Clustering Penyebaran Covid-19 di Kota Bandung dengan Algoritma K-Means

1st Rifki Zaenudin
Fakultas Rekayasa Industri
Universitas Telkom
Bandung, Indonesia
rifkizaenudin@student.telkomuniversity.ac.id

2nd Moh. Deni Akbar
Fakultas Rekayasa Industri
Universitas Telkom
Bandung, Indonesia
denimath@telkomuniversity.ac.id

3rd Vandha Pradwiyasma Widartha
Fakultas Rekayasa Industri
Universitas Telkom
Bandung, Indonesia
vandhapw@telkomuniversity.ac.id

Abstrak—Covid-19 merupakan penyakit yang disebabkan oleh *coronavirus*. Covid-19 pertama kali masuk ke Indonesia pada awal Maret 2020. Pada saat ini penyebaran yang terjadi pada Provinsi Jawa Barat telah mencapai 706.800 jiwa. Penyebaran tertinggi pada Provinsi Jawa Barat berada pada Kota Bandung dengan penyebaran sebanyak 43.269 jiwa. Salah satu strategi untuk mengurangi dampak dari wabah ini, peneliti memanfaatkan *machine learning* yang mampu melakukan *clustering* untuk mengetahui skala prioritas. Metode *clustering* dapat menggunakan algoritma k-means. Salah satu kelebihan algoritma k-means yaitu memiliki hasil evaluasi *cluster* yang baik dan mudah untuk diimplementasikan. Hasil dari penelitian ini memiliki 9 *cluster*, yaitu C0 merupakan *cluster* dalam kategori sedang, C1 merupakan *cluster* dalam kategori sedang, C2 merupakan *cluster* dalam kategori rendah, C3 merupakan *cluster* dalam kategori rendah, C4 merupakan *cluster* dalam kategori tinggi, C5 merupakan *cluster* dalam kategori tinggi, C6 merupakan *cluster* dalam kategori tinggi, C7 merupakan *cluster* dalam kategori rendah, dan C8 merupakan *cluster* dalam kategori tinggi.

Kata Kunci—Covid-19, clustering, algoritma K-Means.

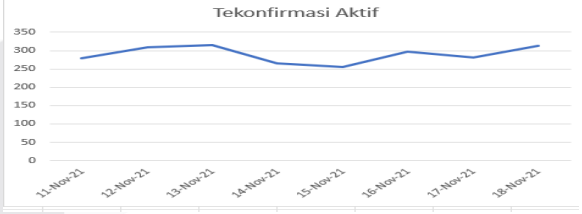
Abstract—Covid-19 is a disease caused by a coronavirus. Covid-19 first entered Indonesia in early March 2020. At this time the spread in the West Java Province has reached 706,800 people. The highest distribution in West Java Province is in the city of Bandung with a distribution of 43,269 people. One strategy to reduce the impact of this outbreak, researchers utilize machine learning that is capable of clustering to determine the priority scale. The clustering method can use the k-means algorithm. One of the advantages of the k-means algorithm is that it has good cluster evaluation results and is easy to implement. The results of this study have 9 clusters, namely C0 is a cluster in the medium category, C1 is a cluster in the medium category, C2 is a cluster in the low category, C3 is a cluster in the low category, C4 is a cluster in the high category, C5 is a cluster in the high category. , C6 is a cluster in the high category, C7 is a cluster in the low category, and C8 is a cluster in the high category.

Keywords—Covid-19, clustering, K-Means algorithm

1. PENDAHULUAN

Covid-19 merupakan penyakit yang disebabkan oleh *coronavirus*, Covid-19 diambil dari kata 'CO' *corona*, 'VI" *virus*, dan 'D" *disease* (penyakit). Covid-19 pertama kali ditemukan pada akhir Desember 2019 di kota Wuhan, Tiongkok. Covid-19 pertama kali masuk ke Indonesia pada awal Maret 2020.

Pada tanggal 11 November 2021 Kota Bandung menjadi Kota dengan posisi paling tinggi pada tingkat penyebaran Covid-19 nya, saat ini total penyebaran Covid-19 di Kota Bandung mencapai 43.269 jiwa menurut *website* Pusat Informasi Covid-19 Kota Bandung (covid19.bandung.go.id, 2021).



GAMBAR 1. 1
PENYEBARAN COVID-19 DI KOTA BANDUNG

Salah satu strategi dalam mengurangi dampak yang besar dari suatu wabah peneliti dapat memanfaatkan dengan kecerdasan buatan. Cabang dari ilmu kecerdasan buatan adalah *machine learning* yang mampu untuk melakukan *clustering* pada data penyebaran Covid-19 untuk mengetahui skala prioritas. *clustering* ini dapat dilakukan dengan menggunakan algoritma k-means berdasarkan kemiripan datanya.

Untuk menyelesaikan permasalahan mengenai penyebaran Covid-19 di atas, maka peneliti

melakukan *clustering* dengan menggunakan algoritma k-means. Penelitian ini diharapkan dapat mendukung dalam pengambilan keputusan secara akurat untuk meminimalisir bahkan memutus rantai penyebaran Covid-19 yang ada di Kota Bandung.

II. KAJIAN TEORI

A. *Data Mining*

Data mining adalah suatu teknik penemuan informasi dalam *database* (KDD). *Knowladge Discovery in Database* (KDD) adalah proses penggalian data dengan beberapa proses seperti penggalian data (*data mining*), pemilihan data, pembersihan data, pengintegrasian data, pentransformasian data, pengevaluasian *pattern* (pola), dan penyajian hasil pengetahuan informasi (Jiawei, Kamber, 2012). Hasil dari pengolahan data dengan metode ini dapat digunakan untuk mengambil suatu keputusan di masa depan. *Data mining* merupakan pengolahan data dengan skala besar, sehingga memiliki peranan penting pada bidang industri, keuangan, teknologi, cuaca dan ilmu. *Data mining* dapat dilakukan di berbagai macam database, jenis pola yang akan digunakan pada *data mining* yaitu *clustering*, asosiasi, klasifikasi, prediksi, dan lain-lain.

