

PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK PENGELOMPOKAN PENYEBARAN COVID-19 DI PROVINSI JAWA BARAT

Nita Mirantika¹, Annisa Tsamratul'Ain², Futry Diviana Agnia³

¹Sistem Informasi Universitas Kuningan,

^{2,3}Magister Sistem Informasi Universitas Komputer Indonesia

¹nita.mirantika@uniku.ac.id, ²ica03tsmratul@gmail.com, ³agniafutry@gmail.com

Abstrak

Covid-19 adalah kategori penyakit baru yang telah menyebar di seluruh penjuru dunia, termasuk di Indonesia. Covid-19 disebabkan oleh virus yang mengakibatkan gangguan pada sistem pernapasan, muncul gejala ringan seperti flu, sampai infeksi paru-paru, seperti pneumonia. Penyebaran covid-19 begitu cepat dan masif, di Indonesia total penyebaran covid-19 sudah mencapai 1.066.313 jiwa dan provinsi Jawa Barat menduduki posisi kedua tertinggi penyebaran covid-19. Oleh karena itu perlu dilakukan pengelompokan kota/kabupaten di provinsi Jawa Barat untuk membuat strategi penanganan penyebaran covid-19. Pengelompokan penyebaran covid-19 dilakukan dengan menggunakan algoritma K-Means clustering yang melakukan pengelompokan data ke beberapa cluster berdasarkan kemiripan datanya dan diimplementasikan pada tools aplikasi bahasa R serta dengan menggunakan metode CRISP-DM sehingga setiap fase dalam data mining lebih terstruktur dan terdefinisi dengan jelas. Berdasarkan hasil penelitian diperoleh 3 klaster, yaitu klaster 1 sebagai klaster dengan jumlah penyebaran covid-19 yang paling tinggi diperoleh 2 kabupaten/kota. Untuk klaster 2 sebagai klaster dengan jumlah penyebaran covid-19 menengah diperoleh 5 kabupaten/kota dan sisanya sebanyak 20 kabupaten/kota masuk dalam klaster 3 dengan jumlah penyebaran covid-19 yang lebih sedikit. Dari hasil penelitian ini diharapkan dapat mendukung pemerintah provinsi Jawa Barat untuk mengambil keputusan secara strategis pada setiap klasternya untuk mengurangi persebaran covid-19 di Jawa Barat.

Kata Kunci: *Algoritma K-Means, Clustering, CRISP-DM*

Abstract

Covid-19 is a new category of disease that has spread all over the world, including in Indonesia. Covid-19 is caused by a virus that causes disturbances in the respiratory system, from mild symptoms such as flu, to lung infections, such as pneumonia. The spread of covid-19 is so fast and massive, in Indonesia the total spread of covid-19 has reached 1,066,313 people and West Java province occupies the second highest position for the spread of covid-19. Therefore, it is necessary to group cities/districts in the West Java Province to create a strategy for handling the spread of COVID-19. The clustering of the spread of COVID-19 is carried out using the K-Means clustering algorithm which groups data into several clusters based on the similarity of the data and is implemented in R language application tools and by using the CRISP-DM method so that each phase in data mining is more structured and clearly defined. Based on the results of the study, there were 3 clusters, namely cluster 1 as the cluster with the highest number of Covid-19 spreads, obtained by 2 districts/cities. For cluster 2 as a cluster with a medium number of Covid-19 spreads, 5 districts/cities are obtained. And the remaining 20 districts/cities are included in cluster 3 with a smaller number of Covid-19 spreads. The results of this study are expected to support the West Java provincial government to make strategic decisions in each cluster to reduce the spread of COVID-19 in West Java.

Key Words: *K-Means Algorithm, Clustering, CRISP-DM*

1. PENDAHULUAN

Pada akhir Desember 2019 virus Covid-19 untuk pertama kali ditemukan di kota Wuhan, Tiongkok. Pandemi virus ini menyebar pada awal Maret 2020 di Indonesia. Covid-19 adalah penyakit yang diakibatkan oleh virus SARS-CoV-2 yang menyebabkan gangguan pada sistem pernapasan, muncul gejala ringan seperti flu, sampai infeksi paru-paru, seperti pneumonia. Virus ini dapat menular dengan cepat dan bahkan menyebar hampir ke seluruh negara, termasuk ke Indonesia. Virus ini bisa menular ke sesama manusia dengan droplet (percikan super kecil) yang dikeluarkan penderita saat berbicara, batuk, atau bersin. Untuk memutuskan rantai penyebaran Covid-19 pemerintah dan masyarakat menerapkan *physical distancing* dan pembatasan kegiatan masyarakat [1,3].

