

ANALISIS DATA MINING UNTUK *CLUSTERING* KASUS COVID-19 DI PROVINSI LAMPUNG DENGAN ALGORITMA K-MEANS

Zulfa Nabila¹, Auliya Rahman Isnain², Permata³, Zaenal Abidin⁴

Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia^{1,2,4}

Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia³

zulfa_nabila@teknokrat.ac.id¹, aulyarahman@teknokrat.ac.id², permata@teknokrat.ac.id³

Received: (7 Juni 2021) Accepted: (14 Juni 2021) Published: (28 Juni 2021)

Abstract

Coronavirus Diseases 2019 or better known as Covid-19 is a case of pneumonia of unknown cause which was first discovered at the end of December 2019 in Wuhan City, China. When infecting humans, this virus causes respiratory infections such as the flu. The purpose of this study was to analyze data on Covid-19 cases to find out the grouping of Covid-19 case problems in Lampung Province. The grouping of data on Covid-19 cases in Lampung Province was carried out using the Clustering method with the K-Means algorithm with the attributes of Regency/City, Suspected, Probable, Positive Confirmation, Completed Isolation, and Death used in the calculation process and divided the data into 4 clusters categorized as Red Zone, Orange Zone, Yellow Zone, and Green Zone. And validation using Davies-Bouldin Index (DBI). The results of DBI validation using manual calculations and using the help of RapidMiner tools have differences, in this case, manual calculations have better results than using RapidMiner tools, but the results of both calculations are both close to 0 which means the evaluated clusters produce good clusters.

Keywords: Covid-19, Clustering, k-Means, Davies-Bouldin Index (DBI)

Abstrak

Coronavirus Diseases 2019 atau yang lebih dikenal dengan sebutan Covid19 adalah kasus pneumonia yang tidak diketahui penyebabnya yang ditemukan pertama kali pada akhir bulan Desember 2019 di Kota Wuhan, China. Ketika menyerang manusia, virus ini menyebabkan infeksi saluran pernafasan seperti flu. Tujuan dari penelitian ini adalah untuk menganalisis data kasus Covid-19 agar dapat mengetahui pengelompokan pada masalah kasus Covid-19 di Provinsi Lampung. Pengelompokan data kasus Covid-19 di Provinsi Lampung dilakukan menggunakan metode Clustering dengan algoritma K-Means dengan atribut Kabupaten/Kota, Suspek, Probable, Konfirmasi Positif, Selesai Isolasi, dan Kematian yang digunakan dalam proses perhitungan dan membagi data ke dalam 4 cluster yang dikategorikan sebagai Zona Merah, Zona Orange, Zona Kuning dan Zona Hijau. Serta validasi menggunakan Davies-Bouldin Index (DBI). Hasil validasi DBI menggunakan perhitungan secara manual dan menggunakan bantuan tools RapidMiner memiliki perbedaan, pada kasus ini perhitungan manual memiliki hasil yang lebih baik dibandingkan dengan menggunakan tools RapidMiner, tetapi hasil dari kedua perhitungan sama-sama mendekati 0 yang berarti kluster yang dievaluasi menghasilkan kluster yang baik.

Kata Kunci: Covid-19, Data Mining, Clustering, K-Means, Davies-Bouldin Index (DBI)

To cite this article:

Nabila, Isnain, Permata, Abidin (2021). Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means, Vol (2), No. 2, 100 – 108 .

1. Pendahuluan

Pada 31 Desember 2019, WHO China Country Office melaporkan kasus pneumonia yang tidak diketahui penyebabnya di Kota Wuhan, Provinsi Hubei, China. Pada tanggal 7 Januari 2020, China mengidentifikasi

pneumonia yang tidak diketahui penyebabnya tersebut sebagai jenis baru coronavirus (novel coronavirus) [1]. Coronavirus disebut dengan virus zoonotik yaitu virus yang ditransmisikan dari hewan ke manusia. Berdasarkan penelitian dari 59 kasus yang diduga, 41 pasien di Wuhan dipastikan terinfeksi coronavirus (2019-nCoV) [2].