B. *Clustering*

Clustering merupakan suatu metode untuk mecari dan mengelompokkan data yang memiliki kemiripan karakteristik (*similarity*) antara satu sama lain. *Clustering* merupakan metode dari *data mining* yang memiliki sifat tanpa arahan (*unsupervised*). Dalam data mining ada dua jenis metode *clustering* yaitu metode *hierarchical clustering* dan *non-hierarchical clustering* (Santosa, dkk, 2007).

Hierarchical clustering merupakan suatu metode *cluster* (pengelompokkan) data yang mengelompokkan dua atau lebih objek berdasarkan karakteristik nya. kemudian proses akan diteruskan ke objek lainnya yang memilik karakteristik yang sama sehingga *cluster* akan membentuk semacam pohon dimana terdapat *hierarki* yang jelas antar objek nya dari paling mirip hingga tidak mirip (Santosa, dkk, 2007).

Non-hierarchical clustering dimulai dengan menentukan terlebih dahulu jumlah *Cluster* yang diinginkan, kemudian proses *cluster* dilakukan tanpa mengikuti proses dari hierarki. metode *non-hierarchical* hierarki ini biasa disebut dengan algoritma k-means *clustering* (Agusta, 2007).

C. Algoritma K-Means

Algoritma k-means adalah algoritma untuk pelatihan *unsupervised*. Algoritma k-means banyak digunakan pada metode *clustering* dan bersifat *non-hierarchical clustering*. Prinsip utama yang digunakan algoritma k-means adalah menyusun sebuah partisi atau pusat (*centroid*), rata-rata (*mean*) dari sekumpulan data. Algoritma K-means dimulai dengan pembentukan partisi klaster di awal kemudian secara iteratif partisi klaster ini diperbaiki hingga tidak terjadi perubahan yang signifikan pada partisi klaster (Mustakim, 2012). Dalam penyelesaiannya, algoritma k-means akan menghasilkan titik *centroid* yang dijadikan tujuan pada algoritma k-means. Setelah iterasi berhenti, setiap objek dalam *dataset* menjadi anggota dari suatu *cluster*, nilai *cluster* dapat ditentukan dengan mencari seluruh objek untuk menemukan *cluster* dengan menghitung jarak terdekat ke objek. Algoritma k-means akan mengelompokkan data dalam suatu *dataset* ke suatu *cluster* berdasarkan jarak terdekat . *Centroid* awal yang dipilih secara acak atau *random* menjadi titik pusat awal, kemudian dihitung jarak dengan semua data menggunakan rumus *euclidean distance*. Proses ini berlanjut hingga tidak terjadi perubahan pada setiap kelompok atau *cluster*.

D. Evaluasi *Cluster*

Metode yang digunakan peneliti dalam menentukan evaluasi *cluster* menggunakan *Davies-Bouldin Index* (DBI). Dalam suatu pengelompokan, kohesi didefinisikan sebagai jumlah dari kedekatan data terhadap *centroid* dari *cluster* yang diikuti. Sedangkan separasi didasarkan pada jarak antar *centroid* dari *clusternya*.

Sum of square within cluster (SSW) merupakan persamaan yang digunakan untuk mengetahui matrik kohesi dalam sebuah *cluster* ke-i yang dirumuskan sebagai berikut:

$$SSWi = \frac{1}{mi}\sum_{j=i}^{mi} d(xi, ci) \dots\dots\dots (1)$$

Keterangan:
m = Jumlah data dalam *cluster* ke-i
c = *Centroid cluster* ke-i
d = Jarak

Sum of square between cluster (SSB) merupakan persamaan yang digunakan untuk mengetahui separasi antar *cluster* yang dihitung menggunakan persamaan:

$$SSBij = d(ci, cj) \dots\dots\dots (2)$$

Keterangan:
d = Jarak
c = *centroid cluster* ke-i

Setelah nilai kohesi dan separasi diperoleh, kemudian dilakukan pengukuran *rasio* (Rij) untuk mengetahui nilai perbandingan antara *cluster* ke-i dan *cluster* ke-j. *Cluster* yang baik adalah *cluster* yang memiliki nilai kohesi sekecil mungkin dan separasi yang sebesar mungkin. Nilai *rasio* dihitung menggunakan persamaan sebagai berikut:

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}} \dots\dots\dots (3)$$

Nilai *rasio* yang diperoleh tersebut digunakan untuk mencari nilai *davies bouldin-index* (DBI) dari persamaan berikut:

$$DBI = \frac{1}{k}\sum_{i=1}^{k} \quad max_{i \neq j}(R_{ij}) \ldots\ldots\ldots (4)$$

Dari persamaan tersebut, k merupakan jumlah *cluster* yang digunakan. Semakin kecil nilai DBI yang diperoleh (non-negatif >= 0), maka semakin baik *cluster* yang diperoleh dari pengelompokan k-means yang digunakan (A. F. Muhammad, 2015).

III. METODE

Dalam Pengerjaan penelitian ini, terdapat tiga tahapan yang akan dilakukan oleh peneliti. Tahapannya terdiri dari tahap inisiasi, implementasi, dan evaluasi. Berikut adalah sistematika penelitian yang digunakan dalam penelitian ini seperti dalam gambar dibawah ini:

GAMBAR 3. 1
SISTEMATIKA PENYELESAIAN MASALAH

A. Insiasi

Pada tahap proses inisiasi ini proses dimulai dengan mengidentifikasi latar belakang masalah. Selanjutnya menentukan perumusan masalah dari studi kasus yang ingin diangkat. Kemudian, tahap selanjutnya dengan menentukan tujuan penelitian. Proses terakhir pada tahap ini adalah menentukan batasan masalah terkait studi kasus yang ingin diangkat.

B. Implementasi

Pada tahap implementasi, dilakukan pengumpulan data mengenai penyebaran Covid-19 yang ada di Kota Bandung kemudian selanjutnya akan di olah dengan menggunakan RapidMiner dengan algoritma k-means *clustering* dan perhitungan secara manual.