Saat ini total penyebaran Covid-19 di Indonesia telah mencapai 1.066.313 jiwa dan di provinsi Jawa Barat sendiri sudah mencapai 147.488 jiwa berada di urutan kedua tertinggi setelah provinsi DKI Jakarta. Oleh karena itu, dibutuhkan suatu strategi yang dilakukan pemerintah daerah untuk melakukan penanganan pada penyebaran covid-19 di provinsi Jawa Barat yang tersebar pada 27 Kota/ Kabupaten [1,3]. Dalam melakukan strategi tersebut perlu dilakukan klusterisasi atau pengelompokan data penyebaran covid-19 untuk mengetahui skala prioritas dan langkah strategis yang dilakukan sesuai dengan klaster penyebaran covid-19 di provinsi Jawa Barat. Pengelompokan data dilakukan menggunakan algoritma K-means, yang melakukan pengelompokan data ke sejumlah *cluster* berdasarkan kemiripan datanya dan metode CRISP-DM [2,4]. Algoritma K-Means adalah salah satu dari algoritma *data mining* yang populer dan umum digunakan. Selain itu, algoritma K-means memiliki kelebihan yaitu diantaranya mudah untuk diimplementasikan dan dijalankan serta cukup tangguh di berbagai jenis data. Sedangkan metode CRISP-DM merupakan salah satu metodologi dalam data mining yang bisa diimplementasikan ke dalam

strategi penyelesaian masalah pada bisnis atau pada unit penelitian. Pada metode CRISP-DM Setiap fase terstruktur dan terdefinisi dengan jelas sehingga lebih lengkap, terdokumentasi dengan baik dan mudah diaplikasikan [4,5,10].

Dengan mendapatkan data penyebaran covid-19 yang sudah dikelompokkan menggunakan Algoritma K-means dan metode CRISP-DM ini diharapkan dapat mendukung dalam pengambilan keputusan strategis yang akurat untuk memangkas penyebaran virus covid-19 di provinsi Jawa Barat.

2. METODE PENELITIAN

2.1. Data Mining

Mengekstraksi pengetahuan dalam bentuk *rule*, *regularities*, *pola*, *constraint* dari data yang tersimpan pada sejumlah basis data. Data mining juga dikenal dengan *Knowledge Discovery in Databases (KDD)* yang mengekstraksi pengetahuan menggunakan data historikal [8,9].

2.2. Data Mining Clustering

Proses mengorganisasikan objek-objek ke dalam kelompok-kelompok (*cluster*) yang anggota kelompoknya memiliki kemiripan di beberapa karakteristiknya. Tujuan dari *clustering* adalah mengelompokkan beberapa data atau objek ke dalam klaster sehingga klaster tersebut berisi data yang mirip. Metode *clustering* ada 2: 1) *Hierarchical Clustering* 2) *Non-Hierarchical Clustering* [7].

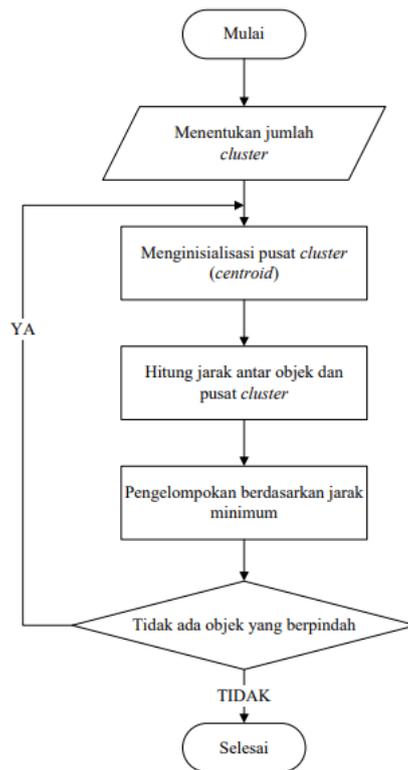
2.3. CRISP-DM

Metode penelitian yang digunakan adalah CRISP-DM (*Cross Industry Standard Process for Data Mining*) merupakan salah satu model dalam proses penambangan data (*data mining*) yang memiliki enam tahap yang berurutan yaitu [2,4,5]:

1. Pemahaman Bisnis (*Business Understanding*), tahap pertama subjek pada penelitian ini adalah provinsi Jawa Barat yang menjadi provinsi kedua dengan jumlah penderita covid-19 terbanyak setelah DKI Jakarta. Oleh karena itu perlu adanya strategi dengan *clustering* atau pengelompokan data

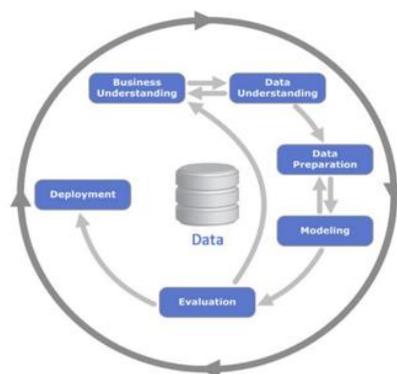
penyebaran covid-19 untuk mengetahui skala prioritas dan langkah strategis yang dilakukan sesuai cluster penyebaran covid-19 di provinsi Jawa Barat.

2. Pemahaman data (*Data Understanding*), peneliti melakukan pemahaman terhadap kebutuhan data terkait dengan penyebaran covid-19 di provinsi Jawa Barat. Data diambil dari Pusat Informasi dan Koordinasi Covid 19 Jawa Barat atau pikobar yaitu website resmi pemerintah provinsi Jawa Barat sebagai pusat informasi dan koordinasi serta tempat bertanya dan pengaduan masyarakat tentang segala hal yang berkaitan dengan wabah virus covid-19. Penelitian ini menggunakan data dari pikobar dengan batas waktu sampai dengan 24 Januari 2021. Data tersebut terdiri dari 27 kota/kabupaten di provinsi Jawa Barat yang terdiri dari data jumlah isolasi/ dalam perawatan, jumlah selesai isolasi/ sembuh dan jumlah meninggal dunia [6].
3. Persiapan data (*Data Preparation*) yaitu mempersiapkan data dilakukan beberapa proses diantaranya pemilihan variabel untuk dianalisis, pembersihan data, persiapan data awal hingga siap untuk transformasi data. Data selection yang diambil di pikobar. Selanjutnya, melakukan data preprocessing. pada proses ini bertujuan untuk melihat kualitas data yang dipilih pada proses *data selection*. Pada tahap ini dilakukan pembersihan data (*Data cleansing*) yaitu proses pembersihan data dari *missing value* atau *noise*. Tahap berikutnya adalah transformasi data dengan cara normalisasi agar tidak ada perbedaan yang terlalu jauh pada skala data.
4. Pemodelan (*Modelling*), pada tahap ini memilih dan menerapkan beberapa teknik pemodelan dan sejumlah parameter disesuaikan sehingga dapat nilai yang optimal. Adapun pemodelan yang dilakukan pada penelitian ini adalah algoritma clustering menggunakan algoritma K-Means yang ditunjukkan oleh Gambar 1 berikut ini:



Gambar 1. Tahapan Algoritma K-Means

5. Evaluasi (*Evaluation*), pada tahap ini fokus terhadap model yang didapatkan telah sesuai dengan standar K-Means Clustering dan tiada yang terlewat pada tahapan awal sampai tahap pemodelan selesai.
6. Penyebaran (*Deployment*), pada tahap terakhir ini CRISP-DM adalah penyebaran pengetahuan atau informasi yang telah diperoleh yang di implementasikan pada laporan dan melakukan analisis tiap cluster yang diperoleh supaya mudah dimengerti oleh pihak yang membutuhkan. Tahapan tersebut terdapat pada Gambar 2 sebagai berikut:



Gambar 2. Tahapan CRISP-DM

2.4. Algoritma K-Means Clustering

K-means clustering adalah metode kluster non hirarki yang mempartisi objek ke dalam satu atau lebih kluster atau kelompok melihat pada karakteristiknya, objek dengan karakteristik yang mirip dikelompokkan pada satu kluster yang sama, objek dengan karakteristik yang berbeda dikelompokkan ke dalam kluster yang berbeda. Tahapan algoritma K-Means:

1. Tentukan nilai k (jumlah kluster) dan centroid dari masing-masing kluster secara acak.
2. Hitung jarak tiap data terhadap centroid menggunakan rumus *Ecludian Distance*

$$dist(x,y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

3. Bentuk cluster dengan memanfaatkan hasil *ecludian distance* terdekat
4. Hitung nilai ratio sebagai bahan perbandingan untuk stop iterasi dengan rumus:

$$Ratio = bcv/wcv \quad (2)$$

- Nilai *bcv (between cluster variation)* merupakan jarak antar centroid terpilih dengan rumus *ecludian distance*.
- Nilai *wcv (within cluster variation)* merupakan jumlah kuadra jarak terdekat setiap data.

$$wcv = \sum_{i=1}^j (jarak\ terdekat\ setiap\ data\ i)^2 \quad (3)$$

5. Cari nilai centroid baru dengan memanfaatkan rata-rata nilai dari setiap anggota kluster
6. Ulangi langkah 1-5 sampai anggota cluster tiada yang pindah cluster atau nilai ratio baru \leq nilai ratio lama.

3. PEMBAHASAN

3.1. Perhitungan dengan Algoritma K-Means

Data yang digunakan untuk pemodelan dengan algoritma K-Means diambil dari website pikobar per tanggal 24 Januari 2021. Berikut adalah sampel data penyebaran COVID-19 di Jawa Barat:

Tabel 1. Data Penyebaran Covid-19 di Jawa Barat

Nama Kota/Kabupaten	Total Terkonfirmasi	Total Isolasi/ Dalam Perawatan	Total Selesai Isolasi/ Sembuh	Total Meninggal
Kab. Bogor	4.447	124	4.307	16
Kab. Sukabumi	2.484	223	2.229	34
Kab. Cianjur	443	234	207	2
Kab. Bandung	6.309	1.889	4.393	27
Kab. Garut	3.991	617	3.313	61
Kab. tasikmalaya	1.143	817	324	2
Kab. Ciamis	1.415	413	973	29
Kab. Kuningan	1.936	271	1.649	16
Kab. Cirebon	2.892	448	2.369	75
Kab. Majalengka	1.514	85	1.399	30
Kab. Sumedang	1.396	327	1.036	33
Kab. Indramayu	1.305	39	1.247	19
Kab. Subang	1.288	486	788	14
Kab. Purwakarta	1.882	170	1.659	53
Kab. Karawang	8.462	1.464	6.844	154
Kab. Bekasi	11.877	2.388	9.430	59
Kab. BandungBarat	2.235	549	1.664	22
Kab. angandaran	137	2	134	1
Kota Bogor	3.719	92	3.585	42
Kota Sukabumi	1.782	731	1.041	10
Kota Bandung	8.453	1.735	6.629	89
Kota Cirebon	585	2	565	18
Kota Bekasi	20.024	3.986	15.866	172
Kota Depok	22.527	3.315	18.936	276
Kota Cimahi	3.283	369	2.869	45
Kota Tasikmalaya	2.929	870	2.039	20
Kota Banjar	488	140	324	24

Pada tahap ini didapatkan hasil :

Jumlah *Cluster* : 3

Jumlah Data : 27

Jumlah Atribut : (Total Isolasi/ Dalam Perawatan, Total Selesai Isolasi/ Sembuh, Total Meninggal)

Iterasi-1

1. Menentukan centeroid awal secara acak
C1 (Kab. Bogor) = (124, 4307,16) , C2 (Kab. Sukabumi) = (223,229,34) , C3(Kab. Cianjur) = (234,207,2)