World Health Organization (WHO) memberi nama virus baru tersebut Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) dan resmi menetapkan penyakit tersebut dengan nama Novel Coronavirus pada manusia dengan sebutan Coronavirus Diseases 2019 atau yang lebih dikenal dengan sebutan Covid-19 [1]. Ketika menyerang manusia, virus ini menyebabkan infeksi saluran pernafasan seperti flu, hingga penyakit yang lebih berat seperti Middle East Respiratory Syndrome (MERS) dan Severe Acute Respiratory Syndrome (SARS) [3]. Sampai saat ini sudah dipastikan 222 Negara yang telah terjangkit, dengan kasus konfirmasi 127.877.462 dengan jumlah kematian 2.796.561 [4].

Setiap hari kasus penyebaran Covid-19 di Indonesia terus meningkat, masyarakat diminta melakukan *social distancing* untuk memutus mata rantai penyebaran Covid-19 yang tersebar di berbagai wilayah di Indonesia [5]. Covid-19 pertama dilaporkan masuk ke Indonesia pada 2 Maret 2020, dengan 2 kasus terkonfirmasi. Wilayah Indonesia yang sangat luas memungkinkan perlunya pengelompokan bagian-bagian berdasarkan wilayah di Indonesia [6]. Menurut data dari Kementerian Kesehatan Republik Indonesia per tanggal 31 Maret 2021 jumlah yang terkonfirmasi positif Covid-19 yaitu 1.511.712 dengan jumlah kematian 40.858 dan sembuh 1.348.330. Negara Indonesia merupakan Negara dengan kasus konfirmasi tertinggi di ASEAN [7]. Sedangkan, kasus Covid-19 di wilayah Provinsi Lampung pertama dilaporkan masuk pada tanggal 18 Maret 2020 dengan jumlah terkonfirmasi positif sebanyak 1 kasus. Di Provinsi Lampung sendiri jumlah yang terkonfirmasi positif Covid-19 per tanggal 31 Maret 2021 yaitu 14.019 dengan jumlah kematian 759 dan sembuh 12.749 [8].

Dinas Kesehatan Provinsi Lampung dan Badan Perencanaan Pembangunan Daerah Provinsi Lampung sebelumnya sudah merincikan lima kriteria zonasi berdasarkan warna. Namun belum adanya perhitungan secara umum dan secara detail hanya berdasarkan pengumpulan data dan kajian maupun analisis dari tim pakar Gugus Tugas Nasional penentuan zona menggunakan indikator yang secara total terdapat 15 indikator utama, diantaranya adalah indikator kesehatan masyarakat, yang terbagi menjadi 11 indikator epidemiologi, 2 indikator surveilans kesehatan masyarakat dan 2 indikator pelayanan kesehatan [9].

Pada penelitian ini penulis ingin menerapkan data mining pada data kasus Covid-19 di Provinsi Lampung menggunakan algoritma K-Means Clustering. Data mining adalah proses menemukan pola dan pengetahuan yang menarik dari suatu data dalam jumlah yang besar. Data mining menggambarkan sebuah pengumpulan teknik-teknik dengan tujuan untuk menemukan pola-pola yang tidak diketahui pada data yang telah dikumpulkan [10]. Algoritma K-Means merupakan salah satu algoritma Clustering yang masuk ke dalam kelompok Unsupervised learning yang digunakan untuk mengelompokkan data ke dalam beberapa kelompok dengan sistem partisi [6] [32].

2. Tinjauan Pustaka

2.1 Data Mining

Data mining adalah proses untuk menemukan korelasi, pola, dan tren baru yang bermakna dengan memilah-milah data dalam jumlah besar yang disimpan di dalam repositori, menggunakan teknologi pengenalan pola serta teknik statistik dan matematika [11]. Data mining adalah proses menemukan pola dan pengetahuan menarik dari data dalam jumlah yang besar [10].

Istilah data mining memiliki beberapa pandangan, seperti *knowledge discovery* ataupun *pattern recognition*. Istilah *knowledge discovery* atau penemuan pengetahuan tepat digunakan karena tujuan utama dari data mining memang untuk mendapatkan pengetahuan yang masih tersembunyi di dalam bongkahan data. Sedangkan istilah untuk *pattern recognition* atau pengenalan pola tepat untuk digunakan karena guna menemukan pola yang tersembunyi di dalam bongkahan data [12].