1. Pengambilan Data

Pada tahap pengambilan data peneliti melakukan pengambilan data dari Dinas Kesehatan Kota Bandung. Proses ini diawali dengan peneliti meminta surat izin untuk mempelajari dan mencari referensi yang berkaitan dengan topik penelitian dari kampus, kemudian peneliti meminta surat izin kepada Badan Kesatuan Bangsa dan Politik Kota Bandung untuk meminta data yang akan dilakukan penelitian, kemudian selanjut nya peneliti menerima data dari Dinas Kesehatan Kota Bandung. Data yang digunakan dalam penelitian ini adalah data rekap Covid-19 pada tanggal 1 Mei 2022 hingga 29 Juni 2022.

2. Data Seleksi

Pada tahap ini peneliti melakukan pengambilan data dari Dinas Kesehatan Kota Bandung. Setelah itu peneliti melakukan seleksi data, pada tahap ini peneliti dilakukan pemilihan data yang akan digunakan untuk penelitian.

3. Menerapkan Algoritma K-Means

Pada tahap menerapkan algoritma k-means dilakukan dengan bantuan *tools* RapidMiner dan perhitungan manual. Implementasi perhitungan manual dilakukan untuk mengetahui bagaimana proses berjalannya algoritma k-means.

C. Evaluasi

Pada tahap evaluasi, terdiri dari 3 tahapan yaitu analisis hasil, evaluasi *cluster* dan menarik kesimpulan.

1. Analisis Hasil *Cluster*

Pada tahap ini peneliti menganalisa terkait hasil yang diperoleh pada saat implementasi algoritma k-means. Hasil yang diperoleh berupa *cluster* yang dikelompokkan berdasarkan kemiripan karakteristik antara satu sama lainnya. Kemudian menentukan kategori *cluster* berdasarkan kondisi setiap *cluster.*

2. Evaluasi *Cluster*

Pada tahap ini peneliti mengevaluasi terkait hasil *cluster* yang di dapat dengan cara mencari nilai DBI dengan bantuan *tools* RapidMiner.

3. Menarik Kesimpulan dan Saran

Pada tahap kesimpulan & saran merupakan tahapan terakhir dimana nanti peneliti akan menarik sebuah kesimpulan dan saran dari permasalahan yang terjadi dalam penelitian tugas akhir ini.

IV. HASIL DAN PEMBAHASAN

A. Data *Preprocessing*

Pada tahap *preprocessing* data, peneliti melakukan pengambilan data dari Dinas Kesehatan Kota Bandung. Data yang di dapat oleh peneliti merupakan data Rekap Covid-19 pada tanggal 1 Mei 2022 hingga 29 Juni 2022. Data Covid-19 yang diperoleh peneliti memiliki kolom Kecamatan, OTG, ODP, PDP, Suspek, dan Positif. Kolom OTG memiliki atribut dipantau, selesai, dan total. Kolom ODP memiliki atribut masih dipantau, selesai, dan total. Kolom PDP memiliki atribut masih dirawat, selesai, dan total. Kolom Suspek memiliki atribut dipantau, discarded, dan total. Kolom Positif memiliki atribut aktif, sembuh, meninggal, dan total.

Pada tahap selanjutnya melakukan *selection* data, pada tahap ini peneliti memilih atribut apa saja yang akan digunakan dalam penelitian. Dalam penelitian ini, peneliti hanya memilih atribut Kecamatan, aktif, sembuh, meninggal, dan kenaikan pasien aktif. Atribut

kenaikan pasien aktif diproleh dari selisih total kolom positif pada tanggal 1 Mei 2022 dan 29 Juni 2022. Atribut ini diperlukan untuk memenuhi penelitian yang dilakukan oleh peneliti.

Pada tahap selanjutnya peneliti memeriksa apakah data yang akan di proses memiliki *missing value* dan *null*. Peneliti memeriksa data Covid-19 dengan bantuan *tools* RapidMiner. Data Covid-19 yang diperoleh peneliti tidak memiliki *missing value* dan kosong sehingga tidak diperlukan untuk pembersihan data. Gambar di bawah ini merupakan hasil pemeriksaan pada data yang akan di proses:

GAMBAR 4. 1
MEMERIKSA DATA *NULL* DAN *MISSING VALUE*

Pada Tabel 4.1 dibawah ini adalah *output* yang dihasilkan setelah melakukan *preprocessing* data. Data inilah yang selanjutnya akan dilakukan *clustering* menggunakan perhitungan manual dan dengan bantuan *tools* RapidMiner. Berikut tampilan hasil datanya:

TABEL 4. 1
DATA SIAP PROSES

NO	KECAMATAN	AKTIF	SEMBUH	MENINGGAL	KENAIKAN PASIEN AKTIF
1	ANDIR	14	3026	61	31
2	ANTAPANI	17	4670	80	35
3	ARCAMANIK	13	4017	63	24
4	ASTANA ANYAR	10	2209	33	23
5	BABAKAN CIPARAY	15	2342	33	28
6	BANDUNG KIDUL	10	2462	43	20
7	BANDUNG KULON	11	2964	33	24
8	BANDUNG WETAN	5	1137	15	22
9	BATUNUNGGAL	9	3385	82	15
10	BOJONGLOA KALER	1	2295	35	7
11	BOJONGLOA KIDUL	9	2318	29	28
12	BUAH BATU	10	3988	73	25
13	CIBEUNYING KALER	11	2268	36	21
14	CIBEUNYING KIDUL	7	3199	86	21
15	CIBIRU	2	2161	41	5
16	CICENDO	20	3219	71	46
17	CIDADAP	4	1567	26	16
18	CINAMBO	1	971	22	1
19	COBLONG	28	4458	91	70
20	GEDEBAGE	3	1846	33	8
21	KIARACONDONG	5	3862	82	10
22	LENGKONG	39	3707	48	54
23	MANDALAJATI	1	2302	35	13
24	PANYILEUKAN	1	2289	49	3
25	RANCASARI	7	3762	78	16
26	REGOL	13	3231	70	36
27	SUKAJADI	15	3711	42	44
28	SUKASARI	23	3479	38	51
29	SUMUR BANDUNG	11	1581	13	24
30	UJUNG BERUNG	14	2835	35	26

B. Algoritma K-Means

Salah satu algoritma yang digunakan untung *clustering* adalah k-means. Algoritma k-means ini berguna untuk mengelompokkan data menjadi beberapa kelompok. Data – data dipilih berdasarkan kriteria yang telah dikumpulkan lalu dijadikan menjadi satu dalam suatu *cluster*. Berikut adalah tahapan-tahapan untuk melakukan optimasi menggunakan algoritma k-means :

1. Menentukan jumlah *cluster*.
2. Menentukan *centroid* awal atau titik pusat *cluster* awal.
3. Menghitung jarak terdekat pada setiap data *centroid*.
4. Mengelompokkan data berdasarkan nilai minimum pada jarak terdekat.
5. Menghitung kembali setiap objek menggunakan *centroid* baru.

