL DAN PEMBAHASAN

Hasil dari penelitian yang dilakukan adalah melakukan analisis cluster dari kasus sebaran covid-19 di Jawa Barat pada bulan April 2021 dengan variabel isolasi, sembuh, dan meninggal. Software yang digunakan yaitu Rstudio.

1. Data Cleaning

Gambar 2 merupakan hasil summary dari dataset yang digunakan, pada hasil summary tersebut tidak ditemukan NA atau missing value, maka tahap penelitian dilanjutkan ke tahap berikutnya.

```
> summary(datacovid)
  Nama.kota.Kabupaten  Isolasi  Sembuh  Meninggal
Length:27             Min.   : 0    Min.   : 79   Min.   : 2.00
Class :character      1st Qu.: 272  1st Qu.: 1908 1st Qu.: 35.00
Mode  :character      Median : 564  Median : 3485 Median : 43.00
                Mean   :1227  Mean   : 7079 Mean   : 81.04
                3rd Qu.:1924  3rd Qu.: 9802 3rd Qu.:104.50
                Max.   :6635  Max.   :35545 Max.   :322.00
```

Gambar 2. Hasil Summary dataset

2. Data Integration

Pada tahap data integration adalah melakukan penggabungan semua dataset yang dibutuhkan ke dalam satu tabel. Pada penelitian ini hanya menggunakan satu tabel, maka tahap ini tidak dilakukan.

3. Data Selection

Penelitian ini menggunakan dataset sebaran kasus covid-19 di provinsi Jawa Barat terhitung sampai 1 April 2021 yang bersumber dari <https://pikobar.jabarprov.go.id/>. Atribut yang digunakan yaitu isolasi, sembuh, dan meninggal.

Tabel 1. Dataset Covid-19 Jabar 1 April 2021

Nama Kota/Kabupaten	Isolasi	Sembuh	Meninggal
Kab. Bogor	2.338	10.77	172
Kab. Sukabumi	225	4.438	34
Kab. Cianjur	813	1.961	2
Kab. Bandung	485	8.831	43
Kab. Garut	802	79	143
Kab. Tasikmalaya	281	1.915	60
Kab. Ciamis	511	2.473	55
Kab. Kuningan	505	3.888	16
Kab. Cirebon	744	574	194
Kab. Majalengka	418	1.902	40
Kab. Sumedang	191	1.922	40
Kab. Indramayu	206	402	43
Kab. Subang	1.964	2.296	14
Kab. Purwakarta	263	342	59
Kab. Karawang	3.583	12.00	237
Kab. Bekasi	4.088	20.81	60

Kab. Bandung Barat	564	3.815	37
Kab. Pangandaran	0	1.024	2
Kota Bogor	1.883	11.73	120
Kota Sukabumi	392	3.103	42
Kota Bandung	2.161	13.78	89
Kota Cirebon	887	3.485	36
Kota Bekasi	2.052	35.54	218
Kota Depok	6.635	35.45	322
Kota Cimahi	250	3.817	54
Kota Tasikmalaya	680	4.103	32
Kota Banjar	218	648	24

4. Data Transformation

Pada tahap ini dilakukan perubahan data agar memudahkan saat melakukan pengolahan data. Perubahan yang dilakukan adalah dengan melakukan standarisasi menggunakan fungsi *scale* terhadap nilai pada semua data, hal ini bertujuan agar beberapa variabel yang digunakan memiliki rentan nilai yang sama, tidak ada yang terlalu besar ataupun terlalu kecil sehingga membuat analisis menjadi lebih mudah dilakukan. Hasil *scalling* dapat dilihat pada Tabel 2.

