

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.

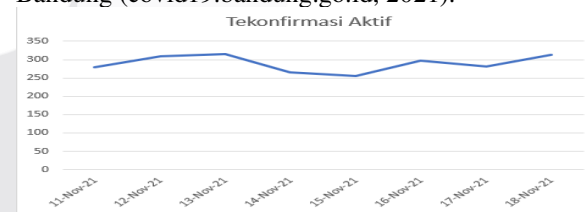
a cluster in the high category, C7 is a cluster in the low category, and C8 is a cluster in the high category.

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). *Knowledge 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 mencari 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 memiliki 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 kluster di awal kemudian secara iteratif partisi kluster ini diperbaiki hingga tidak terjadi perubahan yang signifikan pada partisi kluster (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 *cluster*nya.

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* (R_{ij}) 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:

$$R_{ij} = \frac{SSWi + SSWj}{SSBij} \dots \dots \dots (3)$$

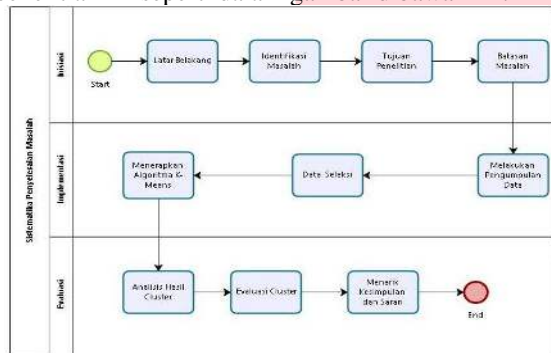
Nilai *rasio* yang diperoleh tersebut digunakan untuk mencari nilai *davies bouldin-index* (DBI) dari persamaan berikut:

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

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. Inisiasi

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 diperoleh 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	ARCAMANI K	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	BATUNUNG GAL	9	3385	82	15
10	BOJONGLO A KALER	1	2295	35	7
11	BOJONGLO A KIDUL	9	2318	29	28
12	BUAH BATU	10	3988	73	25
13	CIBEUNYIN G KALER	11	2268	36	21
14	CIBEUNYIN G KIDUL	7	3199	86	21
15	CIBIRU	2	2161	41	5

NO	KECAMATAN	AKTIF	SEMBUH	MENINGGAL	KENAikan PASIEN AKTIF
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	KIARACON DONG	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 untuk *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 Akt
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 = banyaknya 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	66 83 41	98 21 35	35 70 52 7	21 30 14 1	30 66 1	46 44 84	27 03 11 6	0	12 95 83
Antapani	60 58 88 1	42 68 23	12 48 64 95	96 31 89 9	21 64 08 3	92 87 76	0	27 03 11 6	16 51 63 7
Arcamanik	32 69 76 8	0	82 96 71 6	60 03 94 2	66 96 68	97 25 1	42 68 23	98 21 35	39 98 70
Astana Anyar	0	32 69 76 8	11 49 51 4	41 22 68	98 29 16	22 45 21 9	60 58 88 1	66 83 41	13 85 44 2
Babakan Ciparay	17 71 9	28 06 54 3	14 52 39 5	60 28 09	73 73 15	18 64 15 0	54 21 84 4	46 86 50	10 90 42 5
Bandung Kidul	64 11 8	24 18 44 4	17 56 41 8	80 13 36	54 50 22	15 51 23 5	48 76 86 5	31 85 45	85 34 76
Bandung Kulon	57 00 27	11 09 71 1	33 38 26 3	19 51 72 9	58 04 7	55 32 02	29 12 77 2	46 80	17 97 25
Bandung Wetan	11 49 51 4	82 96 71 6	0	18 50 58 8	42 56 88 8	66 07 04 7	12 48 64 95	35 70 52 7	50 58 04 6
Batununggal	13 85 44 2	39 98 70	50 58 04 6	33 08 26 6	34 65 0	10 63 91	16 51 63 7	12 95 83	0
Bojonegara Kale	76 65	29 66 36 9	13 41 59 3	53 01 49	82 00 19	19 96 16 0	56 43 45 0	53 56 26	11 90 38 1
Bojonegara Kidul	11 92 3	28 87 77 7	13 94 99 7	56 41 59	77 94 61	19 30 38 8	55 34 56 2	50 23 02	11 41 46 7
Buar Batu	31 66 44	94 5 57 9	81 31 53 7	58 63 53 7	62 27 09	80 45 6	46 52 80	92 56 28	36 37 91
Cibeunying Kale	34 95	30 59 74 1	12 79 60 9	49 15 33	86 92 65	20 71 98 2	57 71 74 2	57 52 92	12 49 84 3
Cibeunying Kidul	98 29 16	66 96 68	42 56 88 8	26 67 05 2	0	26 06 29 3	21 64 08 3	30 66 1	34 65 0
Cibiru	27 00	34 45 59 2	10 49 54 4	35 31 84	10 79 73 0	23 92 60 3	62 97 51 7	74 93 13	14 99 96 4