2.2 Clustering

Clustering mengacu pada pengelompokan seperti record, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Cluster adalah kumpulan dari record yang memiliki kemiripan satu sama lain, dan berbeda dengan record di klaster lain. Clustering mencoba untuk membagi seluruh kumpulan data menjadi kelompok-kelompok yang relatif memiliki kemiripan, di mana kemiripan record dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan record dalam kelompok lain akan bernilai minimal [11].

Clustering dalam data mining berguna untuk menemukan pola distribusi di dalam sebuah data set yang berguna untuk proses analisa data. Kesamaan objek biasanya diperoleh dari kedekatan nilai-nilai atribut yang menjelaskan objek-objek data, sedangkan objek-objek data biasanya direpresentasikan sebagai sebuah titik dalam ruang multidimensi [13].

2.3 Covid-19

Coronavirus merupakan keluarga besar virus yang menyebabkan penyakit pada manusia dan hewan. Pada manusia biasanya menyebabkan penyakit infeksi saluran pernapasan, mulai flu biasa hingga penyakit yang serius seperti *Middle East Respiratory Syndrome* (MERS) dan Sindrom Pernafasan Akut Berat/*Severe Acute Respiratory Syndrome* (SARS) [14]. Coronavirus jenis baru yang ditemukan pada manusia sejak kejadian luar biasa muncul di Wuhan Cina, pada Desember 2019, kemudian diberi nama *Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-COV2), dan menyebabkan penyakit Coronavirus Disease-2019 (COVID-19) [1].

2.4 Algoritma K-Means

Algoritma K-Means merupakan salah satu algoritma Clustering yang masuk ke dalam kelompok Unsupervised learning yang digunakan untuk mengelompokkan data ke dalam beberapa kelompok dengan sistem partisi [15]. Unsupervised learning

merupakan algoritma data mining untuk mencari pola dari semua variabel (atribut), variabel (atribut) yang menjadi target/label/class tidak ditentukan (tidak ada) [10] [25] [30]. Algoritma K-Means merupakan teknik Clustering berbasis jarak yang membagi data ke dalam beberapa cluster dan algoritma ini hanya bekerja pada atribut angka atau numerik.

Langkah-langkah algoritma K-Means adalah sebagai berikut [15] [21] [23]:

1. Tentukan nilai k atau jumlah Cluster pada data set
2. Menentukan nilai pusat (centroid). Penentuan nilai centroid pada tahap awal dilakukan secara random, sedangkan pada tahap iterasi digunakan rumus seperti dibawah ini:

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} Xkj \dots \dots \dots (1)$$

Keterangan:

- V_{ij} = Centroid rata-rata cluster ke-I untuk variabel ke j
- N_i = Jumlah anggota cluster ke-i
- i,k = Indeks dari cluster
- j = Indeks dari variabel
- X_{kj} = nilai data ke-k variabel ke-j untuk cluster tersebut

3. Menghitung jarak antara titik centroid dengan titik tiap objek menggunakan Euclidean Distance. Euclidean Distance merupakan jarak garis lurus biasa antara dua titik dalam ruang Euclidean, dengan rumus seperti dibawah ini:

$$De = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \dots \dots \dots (2)$$

Keterangan:

- De = Euclidean Distance
- i = Banyaknya objek
- (x, y) = Koordinat objek
- (s, t) = Koordinat centroid

4. Kelompokkan objek berdasarkan jarak ke Centroid terdekat 5. Ulangi langkah ke-2 hingga ke-4, lakukan iterasi hingga centroid bernilai optimal.

2.5 RapidMiner

RapidMiner merupakan perangkat lunak yang bersifat terbuka (*open source*). RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi. RapidMiner merupakan *software* yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri. RapidMiner ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi. RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik [16].