Tabel 2. Hasil *Scalling*

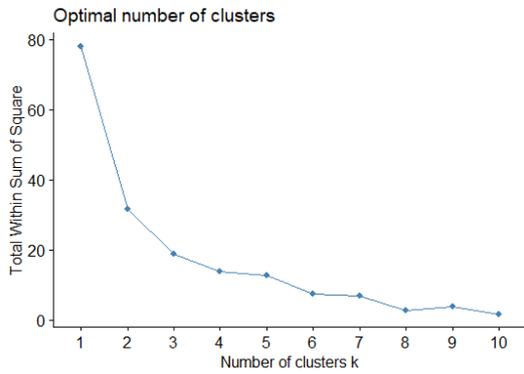
Nama Kota/Kabupaten	Isolasi	Sembuh	Meninggal
Kab. Bogor	0,73	0,38	1,12
Kab. Sukabumi	-0,66	-0,27	-0,57
Kab. Cianjur	-0,27	-0,53	-0,97
Kab. Bandung	-0,49	0,18	-0,46
Kab. Garut	-0,28	-0,73	0,76
Kab. Tasikmalaya	-0,62	-0,54	-0,26
Kab. Ciarnis	-0,47	-0,48	-0,32

Kab. Kuningan	-0,48	-0,33	-0,80
Kab. Cirebon	-0,32	-0,68	1,39
Kab. Majalengka	-0,53	-0,54	-0,50
Kab. Sumedang	-0,68	-0,54	-0,50
Kab. Indramayu	-0,67	-0,69	-0,47
Kab. Subang	0,49	-0,50	-0,83
Kab. Purwakarta	-0,64	-0,70	-0,27
Kab. Karawang	1,56	0,51	1,92
Kab. Bekasi	1,89	1,43	-0,26
Kab. Bandung Barat	-0,44	-0,34	-0,54
Kab. Pangandaran	-0,81	-0,63	-0,97
Kota Bogor	0,43	0,48	0,48
Kota Sukabumi	-0,55	-0,41	-0,48
Kota Bandung	0,62	0,70	0,098
Kota Cirebon	-0,22	-0,37	-0,55
Kota Bekasi	0,54	2,96	1,69
Kota Depok	3,57	2,95	2,97
Kota Cimahi	-0,64	-0,34	-0,33
Kota Tasikmalaya	-0,36	-0,31	-0,60
Kota Banjar	-0,67	-0,68	-0,70

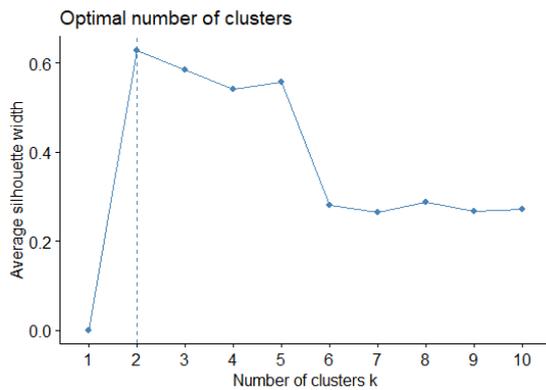
5. Data Mining

Pada tahap ini dilakukan pengolahan data menggunakan algoritma *k-means*. Jumlah cluster yang digunakan adalah hasil dari penentuan dengan *elbow method* dan SSE (*Sum Square of Error*).

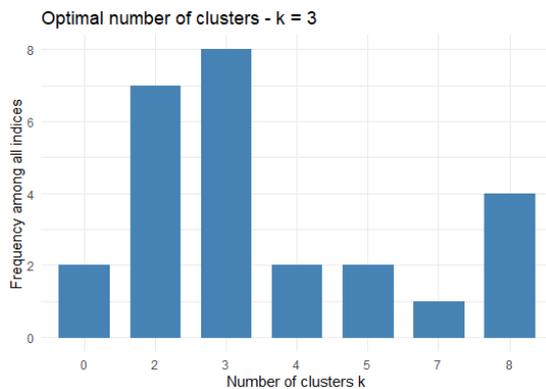
Hasil penentuan nilai C optimal menggunakan *elbow method* dan selisih SSE pada tiap-tiap *cluster*, grafik *silhouette method* dan diagram *factoextra* juga digunakan untuk mempertimbangkan hasil penentuan C optimal, hasil *elbow method* ditunjukkan oleh Gambar 3, grafik *silhouette method* ditunjukkan pada Gambar 4, diagram *factoextra* ditunjukkan oleh Gambar 5, sementara hasil selisih SSE pada tiap-tiap *cluster* ditunjukkan pada Tabel 3.