1. Menentukan Jumlah *Cluster*

Jumlah *cluster* merupakan jumlah kelompok yang akan dihasilkan. Jumlah ini menyesuaikan berdasarkan kebutuhan analisis dalam penelitian. Peneliti menentukan jumlah *cluster* berdasarkan *performance* yang di uji dengan RapidMiner. Dalam penelitian ini peneliti akan menguji dua *cluster* hingga sepuluh *cluster* Berikut merupakan hasil uji *performance* yang dilakukan oleh peneliti :

TABEL 4. 2
PERFORMANCE

Cluster	*Performance*
2 *Cluster*	-0,515
3 *Cluster*	-0,552
4 *Cluster*	-0,457
5 *Cluster*	-0,405
6 *Cluster*	-0,366
7 *Cluster*	-0,382
8 *Cluster*	-0,384
9 *Cluster*	-0,313
10 *Cluster*	-0,458

Berdasarkan Tabel 4.2 *performance* yang memiliki hasil *cluster* terbaik terdapat pada 9 *cluster*. Dalam menentukan jumlah *cluster* semakin kecil nilai yang di dapat maka semakin baik hasil *cluster* nya.

2. Menentukan *Centroid* Awal

Centorid awal merupakan titik pusat pada *cluster* pertama. Pada tahap ini peneliti mengambil satu *sample* data dari setiap *cluster* pada RapidMiner. Dikarenakan peneliti ingin mengetahui alur proses algoritma ini berjalan. Pemilihan titik *centroid* awal akan sangat berpengaruh pada hasil *cluster*. Pada tahap ini peneliti memilih *centroid* awal sebagai berikut :

TABEL 4. 3
CENTROID AWAL

No Data	Kecamatan	Aktif	Sembuh	Meninggal	Kenaikan Pasien Akti
4	Astana Anyar	10	2209	33	23
3	Arcamanik	13	4017	63	24
8	Bandung Wetan	5	1137	15	22
17	Cidadap	4	1567	26	16
14	Cibeunying Kidul	7	3199	86	21
22	Lengkong	39	3707	48	54
2	Antapani	17	4670	80	35
1	Andir	14	3026	61	31
9	Batununggal	9	3385	82	15

3. Menghitung Jarak Terdekat

Pada tahap menghitung *centroid* terdekat dapat dilakukan dengan *euclidean distance*. Pada tahap ini jarak terdekat antara data dengan pusat *cluster* akan menentukan ke dalam *cluster* yang mana. *Euclidean distance* memiliki rumus seperti dibawah ini :

$$de = \sqrt{(xi - si)^2 + (yi - ti)^2} \dots\dots\dots (5)$$

Keterangan :
(x, y) = nilai objek,
(s, t) = nilai *centroid* awal,
i = banyak nya objek.

Peneliti untuk mengetahui jarak terdekat antara data dengan pusat *cluster* dilakukan dengan cara *euclidean distance*. Berikut merupakan hasil dari perhitungan jarak terdekat pada iterasi 1:

TABEL 4. 4
JARAK TERDEKAT ITERASI 1

	Jarak Terdekat *Centroid* Ke *Cluster*								
kecamatan	C0	C1	C2	C3	C4	C5	C6	C7	C8
Andir	6683 41	9821 35	3570 527	2130 141	30661	4644 84	2703 116	0	1295 83
Antapani	6058 881	4268 23	1248 6495	9631 899	2164 083	9287 76	0	2703 116	1651 637
Arcamanik	3269 768	0	8296 716	6003 942	6696 68	9725 1	4268 23	9821 35	3998 70
Astana Anyar	0	3269 768	1149 514	4122 68	9829 16	2245 219	6058 881	6683 41	1385 442
Babakan Ciparay	1771 9	2806 543	1452 395	6008 29	7373 15	1864 150	5421 844	4686 50	1090 425
Bandung Kidul	6411 8	2418 444	1756 418	8013 36	5450 22	1551 235	4876 865	3185 45	8534 76
Bandung Kulon	5700 27	1109 711	3338 263	1951 729	5804 7	5532 02	2912 772	4680	1797 25
Bandung Wetan	1149 514	8296 716	0	1850 58	4256 888	6607 047	1248 6495	3570 527	5058 046
Batununggal	1385 442	3998 70	5058 046	3308 266	3465 0	1063 91	1651 637	1295 83	0
Bojongloa Kaler	7665	2966 369	1341 593	5301 49	8200 19	1996 160	5643 450	5356 26	1190 381
Bojongloa Kidul	1192 3	2887 777	1394 997	5641 59	7794 61	1930 388	5534 562	5023 02	1141 467
Buah Batu	3166 445	945	8131 579	5863 537	6227 09	8045 6	4652 80	9256 28	3637 91
Cibeunying Kaler	3495	3059 741	1279 609	4915 33	8692 65	2071 982	5771 742	5752 92	1249 843
Cibeunying Kidul	9829 16	6696 68	4256 888	2667 052	0	2606 29	2164 083	3066 1	3465 0
Cibiru	2700	3445 592	1049 544	3531 84	1079 730	2392 603	6297 517	7493 13	1499 964