Cicendo	1022083	637359	4338451	273205	1263	238756	210560	37580	28649
Cidadap	412268	600394	185058	0	266705	458156	963189	213014	330826
Cinambo	1533258	928033	28050	355460	496848	748921	136871	422545	583120
Goblong	5063592	197396	110371	836504	158752	566117	46301	205305	115445
Gedebage	1320017	471440	503203	77955	183359	346569	797792	139372	237097
Kiaradong	2734984	24590	743025	527019	439708	27151	653505	699787	227558
Lengekong	2245219	97251	660704	458156	260629	0	928776	464484	106391
Mandalajati	8762	2942	1357	5403	8072	1975	5609	5251	1175
Panyileukan	7065	2986	1328	5219	8297	2013	5671	5441	1202
Rancasari	2413886	65320	689463	482073	317058	5401	824839	542217	142148
Regol	1046025	617989	438806	277124	1511	227410	207082	42132	24305
Sukasadi	2256531	94479	662669	459778	264617	176	921208	469756	108723
Sukasari	1613722	290808	548635	365713	81620	52109	142050	206147	12082
Sumur Bandung	394786	593659	197150	226	2623	4522	9546	2090	325926
Ujung Berung	391893	139791	288362	160801	135129	761362	336933	37182	304835

4. Pengelompokan data

Pada tahap ini peneliti melakukan pengelompokan data berdasarkan nilai minimum jarak terdekat data pada *centroid*. Berikut merupakan hasil

pengelompokan 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 <i>Centroid</i> Ke <i>Cluster</i>								
	C0	C1	C2	C3	C4	C5	C6	C7	C8
Andir	53 66 13, 9	86 45 49, 3	38 90 98 2	18 54 83 0	36 45 3,6 7	49 10 14, 1	23 66 51 5	74 53, 11 1	16 48 43
Antapani	56 47 57 0	51 05 68, 7	13 07 98 05	90 35 53 6	21 13 17 0	89 04 66, 1	11 57 8 3	29 88 57 3	15 33 04 9
Arcamanik	29 69 46 3	38 77, 66 7	87 34 51 6	55 13 06 4	64 35 34, 7	84 54 5,7 8	30 05 37 1	11 56 75 1	34 23 18
Astana Anyar	72 88, 21	30 52 43 0	13 64 37 5	29 16 32, 8	10 16 67 3	23 04 07 6	55 49 66 4	53 69 19, 4	14 96 56 4
Babakan Ciparay	24 62, 76 5	26 05 56 9	16 59 43 9	45 90 14, 4	76 63 21 6	19 17 93 6	49 40 64 8	35 97 03, 1	11 88 85 5
Bandung Kidul	28 27 4,6 5	22 31 92 1	19 83 14 4	63 61 21, 4	57 02 94, 7	15 99 88 5	44 21 27 9	23 01 32, 1	94 13 64
Bandung Kulon	44 89 78, 3	98 49 96, 7	36 48 47 5	16 88 41 7	65 60 1,6 7	58 23 94, 8	25 63 58 0	60 9,7 7, 8	21 98 39
Bandung Wetan	13 39 17 0	79 48 21 7	70 13, 5	27 85 50, 1	43 27 46 8	67 08 32 6	11 75 02 47	32 57 63 9	52 69 18 2
Batununggal	11 92 30 0	32 57 69, 7	54 37 61 2	29 62 91 5	28 86 6,6 7	11 79 52, 4	13 91 47 3	19 82 13, 4	30 24
Bojonglora Kaler	10 0,3 21	27 59 40 1	15 40 37 6	39 75 27, 1	85 12 68, 3	20 51 09 1	51 53 00 3	41 86 53, 8	12 94 08 5
Bojonglora Kidul	77 7,6 54 3	26 83 92 9	15 98 08 5	42 70 16, 4	80 92 25 2	19 85 18 2	50 48 32 2	38 91 61, 1	12 41 98 9
Buar Batu	28 71 00 1	10 74, 66 7	86 11 51 6	54 00 36 4	59 55 67 4	68 76 3,4 4	33 27 01 0	10 95 72 0	30 93 75
Cibeunying Kaler	70 2,3 21	28 49 56 7	14 74 20 1	36 41 85, 1	90 10 89, 7	21 28 40 7	52 75 07 0	45 39 13, 8	13 55 62 1
Cibeunying Kidul	82 14 36, 2	57 27 26, 3	46 05 67 6	23 58 04 9	59 1,3 33 3	27 96 34, 4	18 64 23 3	68 11 1,4 4	55 11 8
Cibiru	17 83	32 22	12 25	24 67	11 15	24 52	57 78	60 99	16 16