Gambar 3. Grafik *Elbow Method*



Gambar 4. Grafik *Silhouette Method*



Gambar 5. Diagram *Factoextra*

Tabel 3. Selisih SSE Pada Setiap Cluster

Cluster	Hasil SSE	Selisih
C2	31,586954	31,586954
C3	18,849531	12,737423
C4	13,8089912	5,0405398

Pada Gambar 3 digambarkan sebuah grafik dari *elbow method* yang menunjukkan nilai C terbaik adalah saat C berada di angka 3, hal ini karena nilai C terbaik adalah ketika terjadi penurunan

di beberapa nilai C dan selanjutnya hasil dari nilai C akan stabil atau turun secara perlahan-lahan, hal ini diperkuat oleh hasil diagram *factoextra* yang juga menunjukkan bahwa C optimal berada saat C = 3. Pada grafik *silhouette method* menunjukkan C optimal berada saat C = 2 kemudian C = 3. Tabel 3 menunjukkan C3 merupakan C terbaik karena memiliki jumlah *cluster* yang tidak terlalu sedikit dan memiliki selisih SSE lebih besar dari cluster di atasnya seperti C4. Berdasarkan hasil *elbow method* dan selisih SSE pada tiap *cluster* serta didukung oleh hasil diagram *silhouette method* dan diagram *factoextra*, maka C terbaik adalah 3. Hasil *clustering* dengan C = 3 dapat dilihat pada Gambar 6.

```
K-means clustering with 3 clusters of sizes 1, 6, 20
Cluster means:
  Isolasi    Sembuh    Meninggal
1  3.5734228  2.954782  2.9693129
2  0.9626675  1.079390  0.8415944
3 -0.4674714 -0.471556 -0.4009440

Clustering vector:
 [1] 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 2 3 3 2 3 2 3 2 1 3 3 3

within cluster sum of squares by cluster:
 [1] 0.000000 10.674424  8.175107
 (between_ss / total_ss = 75.8 %)
```

Gambar 6. Hasil *Clustering*

Ket:

Ss = sum of square, dalam literatur *data science* disebut *Sum Square of Error* (SSE)

Cluster of sizes = ukuran/jumlah titik data pada setiap *cluster*

Cluster means = nilai rata-rata (*centroid*) dari tiap *cluster*

Clustering vector = pembagian *cluster* dari tiap elemen data berdasarkan posisinya

Within_ss = total sum of square per *cluster*

Tot.within_ss = total penjumlahan dari setiap *ss* dari *within*

Total_ss = total sum of square untuk seluruh titik data

Between_ss = perbedaan nilai antara *total_ss* dan *tot.within_ss*

Between_ss/total_ss = rasio antara *between_ss* dibagi dengan *total_ss* atau

disebut juga *variance*. Semakin besar persentasenya, umumnya semakin baik.

Berdasarkan hasil *clustering* menggunakan algoritma *k-means* dengan $C = 3$ menghasilkan *variance* sebesar 75,8% dan SSE sebesar 18,849531. Hasil *clustering* dalam bentuk tabel dapat dilihat pada Tabel 4.

Tabel 4. Hasil *Clustering*

N	Nama	Isol	Semb	Menin	Clus
o	Kota/Kabup	asi	uh	ggal	ter
	aten				
1	Kab. Bogor	0,73	0,38	1,12	2
2	Kab. Sukabumi	-	-0,27	-0,57	3
		0,66			
3	Kab. Cianjur	-	-0,53	-0,97	3
		0,27			
4	Kab. Bandung	-	0,18	-0,46	3
		0,49			
5	Kab. Garut	-	-0,73	0,76	3
		0,28			
6	Kab. Tasikmalaya	-	-0,54	-0,26	3
		0,62			
7	Kab. Ciamis	-	-0,48	-0,32	3
		0,47			
8	Kab. Kuningan	-	-0,33	-0,80	3
		0,48			
9	Kab. Cirebon	-	-0,68	1,39	3
		0,32			
10	Kab. Majalengka	-	-0,54	-0,50	3
		0,53			
11	Kab. Sumedang	-	-0,54	-0,50	3
		0,68			
12	Kab. Indramayu	-	-0,69	-0,47	3
		0,67			
13	Kab. Subang	0,49	-0,50	-0,83	3
14	Kab. Purwakarta	-	-0,70	-0,27	3
		0,64			