2. Menghitung centroid terdekat

$$C1 (Kab. Bogor) = \sqrt{(124 - 124)^2 + (4307 - 4307)^2 + (16 - 16)^2} = 0,000$$

$$C2 (Kab. Bogor) = \sqrt{(124 - 223)^2 + (4307 - 229)^2 + (16 - 34)^2} = 2080,435$$

$$C3 (Kab. Bogor) = \sqrt{(124 - 234)^2 + (4307 - 207)^2 + (16 - 2)^2} = 4101,499$$

Langkah selanjutnya yaitu menggunakan persamaan *Euclidean Distance* untuk menghitung jarak tiap data dengan *Centroid*. Tahapan ini merupakan penentuan suatu data masuk ke dalam *Cluster* mana dilihat dari jarak terdekat antara data dengan *Cluster*.

3. Membuat kelompok berdasarkan cluster. Berikut merupakan hasil pengelompokkan *cluster* pada iterasi ke 1:

Tabel 2. Hasil pengelompokan kluster pada iterasi 1

Nama Kota/Kabupaten	C1	C2	C3	Terdekat
Kab. Bogor	0	2080,43481	4.101	C1
Kab. Sukabumi	2080,43481	0	2.022	C2
Kab. Cianjur	4101,499238	2022,283116	0	C3
Kab. Bandung	1767,128179	2731,025632	4.501	C1
Kab. Garut	1110,454862	1153,698834	3.130	C1
Kab. Tasikmalaya	4042,862105	1995,716663	595	C3
Kab. Ciamis	3346,527454	1270,299571	787	C3
Kab. Kuningan	2662,061795	582,2611098	1.443	C2
Kab. Cirebon	1965,782541	268,1529414	2.174	C2
Kab. Majalengka	2908,295205	841,4035892	1.202	C2
Kab. Sumedang	3277,337181	1197,524948	835	C3
Kab. Indramayu	3061,181798	999,2021817	1.058	C2
Kab. Subang	3537,571059	1464,940272	633	C3
Kab. Purwakarta	2648,657962	572,7739519	1.454	C2
Kab. Karawang	2872,457659	4780,450397	6.752	C1
Kab. Bekasi	5601,131493	7519,458159	9.471	C1
Kab. Bandung Barat	2676,959096	652,4147454	1.491	C2
Kab. Pangandaran	4174,809936	2106,882768	243,2159534	C3
Kota Bogor	723,176327	1362,336596	3.381	C1
Kota Sukabumi	3321,933323	1292,278608	971	C3
Kota Bandung	2827,07163	4652,866751	6.596	C1
Kota Cirebon	3743,988782	1678,68788	426,9004568	C3
Kota Bekasi	12188,10326	14147,33127	16.103	C1
Kota Depok	14975,23696	16992,43588	18.983	C1
Kota Cimahi	1459,009938	656,5340814	2.666	C2
Kota Tasikmalaya	2387,541832	674,4664558	1.939	C2
Kota Banjar	3983,040171	1906,833501	151,6871781	C3

4. Hitung nilai ratio sebagai bahan perbandingan untuk stop iterasi dengan rumus:

$$C1 \text{ (Kab. Bogor)} = (124, 4307,16) \quad D$$

$$(C1, C2) \quad 2080,435$$

$$C2 \text{ (Kab. Sukabumi)} = (223,229,34) \quad D$$

$$(C1, C3) \quad 2022,283$$

$$C3 \text{ (Kab. Cianjur)} = (234,207,2) \quad D$$

$$(C2, C3) \quad 8204,217$$

$$bcv = 2080,435 + 4101,499 + 2022,283$$

$$bcv = 8204,217$$

$$wcv = 432337596$$

maka, rasionya

$$Ratio = 8203,217/432337596$$

$$Ratio = 0,00001897$$

5. Memilih Centroid baru pada iterasi selanjutnya dengan menghitung nilai rata-rata tiap cluster. Berikut merupakan nilai centroid baru pada iterasi ke 2

Tabel 3. Centroid baru iterasi ke 2

Centeroid			
C1	1734,444	8144,778	99,556
C2	336	1902,7	34,9
C3	350,222	599,111	14,778