Cicendo	102208 83	637359	433845 1	273204 5	1263	238756	210560 6	37580	28649
Cidadap	412268	600394 2	185058	0	266705 2	458156 33	963189 9	213014 1	330826 6
Cinambo	153325 8	928033 8	28050	355460	496848 6	748921 9	136871 37	422545 9	583120 0
Goblong	506359 2	197396	110371 44	836504 6	158752 8	566117	463011	205305 9	115454 4
Gedebage	132001	471440 7	503203	779555	183359 1	346569 8	797792 8	139372 4	237097 7
Kiaracondong	273498 4	245900	743025 8	527019 8	439708	271511	653505	699787	227558
Lengkong	224521 9	972511	660704 7	458156 33	260629	0	928776	464484	106391
Mandalajati	8762	294214 2	135771 0	540318	807280	197591 33	560994 9	525189	117511 0
Panyileukan	7065	298663 3	132862 5	521985	829799	201336 4	567116 2	544110	120245 7
Rancasari	241388 6	65320	689463 2	482073 2	317058	5401	824839	542217	142148
Regol	104602 5	617989	438806 5	277124 1	1511	227410	207082 6	421322	24305
Sukajsadi	225653 31	94479	662669 9	459787 87	264617	176	921208	469756	108723
Sukasari	161372 2	290808	548635 2	365713 2	816220	521099	142050 7	206147	12082
Sumur Bandung	394786	593659 8	197150	436	262326 6	452202 9	954653 7	209038 1	325926 0
Ujung Berung	391893	139791 33	288362 9	160801 5	135129	761362	336933 4	37182	304835

4. Pengelompokkan data

Pada tahap ini peneliti melakukan pengelompokkan data berdasarkan nilai minimum jarak terdekat data pada *centroid*. Berikut merupakan hasil pengelompokkan data berdasarkan jarak data dengan *centroid* pada iterasi 1:

TABEL 4. 5
HASIL *CLUSTER* ITERASI 1

Kecamatan	*Cluster*
Andir	C7
Antapani	C6
Arcamanik	C1
Astana Anyar	C0
Babakan Ciparay	C0
Bandung Kidul	C0
Bandung Kulon	C7
Bandung Wetan	C2
Batununggal	C8
Bojongloa Kaler	C0
Bojongloa Kidul	C0
Buah Batu	C1
Cibeunying Kaler	C0
Cibeunying Kidul	C4
Cibiru	C0
Cicendo	C4
Cidadap	C3
Cinambo	C2
Goblong	C6
Gedebage	C3
Kiara Condong	C1
Lengkong	C5
Mandalajati	C0
Panyileukan	C0
Rancasari	C5
Regol	C4
Sukajadi	C5
Sukasari	C8
Sumur Bandung	C3
Ujung Berung	C7

5. Menghitung *Centroid* Baru

Menentukan *centroid* baru untuk iterasi berikutnya dengan mencari nilai rata-rata nilai dari setiap *cluster* yang ada. Berikut adalah nilai *centroid* baru untuk iterasi 2 :

TABEL 4. 6
CENTROID BARU

Cluster berdasarkan jarak terdekat	Aktif	Sembuh	Meninggal	Kenaikan Pasien Aktif
Cluster 0	6,66	2294	37,11	16,44
Cluster 1	9,33	3955,66	72,66	19,66
Cluster 2	3	1054	18,5	11,5
Cluster 3	6	1664,66	24	16
Cluster 4	13,33	3216,33	75,66	34,33
Cluster 5	20,33	3726,66	56	38
Cluster 6	22,5	4564	85,5	52,5
Cluster 7	13	2941,66	43	27
Cluster 8	16	3432	60	33

Langkah selanjut nya sama dengan sebelumnya yaitu menghitung kembali jarak pada setiap data dengan *centroid* dengan cara persamaan *euclidean distance*. Berikut merupakan hasil dari perhitungan jarak pada iterasi ke 2:

TABEL 4. 7
JARAK TERDEKAT ITERASI 2

kecamatan	Jarak Terdekat *Centroid* Ke *Cluster*								
	C0	C1	C2	C3	C4	C5	C6	C7	C8
Andir	5366613,9	8645493	3890982	1854830	3645 3,67	4910 14,11	2366 5155	7453,1111	1648443
Antapani	5647570	5105687	1307980 5	9035536	2113170	8904 66,1	1157 8	2988573	1533049
Arcamanik	2969463	3877,667	8781516	5535064	6413 34,7	8454 5,78	3005 37	1156751	3423 18
Astana Anyar	7288,21	3052430	1334375	2964 32,8	1016673	2304076	5549664	5369 19,4	1496564
Babakan Ciparay	2462,765	2605569	1659439	4590 14,4	76 6321	1917936	4940648	3597 03,1	1188855
Bandung Kidul	2827 4,65	2231921	1983144	6361 21,4	5702 94,7	1599885	4421279	2301 32,1	9413 64
Bandung Kulon	4489 78,3	9849 96,7	3648475	1688417	6560 1,67	5823 94,8	2563580	609,7778	2198 39
Bandung Wetan	1339170	7948217	7013,5	2785 50,1	4327468	6708326	1175 0247	3257639	52691 82
Batununggal	1192300	3257 69,7	5437612	2962915	2886 6,67	1179 52,4	1391473	1982 13,4	3024
Bojongloa Kaler	100,321	2759401	1540376	3975 27,1	8512 68,3	2051091	5153003	4186 53,8	1294085
Bojongloa Kidul	777,6543	2683929	1598085	4270 16,4	8092 25	1985182	5048322	3891 61,1	1241989
Buah Batu	2871001	1074,667	8611516	5400364	595567	6876 3,44	3327 01	1095720	3093 75
Cibeunying Kaler	702,321	2849567	1474201	3641 85,1	9010 89,7	2128407	5275070	4539 13,8	1355621
Cibeunying Kidul	8214 36,2	5727 26,3	4605676	2358049	591,3333	2796 34,4	1864233	6811 1,44	55118
Cibiru	1783	3222	1225	2467	1115	2452	5778	6099	1616
	9,77	054	999	60,8	802	644	666	39,4	600
Cicendo	8576 60,3	5433 84,7	4691189	2419075	171,6667	2580 14,8	1809280	7806 5,78	4566 63
Cidadap	5286 55,3	5707925	2632 46,5	9544, 778	2723113	4665560	8986900	1890127	3479682
Cinambo	1750802	8911159	7013,5	4814 07,4	5045525	7596243	1291 6355	3884656	6059004
Goblong	4688690	2552 27	1159 5920	7810 1388	1543258	5371 05,1	1157 8	2303435	1055018
Gedebage	2007 95,9	4452409	6274 86,5	3302 9,78	1880338	3538353	7392280	1200956	2516763
Kiaracondong	2460682	8958, 333	7888 9901	4831 6755	417526	1979 0,44	4946 40	8488 31,4	1859 24
Lengkong	1998 130	6365 2	7041 1222	4173 178	2419 31,7	725,4444	7358 74	5865 15,11	76233
Mandalajati	85,98 765	2736085	1557 781	4063 28,8	8381 26,7	2030 760	5120 776	4094 45,4	1277940
Panyileukan	352,7654	2778624	1526 230	3905 91,1	8616 52,3	2068179	5179 429	4265 97,8	1307485
Rancasari	2156696	37551	7336 829	4401 724	2981 00	2229, 778	6446 08	6742 98,8	1095 22
Regol	8794 39,4	5254 19,3	4742 592	2455 923	250,3333	2458 92,8	1777 411	8452 3,78	4051 3
Sukajsadi	2008 681	614000	7061 270	4188 597	2459 23,7	482,778	7295 81	5921 65,8	7828 7
Sukasari	1405 436	2294 08,3	5882 586	3293 243	70700	6183 4,44	1179 484	2893 38,1	3024
Sumur Bandung	5090 11,8	5642 622	2779 23,5	7190, 111	26 7835 1	4605 940	8904 369	1852 325	3428 496
Ujung Berung	2927 84,1	1257 357	3172 455	1369 909	1471 39	7956 60,8	2992 702	1144 3,78	3570 85