15	Kab. Karawang	1,56	0,51	1,92	2
16	Kab. Bekasi	1,89	1,43	-0,26	2
17	Kab. Bandung Barat	-	-0,34	-0,54	3
		0,44			
18	Kab. Pangandaran	-	-0,63	-0,97	3
		0,81			
19	Kota Bogor	0,43	0,48	0,48	2
20	Kota Sukabumi	-	-0,41	-0,48	3
		0,55			
21	Kota Bandung	0,62	0,70	0,098	2
22	Kota Cirebon	-	-0,37	-0,55	3
		0,22			
23	Kota Bekasi	0,54	2,96	1,69	2
24	Kota Depok	3,57	2,95	2,97	1
25	Kota Cimahi	-	-0,34	-0,33	3
		0,64			
26	Kota Tasikmalaya	-	-0,31	-0,60	3
		0,36			
27	Kota Banjar	-	-0,68	-0,70	3
		0,67			

6. Evaluation

Pada tahap ini dilakukan evaluasi menggunakan indeks validitas *silhouette*, *dunn*, dan *connectivity*. Hasil evaluasi dapat dilihat pada Gambar 7.

```

Clustering Methods:
kmeans

Cluster sizes:
3

Validation Measures:
3

kmeans Connectivity 9.4738
      Dunn          0.4776
      Silhouette    0.5839

Optimal scores:

      Score Method Clusters
Connectivity 9.4738 kmeans 3
Dunn         0.4776 kmeans 3
Silhouette   0.5839 kmeans 3
    
```

Gambar 7. Hasil Evaluasi

Gambar 7 menunjukkan nilai evaluasi dari masing-masing indeks validitas, menghasilkan indeks *connectivity* = 9,4738, *dunn* = 0,4776, dan *silhouette* = 0,5839 (struktur data beralasan). Hasil evaluasi dalam bentuk tabel dapat dilihat pada Tabel 5.

Tabel 5. Hasil Evaluasi

Indeks Validitas	Hasil Evaluasi
<i>Silhouette</i>	0,5839
<i>Dunn</i>	0,4776
<i>Connectivity</i>	9,4738

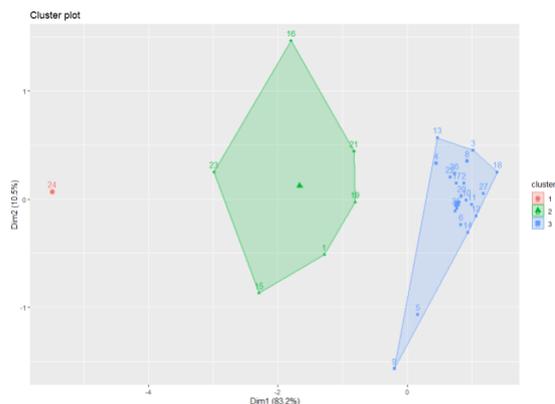
6. Knowledge Presentation

Pada tahap ini merupakan penyajian hasil yang diperoleh dari penelitian yang dilakukan. Hasil *clustering final* dapat dilihat pada Tabel 6 serta plot visualisasinya dapat dilihat pada Gambar 7.

Tabel 6. Hasil Clustering Final

No	Nama Kota/Kabupaten	Cluster
1	Kab. Bogor	2
2	Kab. Sukabumi	3
3	Kab. Cianjur	3
4	Kab. Bandung	3
5	Kab. Garut	3

6	Kab. Tasikmalaya	3
7	Kab. Ciamis	3
8	Kab. Kuningan	3
9	Kab. Cirebon	3
10	Kab. Majalengka	3
11	Kab. Sumedang	3
12	Kab. Indramayu	3
13	Kab. Subang	3
14	Kab. Purwakarta	3
15	Kab. Karawang	2
16	Kab. Bekasi	2
17	Kab. Bandung Barat	3
18	Kab. Pangandaran	3
19	Kota Bogor	2
20	Kota Sukabumi	3
21	Kota Bandung	2
22	Kota Cirebon	3
23	Kota Bekasi	2
24	Kota Depok	1
25	Kota Cimahi	3
26	Kota Tasikmalaya	3
27	Kota Banjar	3