2.6 Evaluasi Clustering

Metode yang digunakan dalam menentukan evaluasi cluster ini menggunakan *Davies-Bouldin Index* (DBI). [22] [28] *Davies-bouldin index* merupakan salah satu metode evaluasi internal yang mengukur evaluasi cluster pada suatu metode pengelompokan yang didasarkan pada nilai kohesi dan separasi. 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:

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_i, c_i) \dots \dots \dots (3)$$

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:

$$SSB_{ij} = d(c_i, c_j) \dots \dots \dots (4)$$

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 kei 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:

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}} \dots \dots \dots (5)$$

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

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{ij}) \dots \dots \dots (6)$$

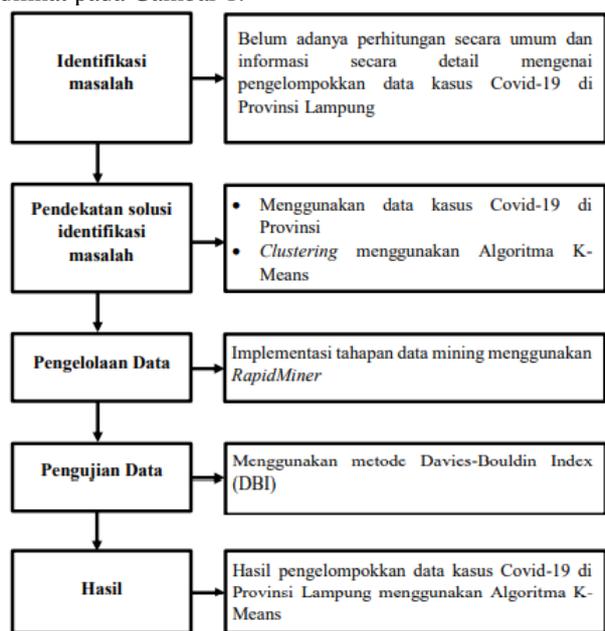
Dari persamaan tersebut, k merupakan jumlah cluster yang digunakan. Semakin kecil nilai DBI yang diperoleh (non-negatif >= 0), maka semakin baik cluster yang diperoleh dari pengelompokan K-means yang digunakan [13] [26] [29].

3. Metode Penelitian

3.1 Kerangka Penelitian

Kerangka penelitian atau suatu rancangan alur sebuah penelitian dibuat oleh penulis disampaikan melalui gambar yang berurutan sesuai dengan tahapan apa saja

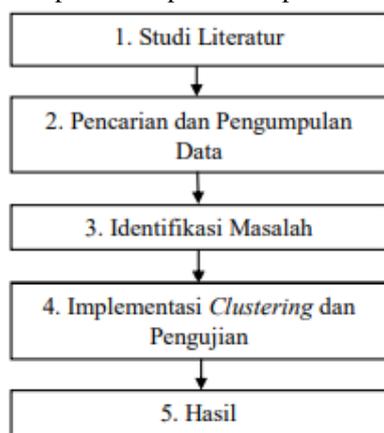
yang akan dilakukan dalam melakukan penelitian dapat dilihat pada Gambar 1.



Gambar 1. Kerangka Penelitian

3.2 Tahapan Penelitian

Tahapan penelitian merupakan langkah-langkah yang dilakukan oleh peneliti, berikut adalah tahapan yang dilakukan oleh peneliti dapat dilihat pada Gambar 2:



Gambar 2. Tahapan Penelitian

4. Hasil dan Pembahasan

4.1 Pengolahan Data dan Informasi

Data diperoleh dari Badan Penanggulangan Bencana Daerah (BPBD) Provinsi Lampung yang bersumber dari Dinas Kesehatan Provinsi Lampung. Data tersebut diminta langsung oleh peneliti melalui Kepala Divisi Darurat dan Logistik. Data yang digunakan dalam penelitian ini adalah data dari 15 kabupaten/kota di Provinsi Lampung pada akhir bulan Maret 2020 sampai Maret 2021 [17] [31].