Tahap selanjut nya sama seperti sebelum nya yaitu mengelompokkan data berdasarkan nilai minimum jarak terdekat data dengan *centroid*. Berikut

merupakan hasil pengelompokkan data berdasarkan nilai minimum jarak data dengan *centroid* pada iterasi 2:

TABEL 4. 8
HASIL *CLUSTER* ITERASI 2

Kecamatan	*Cluster*
Andir	C7
Antapani	C6
Arcamanik	C1
Astana Anyar	C0
Babakan Ciparay	C0
Bandung Kidul	C0
Bandung Kulon	C7
Bandung Wetan	C2
Batununggal	C8
Bojongloa Kaler	C0
Bojongloa Kidul	C0
Buah Batu	C1
Cibeunying Kaler	C0
Cibeunying Kidul	C4
Cibiru	C0
Cicendo	C4
Cidadap	C3
Cinambo	C2
Goblong	C6
Gedebage	C3
Kiara Condong	C1
Lengkong	C5
Mandalajati	C0
Panyileukan	C0
Rancasari	C5
Regol	C4
Sukajadi	C5
Sukasari	C8
Sumur Bandung	C3
Ujung Berung	C7

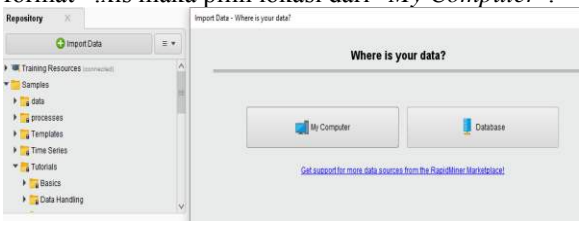
Berdasarkan Tabel 4.5 dengan Tabel 4.8 pengelompokkan data berdasarkan nilai minimum jarak terdekat pada *centroid* mencapai hasil yang sama tanpa ada perpindahan, maka dari itu implementasi algoritma k-means pada perhitungan manual dapat dihentikan.

C. Implementasi Algoritma K-Means Menggunakan RapidMiner

Penerapan algoritma k-means menggunakan RapidMiner untuk bertujuan mengelompokkan data penyebaran Covid-19 di Kota Bandung. Untuk menghasilkan *cluster* dari data yang telah peneliti siapkan, data terlebih dahulu di *import* ke dalam RapidMiner. Kemudian selanjutnya pembuatan lembar proses, dimana pada proses ini berisi mengenai proses penerapan algoritma maupun proses pengujian menggunakan metode-metode pengujian yang ada pada RapidMiner. Langkah-langkah pengerjaan pada RapidMiner dapat dilihat pada penjelasan-penjelasan dibawah ini :

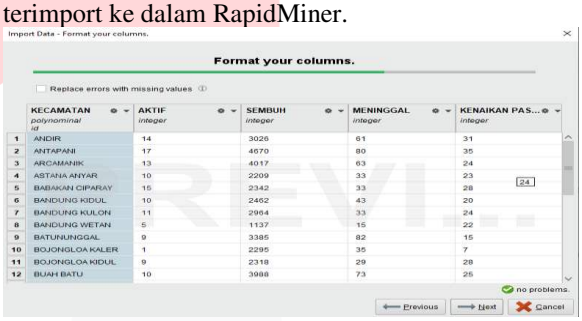
1. *Import* data

Sebelum menerapkan algoritma k-means data dengan format *.xls di import terlebih dahulu ke dalam RapidMiner dengan cara klik "*import* data", selanjutnya pilih lokasi penyimpanan yang ingin di uji. Dikarenakan data yang dimiliki oleh peneliti memiliki format *.xls maka pilih lokasi dari "*My Computer*".



GAMBAR 4. 2
IMPORT DATA

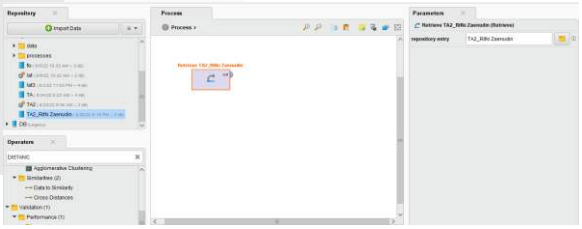
Selanjutnya pilih data yang ingin di *import* ke dalam RapidMiner. Apabila data yang dipilih terindikasi bahwa tidak ada masalah maka akan muncul dibagian bawah kanan "*no problems*" pada proses *import* data. Lalu lanjutkan hingga *finish*, sehingga data berhasil terimport ke dalam RapidMiner.



GAMBAR 4. 3
PROSES *IMPORT* DATA

2. Lembar Proses

Setelah data di *import*, data dimasukkan ke halaman proses. Dengan menyeret atau *drag and drop* data yang telah di *import*. Nama file yang data yang ingin digunakan dalam penelitian disini adalah TA2_Rifki Zaenudin. Kemudian peneliti *drag and drop* file data TA2_Rifki Zaenudin ke halaman proses.