Gambar 7. Plot Clustering

Pada Tabel 6 merupakan hasil penyajian dari clustering dengan $C = 3$ dan Gambar 7 merupakan plot dari hasil clustering. Untuk lebih memudahkan dalam membaca hasil *clustering* dapat dilihat pada Tabel 7.

Tabel 7. Hasil *Clustering* Terurut menurut *Cluster*

Nama Kota/Kabupaten	Cluster
Kota Depok	1
Kab. Bogor	2
Kab. Karawang	2
Kab. Bekasi	2
Kota Bekasi	2
Kota Bandung	2
Kota Bogor	2
Kab. Garut	3
Kab. Tasikmalaya	3
Kab. Ciamis	3
Kab. Kuningan	3
Kab. Cirebon	3
Kab. Majalengka	3
Kab. Sumedang	3
Kab. Indramayu	3
Kab. Subang	3
Kab. Purwakarta	3
Kab. Sukabumi	3
Kab. Bandung Barat	3
Kab. Pangandaran	3
Kota Sukabumi	3
Kota Cirebon	3
Kab. Cianjur	3
Kab. Bandung	3
Kota Cimahi	3
Kota Tasikmalaya	3
Kota Banjar	3

Untuk memudahkan dalam menganalisis kategori masing-masing *cluster*, deskripsi dari hasil *clustering* dapat dilihat pada Gambar 8.

```
# A tibble: 3 x 4
  cluster isolasi sembuh meninggal
  <int>   <dbl> <dbl>   <dbl>
1     1    6635  35453    322
2     2   2684. 17444.   149.
3     3    520.  2551.   48.5
```

Gambar 8. Deskripsi Hasil *Clustering*

Berdasarkan kesimpulan deskripsi hasil *clustering* pada Gambar 8, didapatkan bahwa *cluster* 1 merupakan kategori dengan tingkat isolasi, sembuh, dan meninggal tertinggi, *cluster* 2 merupakan kategori tingkat isolasi, sembuh, dan meninggal menengah dan *cluster* 3 adalah kategori dengan tingkat isolasi, sembuh, dan meninggal terendah. Hasil penggolongan kota/kabupaten ke masing-masing *cluster* dapat dilihat pada Tabel 8.

Tabel 8. Hasil *Clustering*

Cluster	Karakteristik			Jumlah	Kategori
	Isolasi	Sembuh	Meninggal		
C1	6.635	35.453	322	1	Tinggi
C2	1.883 - 4088	10.772 - 35.545	60 - 237	6	Menengah
C3	0 - 1.964	79 - 8.831	2 - 194	20	Rendah

Tabel 8 menyajikan hasil dari pembagian *cluster* sebaran covid-19 untuk kota/kabupaten di provinsi Jawa Barat dimana *cluster* dengan jumlah kota/kabupaten terbanyak adalah *cluster* 3 sebanyak 20 kota/kabupaten dan *cluster* dengan jumlah kota/kabupaten paling sedikit adalah *cluster* 1 sebanyak 1 kota/kabupaten, dengan masing-masing urutan *cluster* yaitu: *cluster* 1 sebanyak 1 kota/kabupaten, *cluster* 2 sebanyak 6 kota/kabupaten, dan *cluster* 3 sebanyak 20 kota/kabupaten.

Cluster 1 merupakan kategori dengan tingkat isolasi, sembuh, dan meninggal tertinggi, *cluster* 1 hanya

terdiri dari 1 kota/kabupaten yaitu kota Depok. Karena *cluster* ini hanya berjumlah 1, maka jumlah kasus, rata-rata kasus, angka kasus terkecil dan terbesar pada variabel isolasi adalah 6.635 kasus. Jumlah kasus, rata-rata kasus, angka kasus terkecil dan terbesar pada variabel sembuh adalah 35.453 kasus. Dan jumlah kasus, rata-rata kasus, angka kasus terkecil dan terbesar pada variabel meninggal adalah 322 kasus. Rincian karakteristik dari *cluster* 1 dapat dilihat pada Tabel 9.