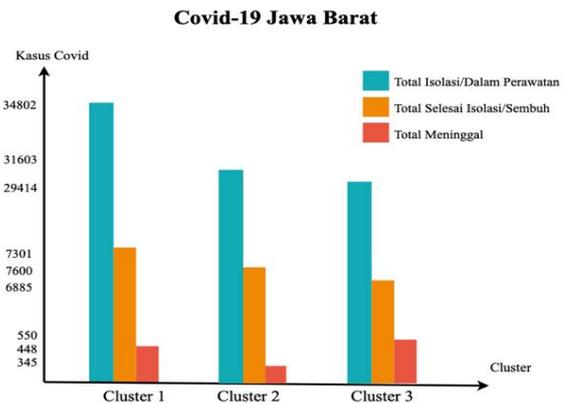
Perhitungan dihentikan apabila tahapan iterasi sudah mendapatkan hasil sama, tidak ada objek yang pindah lagi. Pada penelitian ini, proses iterasi yang dilakukan sampai dengan iterasi ke 9. Hasil akhir clustering iterasi ke 9 dapat dilihat pada table berikut ini:

Tabel 4. Hasil akhir clustering iterasi 9

Kluster	Jumlah	Nama Kota/Kabupaten
C1	2	Kota Bekasi, Kota Depok
C2	5	Kab. Bogor, Kab. Bandung, Kab. Karawang, Kab. Bekasi, Kota Bandung
C3	20	Kab. Sukabumi, Kab. Cianjur, Kab. Garut, Kab. Tasikmalaya, Kab. Ciamis, Kab. Kuningan, Kab. Cirebon, Kab. Majalengka, Kab. Sumedang, Kab. Indramayu, Kab. Subang, Kab. Purwakarta, Kab. Bandung Barat, Kab. Pangandaran, Kota Bogor, Kota Sukabumi, Kota Cirebon, Kota Cimahi, Kota Tasikmalaya, Kota Banjar

Hasil Untuk kluster 1 sebagai kluster dengan jumlah penyebaran covid-19 yang besar terdapat di 2 kabupaten/kota. Untuk kluster 2 sebagai kluster dengan jumlah penyebaran covid-19 menengah diperoleh 5 kabupaten/kota Dan sisanya sebanyak 20 kabupaten/kota masuk dalam kluster 3 dengan jumlah penyebaran covid-19 yang lebih sedikit.

Hasil olah data secara manual dengan 27 kota/kabupaten dapat dilihat pada gambar berikut ini:



Gambar 3. Kluster Penyebaran Covid-19

Dari grafik tersebut diperoleh hasil cluster 1 dengan penyebaran dalam jumlah besar yaitu sebanyak 34802 kasus total isolasi/dalam perawatan, 7301 total selesai isolasi/sembuh dan 448 kasus meninggal. Cluster 2 dengan penyebaran dalam jumlah menengah yaitu dengan 31603 kasus total isolasi/ dalam perawatan, 7600 total selesai isolasi/sembuh dan 345 kasus meninggal. Cluster 3 dengan penyebaran jumlah kecil yaitu dengan 29414 kasus total isolasi/dalam perawatan, 6885 total selesai isolasi/sembuh dan 550 kasus meninggal.

3.2. Implementasi pada aplikasi Tools R

Pengimplementasian pada aplikasi Tools R, uji coba data mining k-means dapat ditunjukkan sebagai berikut;

- Memasukan data set Covid-19-Data-Jabar.csv

```
datacovid<-
read.csv(file.choose(),header = TRUE)
```
- Melihat data covid

```
View(datacovid)
```
- Memasukan data frame data covid ke dalam variable datacovid

```
datacovid<-
as.data.frame(datacovid)
```
- Memasukkan ke data frame data covid ke dalam data covid ready

```
(datacovidready<data.frame(datacovid$TotalIsolasiDalamPerawatan,datacovid$TotalSelesaiIsolasiSembuh,datacovid$TotalMeninggal))
```
- Merangkum data covid

```
summary(datacovidready)
```