GAMBAR 4. 4
LEMBAR PROSES

3. Menerapkan Algoritma K-Means

Pada tahap ini memasukkan algoritma k-means pada lembar proses yang telah terbentuk sebelumnya untuk mengetahui pengelompokkan dari proses yang telah dibuat. Cari "k-means" pada Operator lalu *drag and drop* disebelah data "TA2_Rifki Zaenudin" lalu hubungkan ke "*clustering*" dan "*res*".

GAMBAR 4. 5
MENERAPKAN ALGORITMA

D. Analisis Hasil *Cluster*

Berdasarkan penelitian yang dilakukan oleh peneliti pada perhitungan manual proses k-means *clustering* dilakukan sebanyak 2 iterasi. Pada iterasi kedua memiliki hasil yang sama pada iterasi kesatu, maka proses pengolahan berhenti sampai disini. Hasil pengolahan dari data Covid-19 Kota Bandung didapatkan C0 sebanyak 9 Kecamatan, C1 sebanyak 3 Kecamatan, C2 sebanyak 2 Kecamatan, C3 sebanyak 3 Kecamatan, C4 sebanyak 3 Kecamatan, C5 sebanyak 3 Kecamatan, C6 sebanyak 2 Kecamatan, C7 sebanyak 3 Kecamatan, dan C8 sebanyak 2 Kecamatan, sehingga total keseluruhan Kecamatan pada Kota Bandung memiliki 30 Kecamatan.

Peneliti dapat melihat karakteristik masing-masing *cluster* berdasarkan nilai *centroidnya*. Berdasarkan penelitian, peneliti melihat bahwa nilai terbesar untuk setiap atribut ada pada atribut yang sembuh, tetapi setiap *cluster* memiliki karakteristik yang berbeda. Berikut merupakan tabel karakteristik berdasarkan nilai *centroid* :

TABEL 4. 9
KARAKTERISTIK NILAI *CENTROID*

Karakteristik Berdasarkan Nilai *Centroid*				
Cluster	Aktif	Sembuh	Meninggal	Kenaikan Pasien Aktif
Cluster 0	3,33	85	4,11	6,55
Cluster 1	3,66	61,33	9,66	4,33
Cluster 2	2	83	3,5	10,5
Cluster 3	2	97,66	2	0
Cluster 4	6,33	17,33	10,33	13,33
Cluster 5	18,66	19,66	8	16
Cluster 6	5,5	106	5,5	17,5
Cluster 7	1	84,33	18	4
Cluster 8	7	47	22	18

Peneliti untuk mengetahui kondisi hasil *cluster* yang di dapat, peneliti melakukan list *ranking*. Dalam penelitian ini peneliti membagi menjadi 3 kategori yaitu *cluster* tinggi, *cluster* sedang, dan *cluster* rendah. Berikut merupakan tabel list *ranking* beserta kategori berdasarkan nilai dari yang tertinggi hingga yang terendah.

TABEL 4. 10
MENENTUKAN KATEGORI

Ranking	Kategori
1-3	Tinggi
4-6	Sedang
7-9	Rendah

Pada tahap selanjut nya peneliti melakukan *ranking* pada setiap atribut berdasarkan nilai tertinggi hingga yang terendah. Berikut merupakan tabel *ranking* berdasarkan nilai *centroid* :

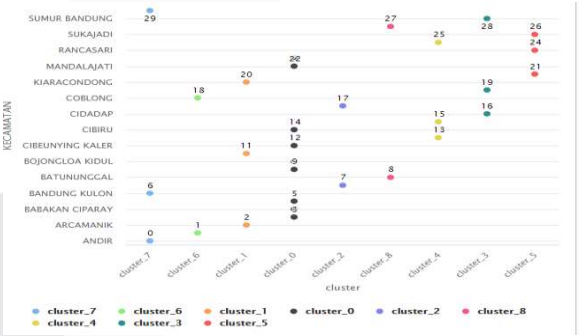
TABEL 4. 11
RANKING

Aktif	Sembuh	Meninggal	Kenaikan Pasien Aktif
6	3	7	6
5	6	4	7
7	5	8	5
8	2	9	9
3	9	3	4
1	8	5	3
4	1	6	2
9	4	2	8
2	7	1	1

Berdasarkan Tabel 4.11 dapat diketahui bahwa C0 merupakan *Cluster* dalam kategori sedang, C1 merupakan *Cluster* dalam kategori sedang, C2 merupakan *Cluster* dalam kategori rendah, C3 merupakan *Cluster* dalam kategori rendah, C4 merupakan *Cluster* dalam kategori tinggi, C5 merupakan *Cluster* dalam kategori tinggi, C6 merupakan *Cluster* dalam kategori tinggi, C7 merupakan *Cluster* dalam kategori rendah, dan C8 merupakan *Cluster* dalam kategori tinggi.

E. Hasil Visualisasi

Pada tahap hasil visualisasi merupakan penggambaran hasil *cluster* yang telah dilakukan oleh peneliti. *Plot type* visualisasi yang digunakan peneliti yaitu *scatter bubble*. Berikut merupakan hasil visualisasi *cluster* pada saat penelitian :



GAMBAR 4. 6
VISUALISASI

Berdasarkan Gambar 5.6 dapat disimpulkan bahwa *cluster* 0 memiliki 9 data yang terdapat pada data ke 3, 4, 5, 9, 10, 12, 14, 22, dan 23. *Cluster* 1 memiliki 3 data yang terdapat pada pada data ke 2, 11, dan 20. *Cluster* 2 memiliki 2 data yang terdapat pada pada data ke 7, dan 17. *Cluster* 3 memiliki 3 data yang terdapat pada pada data ke 16, 19, dan 28. *Cluster* 4 memiliki 3 data yang terdapat pada pada data ke 13, 15, dan 25. *Cluster* 5 memiliki 3 data yang terdapat pada pada data ke 21, 24, dan 26. *Cluster* 6 memiliki 2 data yang terdapat pada pada data ke 1, dan 18. *Cluster* 7 memiliki 3 data yang terdapat pada pada data ke 0, 6,

dan 29. *Cluster* 8 memiliki 2 data yang terdapat pada pada data ke 8, dan 27.