Tabel 9. Rincian Karakteristik *Cluster* 1

C1	Isolasi	Sembuh	Meninggal
Min	6.635	35.453	322
Max	6.635	35.453	322
Rata-rata	6.635	35.453	322
Jumlah	6.635	35.453	322

Cluster 2 merupakan kategori dengan tingkat isolasi, sembuh, dan meninggal menengah, *cluster* 2 terdiri dari 6 kota/kabupaten yaitu Kab. Bogor, Kab. Karawang, Kab. Bekasi, Kota Bekasi, Kota Bandung, Kota Bogor. Pada variabel isolasi angka kasus terkecil = 1.883 kasus, angka kasus terbesar = 4.088 kasus, dengan jumlah kasus = 16.105 kasus dan rata-rata kasus = 2.684,167 kasus. Pada variabel sembuh angka kasus terkecil = 10.772 kasus, angka kasus terbesar = 35.545 kasus, dengan jumlah kasus = 104.665 kasus dan rata-rata kasus = 17.444,17 kasus. Sementara pada variabel meninggal angka kasus terkecil = 60 kasus, angka kasus terbesar = 237 kasus, dengan jumlah kasus = 896 kasus dan rata-rata kasus = 149,3333 kasus. Rincian karakteristik dari *cluster* 2 dapat dilihat pada Tabel 10.

Tabel 10. Rincian Karakteristik *Cluster* 2

C2	Isolasi	Sembuh	Meninggal
Min	1.883	10.772	60
Max	4.088	35.545	237
Rata-rata	2.684,16	17.444,1	149,3333
Jumlah	16.105	104.665	896

Cluster 3 merupakan kategori dengan tingkat isolasi, sembuh, dan meninggal terendah dari dua *cluster* lainnya, *cluster* 3 memiliki jumlah anggota terbanyak yaitu 20 kota/kabupaten yang terdiri atas Kab. Garut, Kab. Tasikmalaya, Kab. Ciamis, Kab. Kuningan, Kab. Cirebon, Kab. Majalengka, Kab. Sumedang, Kab. Indramayu, Kab. Subang, Kab. Purwakarta, Kab. Sukabumi, Kab. Bandung Barat, Kab. Pangandaran, Kota Sukabumi, Kota Cirebon, Kab. Cianjur, Kab. Bandung, Kota Cimahi, Kota Tasikmalaya dan Kota Banjar. Pada variabel isolasi *cluster* 3, angka kasus terkecil = 0 kasus, angka kasus terbesar = 1.964 kasus, jumlah kasus pada variabel isolasi di *cluster* 3 = 10.399 kasus dengan rata-rata kasus = 519,95 kasus. Pada variabel sembuh angka kasus terkecil = 79 kasus, angka kasus terbesar = 8.831 kasus, dengan jumlah kasus = 51.018 kasus dan rata-rata kasus = 2.550,9 kasus. Sementara pada variabel meninggal angka kasus terkecil = 2 kasus, angka kasus terbesar = 194 kasus, dengan jumlah kasus = 970 kasus dan rata-rata kasus = 48,5 kasus. Rincian karakteristik dari *cluster* 3 dapat dilihat pada Tabel 11.

Tabel 11. Rincian Karakteristik *Cluster* 3

C3	Isolasi	Sembuh	Meninggal
Min	0	79	2
Max	1.964	8.831	194

Rata-rata	519,95	2.550,9	48,5
Jumlah	10.399	51.018	970

SIMPULAN

Berdasarkan penelitian yang telah dilakukan, kesimpulan yang dapat diambil sebagai berikut.