datacovid.TotalIsolasi DalamPerawatan	datacovid.TotalSelesaiIsolasiSembuh	datacovid.TotalMeninggal
Min. : 2.0	Min. : 134	Min. : 1.00
1st Qu.: 155.0	1st Qu.: 1004	1st Qu.: 17.00
Median : 413.0	Median: 1664	Median : 29.00
Mean : 806.9	Mean : 3549	Mean : 49.74
3rd Qu.: 843.5	3rd Qu.: 3946	3rd Qu.: 56.00
Max. :3986.0	Max. :18936	Max. :276.00
- Percobaan algoritma Hartigan-Wong di K-Means

```
(hasilkmeans<-
kmeans(datacovidready,3,algorithm =
"Hartigan-Wong"))
```
- melihat hasil pemetaan data kota/kabupaten per cluster

```
table(datacovid$NamaKotaKabupaten, hasilkmeans$cluster)
```
- Percobaan algoritma Llyod di K-Means

```
(hasilkmeans<-
kmeans(datacovidready,3,algorithm =
"Lloyd"))
```
- melihat hasil pemetaan data kota/kabupaten per cluster

```
table(datacovid$NamaKotaKabupaten, hasilkmeans$cluster)
```
- Percobaan algoritma Forgry di K-Means

```
(hasilkmeans<-
kmeans(datacovidready,3,algorithm =
"Forgry"))
```
- melihat hasil pemetaan data kota/kabupaten per cluster

```
table(datacovid$NamaKotaKabupaten, hasilkmeans$cluster)
```
- Percobaan algoritma MacQueen di K-Means

```
(hasilkmeans<-
kmeans(datacovidready,3,algorithm =
"MacQueen"))
```

- melihat hasil pemetaan data kota/kabupaten per cluster

```
table(datacovid$NamaKotaKabupaten, hasilkmeans$cluster)
```

Tabel 4. Hasil Pemetaan per Cluster pada algoritma di K-Means

	Hartigan-Wong	Llyod	Forgry	MacQueen
	1 2 3	1 2 3	1 2 3	1 2 3
Kab. Bandung	1 0 0	0 1 0	0 1 0	1 0 0
Kab. Bandung Barat	0 1 0	1 0 0	0 0 1	0 1 0
Kab. Bekasi	1 0 0	0 1 0	0 1 0	1 0 0
Kab. Bogor	1 0 0	0 1 0	0 1 0	1 0 0
Kab. Ciamis	0 1 0	1 0 0	0 0 1	0 1 0
Kab. Cianjur	0 1 0	1 0 0	0 0 1	0 1 0
Kab. Cirebon	0 1 0	1 0 0	0 0 1	0 1 0
Kab. Garut	0 1 0	1 0 0	0 0 1	0 1 0
Kab. Indramayu	0 1 0	1 0 0	0 0 1	0 1 0
Kab. Karawang	1 0 0	0 1 0	0 1 0	1 0 0
Kab. Kuningan	0 1 0	1 0 0	0 0 1	0 1 0
Kab. Majalengka	0 1 0	1 0 0	0 0 1	0 1 0
Kab. Pangandaran	0 1 0	1 0 0	0 0 1	0 1 0
Kab. Purwakarta	0 1 0	1 0 0	0 0 1	0 1 0
Kab. Subang	0 1 0	1 0 0	0 0 1	0 1 0
Kab. Sukabumi	0 1 0	1 0 0	0 0 1	0 1 0
Kab. Sumedang	0 1 0	1 0 0	0 0 1	0 1 0
Kab. Tasikmalaya	0 1 0	1 0 0	0 0 1	0 1 0
Kota Bandung	1 0 0	0 1 0	0 1 0	1 0 0
Kota Banjar	0 1 0	1 0 0	0 0 1	0 1 0
Kota Bekasi	0 0 1	0 0 1	1 0 0	0 0 1
Kota Bogor	0 1 0	1 0 0	0 0 1	0 1 0
Kota Cimahi	0 1 0	1 0 0	0 0 1	0 1 0
Kota Cirebon	0 1 0	1 0 0	0 0 1	0 1 0
Kota Depok	0 0 1	0 0 1	1 0 0	0 0 1
Kota Sukabumi	0 1 0	1 0 0	0 0 1	0 1 0
Kota Tasikmalaya	0 1 0	1 0 0	0 0 1	0 1 0

Berdasarkan hasil pengelompokan memakai program aplikasi tools R, didapatkan hasil pengelompokan yang sama dengan perhitungan manual. Pengelompokan menggunakan package pada tools R terdapat empat algoritma yang digunakan untuk memperhalus data yaitu algoritma Hartigan-Wong, algoritma Llyod, algoritma Forgry, dan algoritma MacQueen. Perbedaan hanya terdapat pada posisi nama klusternya saja, sedangkan anggota masing-masing klaster tidaklah berbeda.