	9,7 7	05 4	99 9	60, 8	80 2	64 4	66 6	39, 4	60 0
Cicendo	85 76 60, 3	54 33 84, 7	46 91 18 9	24 19 07 5	17 1,6 66 7	25 80 14, 8	18 09 28 0	78 06 5,7 8	45 66 3
Cidadap	52 86 55, 3	57 07 92 5	26 32 46, 5	95 44, 77 8	27 23 11 3	46 65 56 0	89 86 90 0	18 30 12 7	34 79 68 2
Cinambo	17 50 80 2	89 11 15 9	70 13, 5 4	48 14 07, 4	80 52 5 5	75 96 24 3	12 61 63 55	38 84 65 6	60 59 00 4
Goblong	46 88 69 0	25 52 27 20	11 59 59 8	78 10 13 8	15 43 25 8	53 71 05, 1	11 57 8 5	23 03 43 5	10 55 01 8
Gedebage	20 07 95, 9	44 52 40 9	62 74 86, 5	33 02 9,7 8	18 80 33 8	35 38 35 3	73 92 28 0	12 00 95 6	25 16 76 3
Kiaradondong	24 60 68 2	89 58, 33 3	78 88 90 5	48 31 67 5	41 75 26 5	19 79 0,4 4	49 46 40 31	84 88 4, 1	18 59 24
Lengekong	19 98 13 0	63 65 12 2	70 41 17 8	41 73 31, 7	24 19 31, 4	72 5,4 44 4	73 58 74 15,	58 65 15, 1	76 23 3
Manadajati	85, 98 76 5	27 36 08 5	15 57 78 1	40 63 28, 8	83 81 26, 7	20 30 76 0	51 20 77 6	40 94 45, 4	12 77 94 0
Panyileukan	35 2,7 65 4	27 78 62 4	15 26 23 0	39 05 91, 1	86 16 52, 3	20 68 17 9	51 79 42 9	42 65 97, 8	13 07 48 5
Rancasari	21 56 69 6	37 55 1 9	73 36 82 4	44 01 72 4	29 81 00 8	22 29, 77 8	64 46 08 8	67 62 98, 8	10 95 22
Regol	87 94 39, 4	52 54 19, 3	47 42 59 2	24 55 92 3	0,3 33 3 3	24 58 92, 8	17 77 41 1	84 52 3,7 8	40 51 3
Sukaajadi	20 08 68 1	61 40 0 0	70 61 27 0	41 88 59 7	24 59 23, 7	48 2,7 77 8	72 95 81 8	59 21 65, 8	78 28 7
Sukasari	14 05 43 6	22 94 08, 3	58 82 58 6	32 93 24 3	70 70 0 4	61 83 4,4 4	11 79 48 4	28 93 38, 1	30 24
Sumur Bandung	50 90 11, 8	56 42 62 2	27 79 23, 5	71 90, 11 1	26 78 35 1	46 05 94 0	89 04 36 9	18 52 32 5	34 28 49 6
Ujung Berung	29 27 84, 1	12 57 35 7	31 72 45 5	13 69 90 9	14 71 39 9	14 79 56 8	29 92 60, 2	11 4,7 3,7 8	35 70 85

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

merupakan hasil pengelompokan 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 pengelompokan 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.