1. Pada penelitian ini didapatkan jumlah *cluster* optimum yaitu 3, setelah melakukan evaluasi dengan beberapa indeks validitas, didapatkan nilai indeks *connectivity* = 9,4738, *dunn* = 0,4776, dan *silhouette* = 0,5839 (struktur data beralasan). Berdasarkan hal tersebut dapat dikatakan bahwa struktur data pada penelitian ini cukup baik dan pada hasil *clustering* dengan $C = 3$ menghasilkan *variance* yang baik yaitu dengan presentase sebesar 75,8%.
2. Jumlah kota/kabupaten paling sedikit adalah *cluster* 1 dengan kategori tingkat isolasi, sembuh, dan meninggal tertinggi yaitu berjumlah 1 kabupaten/kota, selanjutnya *cluster* 2 yaitu kategori kota/kabupaten dengan tingkat isolasi, sembuh, dan meninggal menengah berjumlah 6 kota/kabupaten, dan *cluster* 3 yaitu kategori tingkat isolasi, sembuh, dan meninggal terendah berjumlah 20 kota/kabupaten. *Cluster* 3 merupakan *cluster* dengan jumlah kota/kabupaten terbanyak diantara *cluster* lain.
3. Pada setiap *cluster* terdapat variabel dengan jumlah kasus bervariasi. Variabel dengan jumlah kasus terbanyak pada setiap *cluster* adalah variabel sembuh dengan jumlah kasus terbanyak berada di *cluster* 3 sebanyak 51.018 kasus, sementara variabel dengan jumlah kasus paling sedikit adalah variabel isolasi pada *cluster* 3 dengan jumlah kasus sebanyak 0 kasus.

DAFTAR PUSTAKA

- Aditya, A., Jovian, I., & Sari, B. N. (2020). Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama Di Indonesia Tahun 2018/2019. *Jurnal Media Informatika Budidarma*, 4(1): 51-58.
- Dwitri, N., Tampubolon, J.A., Prayoga, S., Zer, F. I. R. H., & Hartama, D. (2020). Penerapan Algoritma K-Means Clustering Dalam Menentukan Tingkat Penyebaran Pandemi Covid-19 Di Indonesia. *Jurnal Teknologi Informasi*, 4(1): 128-132.
- Hertina, H., Nurwahid, M., Haswir, H., Sayuti, H., Darwis, A., Rahman, M., ... & Hamzah, M. L. (2021). Data mining applied about polygamy using sentiment analysis on Twitters in Indonesian perception. *Bulletin of Electrical Engineering and Informatics*, 10(4), 2231-2236.
- Khotimah, T., & Darsin. (2020). Clustering Perkembangan Kasus Covid-19 Di Indonesia Menggunakan Self Organizing Map. *Jurnal Dialektika Informatika (Detika)*, 1(1): 23-26.
- Liu, J., Kantarci, B., & Adams, C. (2020, July). Machine learning-driven intrusion detection for Contiki-NG-based IoT networks exposed to NSL-KDD dataset. In *Proceedings of the 2nd ACM Workshop on Wireless Security and Machine Learning* (pp. 25-30).
- Mardi, Y. (2016). Data Mining: Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*, 2(2): 213-219.
- Riedel, S., Morse, S., Mietzner, T., Miler, S. J., Melnick., & Adelberg's Medical Microbiology. (2019). *Adelberg's Medical Microbiology* (28thed.). New York: McGraw-Hill Education/Medical.

- Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means clustering algorithm. *IEEE Access*, 8, 80716-80727.
- Sindi, S., Ningse, W. R. O., Sihombing, I. A., Ilmi, F., & Hartama, D. (2020). Analisis Algoritma K-Medoids Clustering Dalam Pengelompokan Penyebaran Covid-19 Di Indonesia. *Jurnal Teknologi Informasi*, 4(1): 166-173.
- Plotnikova, V., Dumas, M., & Milani, F. (2020). Adaptations of data mining methodologies: a systematic literature review. *PeerJ Computer Science*, 6, e267.
- Yogi Yunefri, Nizwardi Jalinus, Syahril, Muhammad Luthfi Hamzah. (2020). Grouping Systems In Cooperative Oriented Problem Learning Model Using K-Means Clustering. *International Journal of Advanced Science and Technology*, 29(05), 5963 - 5971. Retrieved from <http://serisc.org/journals/index.php/IJAST/article/view/15570>