4. PENUTUP

Berdasarkan hasil penelitian yang dilakukan, dapat kami simpulkan bahwa perhitungan secara manual algoritma k-means clustering dengan menggunakan metode CRISP-DM dan perhitungan menggunakan tools aplikasi R diperoleh klaster atau pengelompokan yang sama. Untuk klaster 1 sebagai klaster dengan jumlah penyebaran covid-19 yang paling tinggi diperoleh 2 kota/ kabupaten yaitu kota Depok dan kota Bekasi. Untuk klaster 2 sebagai klaster dengan jumlah penyebaran covid-19 menengah diperoleh 5 kota/ kabupaten yaitu Kota Bandung, Kabupaten Bandung, Kabupaten Bekasi, Kabupaten Bogor dan Kabupaten Karawang. Sisanya sebanyak 20 kota/kabupaten masuk dalam

klaster 3 dengan jumlah penyebaran covid-19 yang lebih sedikit.

Dengan pengelompokan penyebaran covid-19 bisa menjadi bahan pertimbangan pemerintah provinsi Jawa Barat dalam mengambil keputusan untuk strategi penanganan covid-19 di Jawa Barat pada setiap klasternya. Klaster 1 dan klaster 2 perlu mendapatkan perhatian serius dan prioritas dalam penanganan covid-19 karena mempunyai jumlah penyebaran yang tinggi. Dengan membuat strategi penanganan yang tepat pada setiap klasternya diharapkan dapat mengakhiri rantai penyebaran covid-19 di provinsi Jawa Barat.

DAFTAR PUSTAKA

- [1] 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.
- [2] Astuti, D. (2019). Penentuan Strategi Promosi Usaha Mikro Kecil dan Menengah (UMKM) Menggunakan Metode CRISP-DM dengan Algoritma K-Means Clustering. *Journal of Informatics, Information System, Software Engineering and Applications* 1 (2) 60-72.
- [3] Dwitri, N., Tampubolon, J., Prayoga, S. Zer F., dan Hartama, D. (2020). Penerapan algoritma K-Means dalam menentukan tingkat penyebaran pandemi COVID-19 di Indonesia. *JurTI (Jurnal Teknologi Informasi)*, 4 (1) 128-132.
- [4] Fadillah, A. (2015). Penerapan Metode CRISP-DM untuk Prediksi Kelulusan Studi Mahasiswa Menempuh Mata Kuliah (Studi Kasus Universitas XYZ). *Jurnal Teknik Informatika Dan Sistem Informasi*, 1 (3).
- [5] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., and Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide *SPSS inc* 9 13
- [6] pikobar.jabarprov.go.id. 2021 Sebaran Kasus Covid-19 di Jawa Barat *Diakses pada 24 Januari 2021*, dari <https://pikobar.jabarprov.go.id/distributioon-case>
- [7] Bachtiar, Adam M. (2017). Manajemen Data Data Mining Association Rule. *Diakses pada 25 Januari 2021*, dari <https://www.slideshare.net/adfbipotter/data-management-data-mining-association-rule>
- [8] Bachtiar, Adam M. (2021). Data Mining Clustering. *Diakses pada 25 Januari 2021* dari <https://www2.slideshare.net/adfbipotter/data-mining-clustering>
- [9] Firdaus, E.A., Maulani, S., Dharmawan, A.B. (2021). Pengukuran Minat Baca Mahasiswa dengan Metode Clustering di Perpustakaan Akademi Keperawatan RS. Dustira Cimahi Menggunakan Data Mining. *Jurnal Nuansa Informatika*, Vol 15 (1).
- [10] Suhandi, Y., Kurniati, I., & Norma, S. (2020). Penerapan Metode Crisp-DM Dengan Algoritma K-Means Clustering Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik. *Jurnal Teknologi Informatika dan Komputer*, 6(2), 12-20.