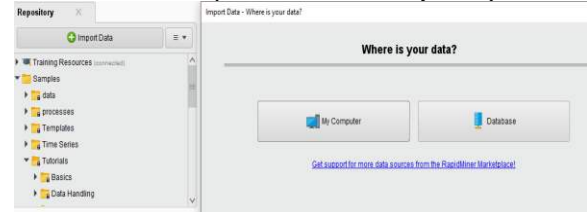
C. Implementasi Algoritma K-Means Menggunakan RapidMiner

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

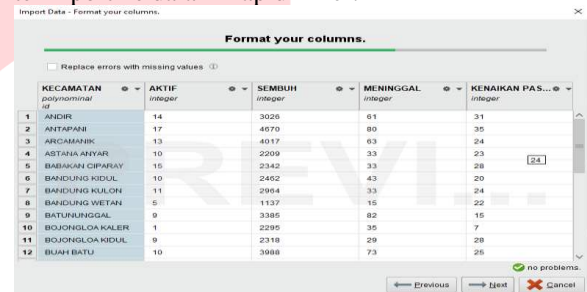
Sebelum menerapkan algoritma k-means data dengan format *.xls di import terlebih dahulu ke dalam RapidMiner dengan cara klik "*import data*",

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

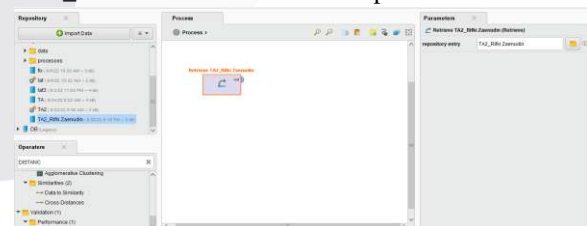
Selanjutnya pilih data yang ingin di *import* ke dalam RapidMiner. Apabila data yang dipilih terindikasi bahwa tidak ada masalah maka akan muncul dibagian bawah kanan "*no problems*" pada proses *import* data. Lalu lanjutkan hingga *finish*, sehingga data berhasil terimport ke dalam RapidMiner.



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 pengelompokan 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 :

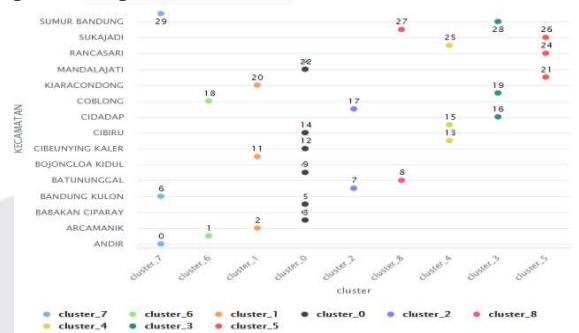
TABEL 4.11
RANKING

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 data ke 2, 11, dan 20. Cluster 2 memiliki 2 data yang terdapat pada data ke 7, dan 17. Cluster 3 memiliki 3 data yang terdapat pada data ke 16, 19, dan 28. Cluster 4 memiliki 3 data yang terdapat pada data ke 13, 15, dan 25. Cluster 5 memiliki 3 data yang terdapat pada data ke 21, 24, dan 26. Cluster 6 memiliki 2 data yang terdapat pada data ke 1, dan 18. Cluster 7 memiliki 3 data yang terdapat 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 dengan peneliti mencari nilai *performance* dengan bantuan *tools* RapidMiner. Selanjutnya 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 berbeda misalnya fuzzy c-means atau k-medoids.
2. Penelitian selanjutnya dalam pengujian algoritma k-means dapat menggunakan bantuan *tools* lainnya

seperti *python* dan *matlab* agar pembahasannya 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 Pengelompokan 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 Pengelompokan 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 Pengelompokan 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.