V. KESIMPULAN

A. Kesimpulan

Adapun kesimpulan dari peneliti untuk pengembangan lebih lanjut dari penelitian ini yaitu:

1. Berdasarkan penelitian yang telah dilakukan clustering penyebaran Covid-19 di Kota Bandung memiliki 9 cluster, yaitu C0 merupakan *cluster* dalam kategori sedang, C1 merupakan *cluster* dalam kategori sedang, C2 merupakan *cluster* dalam kategori rendah, C3 merupakan *cluster* dalam kategori rendah, C4 merupakan *cluster* dalam kategori tinggi, C5 merupakan *cluster* dalam kategori tinggi, C6 merupakan *cluster* dalam kategori tinggi, C7 merupakan *cluster* dalam kategori rendah, dan C8 merupakan *cluster* dalam kategori tinggi.
2. Berdasarkan penelitian yang telah dilakukan oleh peneliti pengujian terhadap *cluster* menggunakan bantuan *tools* RapidMiner. Nilai yang didapat pada saat peneliti menguji *performance* merupakan nilai DBI. Pada penelitian ini peneliti menguji terhadap 2 *cluster* hingga 10 *cluster*. Pada penelitian ini yang memiliki hasil *cluster* terbaik terdapat pada 9 *cluster*.
3. Berdasarkan penelitian yang dilakukan oleh peneliti analisis *clustering* dilakukan dengan cara melihat hasil *cluster* yang diperoleh pada saat penelitian. Kemudian peneliti melihat karakteristik setiap *cluster* berdasarkan nilai *centroid*. Kemudian selanjutnya peneliti melakukan *ranking* nilai *centroid* setiap *cluster* untuk mengetahui *cluster* tersebut dalam kategori tinggi, sedang, dan rendah.
4. Berdasarkan penelitian yang telah dilakukan, implementasi algoritma k-means dimulai dengann peneliti mencari nilai *performance* dengan bantuan *tools* RapidMiner. Selanjut nya peneliti menentukan *centroid* awal atau titik pusat pada *cluster*. Selanjutnya peneliti dapat menghitung jarak data pada *centroid* dengan menggunakan rumus *euclidean distance*. Selanjutnya peneliti dapat mengelompokkan data berdasarkan nilai minimum antara jarak data dengan *centroid*.

B. Saran

Adapun saran dari peneliti untuk pengembangan lebih lanjut dari penelitian ini yaitu:

1. Penelitian selanjutnya dapat dilakukan secara lebih detail dengan menggunakan data per Kelurahan yang ada di Kota Bandung dengan data kasus Covid-19 yang lebih detail lagi, dengan menggunakan algoritma yang bebeda misalnya fuzzy c-means atau k-medoids.
2. Penelitian selanjutnya dalam pengujian algoritma k-means dapat menggunakan bantuan *tools* lainnya seperti *phyton* dan *matlab* agar pembahasanya dapat lebih luas lagi.
3. Penelitian selanjutnya untuk pengujian *cluster* dapat menggunakan metode lainnya tidak hanya DBI sama melainkan dapat menggunakan metode *elbow* dan dB untuk mengetahui *cluster* yang terbaik.

REFERENSI

Bastian, A. (2018). Penerapan algoritma k-means clustering analysis pada penyakit menular manusia (studi kasus kabupaten Majalengka). *Jurnal Sistem Informasi*, *14*(1), 28-34.

Damanik, Y. F. S. Y., Sumarno, S., Gunawan, I., Hartama, D., & Kirana, I. O. (2021). Penerapan Data Mining Untuk Pengelompokan Penyebaran Covid-19 Di Sumatera Utara Menggunakan Algoritma K-Means. *Jurnal Ilmu Komputer dan Informatika*, *1*(2).

Darmansah, D. D., & Wardani, N. W. (2021). Analisis Pesebaran Penularan Virus Corona di Provinsi Jawa Tengah Menggunakan Metode K-Means Clustering. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, *8*(1), 105-117.

Dwitri, N., Tampubolon, J. A., Prayoga, S., Zer, F. I. R., & Hartama, D. (2020). Penerapan algoritma K-Means dalam menentukan tingkat penyebaran pandemi COVID-19 di Indonesia. *(JurTI) Jurnal Teknologi Informasi*, *4*(1), 128-132.

Fira, A., Rozikin, C., & Garno, G. (2021). Komparasi Algoritma K-Means dan K-Medoids Untuk Pengelompokkan Penyebaran Covid-19 di Indonesia. *Journal of Applied Informatics and Computing*, *5*(2), 133-138.

Fitriyani, N. K., & Abdulloh, F. F. (2021). Analisis Algoritma K-Means dalam Pengelompokkan Persebaran Covid-19 di Indonesia. *MEANS (Media Informasi Analisa dan Sistem)*, 180-183.

Fitriyani, V. (2021, May). Analisis Clustering Provinsi Indonesia Berdasarkan Persebaran Virus Corona (Covid-19) Menggunakan Algoritma K-Means. In *Prosiding Seminar Pendidikan Matematika dan Matematika* (Vol. 3).

Gayatri, L., & Hendry, H. (2021). Pemetaan Penyebaran Covid-19 Pada Tingkat Kabupaten/Kota Di Pulau Jawa Menggunakan Algoritma K-Means Clustering. *Sebatik*, *25*(2), 493-499.

Marlina, D., Putri, N. F., Fernando, A., & Ramadhan, A. (2018). Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokkan Wilayah Sebaran Cacat pada Anak. *J. CoreIT*, *4*(2), 64.

Mirantika, N. (2021). Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Covid-19 di Provinsi Jawa Barat. *Nuansa Informatika*, *15*(2), 92-98.

Nabila, Z., Isnain, A. R., Permata, P., & Abidin, Z. (2021). Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means. *Jurnal Teknologi Dan Sistem Informasi*, *2*(2), 100-108.

Sepri, D., & Fimazid, Y. (2021). Pengelompokan Penyebaran Covid-19 di Kota Padang Menggunakan Algoritma K-Medoids. *Insearch: Information System Research Journal*, *1*(02), 39-45.

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.