4.2 Perhitungan dengan Algoritma K-Means

Saat melakukan perhitungan manual, sampel data kasus Covid-19 sebanyak 15 kabupaten/kota di Provinsi

Lampung. Dari semua data yang diperoleh peneliti menggunakan 6 atribut, hal ini dilakukan dengan mempertimbangkan kebutuhan kasus. Berikut contoh data kasus Covid-19 di Provinsi Lampung per 31 Maret 2021:

Tabel 1. Data Kasus Covid-19

Kab/Kota	Suspek	Probable	Konfirmasi Positif	Selesai Isolasi	Kematian
Bandar Lampung	12	2	5109	4627	325
Lampung Selatan	29	2	853	781	45
Lampung Timur	19	1	1051	951	57
Lampung Utara	25	13	1134	1051	31
Metro	0	0	700	628	37
Tulang Bawang	0	0	199	182	13
Tulang Bawang Barat	0	0	197	175	18
Mesuji	0	0	100	96	4
Tanggamus	0	0	551	520	27
Pringsewu	0	16	465	400	27
Pesawaran	30	10	589	537	30
Pesisir Barat	25	1	267	242	8
Lampung Barat	44	2	474	431	15
Lampung Tengah	28	38	2217	2020	117
Way Kanan	0	0	113	108	5

Proses perhitungan data menggunakan algoritma K-Means dilakukan dengan menggunakan Microsoft Excel. Berikut langkah-langkah penyelesaiannya:

4.2.1 Menentukan Jumlah Cluster

Jumlah cluster merupakan jumlah kelompok yang akan dihasilkan. Jumlah ini menyesuaikan dengan kebutuhan analisis dalam penelitian ini jumlah cluster yang akan digunakan sebanyak 4 cluster.

4.2.2 Menentukan Centroid Awal

Centroid awal merupakan titik pusat cluster pertama yang diperoleh secara acak dari data training. Pemilihan titik centroid awal akan sangat berpengaruh terhadap hasil cluster. Berikut merupakan centroid awal dari penelitian ini:

Tabel 2. Centroid Awal

Kab/Kota	Suspek	Probable	Konfirmasi Positif	Selesai Isolasi	Kematian
C0	4	7	13	552	425
C1	1	86	1	2786	1971
C2	14	13	25	920	744
C3	8	0	4	45	39

4.2.3 Menghitung Jarak Data ke Titik Pusat Cluster

Langkah selanjutnya adalah menghitung jarak setiap data dengan centroid awal menggunakan persamaan *Euclidean Distance*. Pada tahap ini, jarak terdekat antara data dan cluster akan menentukan data mana yang termasuk dalam kelompok data mana. Berikut adalah hasil perhitungan jarak pada iterasi ke-1.

Tabel 3. Hasil Iterasi ke-1

Kab/Kota	Jarak ke Centroid				Kelompok Data
	C0	C1	C2	C3	
Bandar Lampung	0	6540,1	3899,3	5481,9	C0
Lampung Selatan	5743,1	797,12	1844,5	261,43	C3
Lampung Timur	5481,9	1058,2	1583,5	0	C3
Lampung Utara	5354,9	1186,1	1456,0	133,21	C3
Metro	5959,3	581,37	2060,9	477,80	C3
Tulang Bawang	6630,5	94,398	2731,9	1148,7	C1
Tulang Bawang Barat	6636,4	100,69	2737,9	1154,7	C1
Mesuji	6761,8	223,30	2863,3	1280,0	C1
Tanggamus	6142,6	398,66	2244,0	661,10	C1
Pringsewu	6286,7	255,70	2388,1	805,31	C1
Pesawaran	6102,9	437,37	2204,0	621,15	C1
Pesisir Barat	6540,1	0	2641,4	1058,2	C1
Lampung Barat	6259,9	281,03	2361,1	778,34	C1
Lampung Tengah	3899,3	2641,4	0	1583,5	C2
Way Kanan	6744,1	205,70	2845,5	1262,4	C1

4.2.4 Menghitung Titik Pusat Centroid Baru

Pada iterasi pertama data telah menghasilkan keanggotaan cluster, selanjutnya penulis melakukan perhitungan untuk menentukan nilai titik pusat cluster baru yang diperoleh dari menghitung rata-rata keanggotaan dari cluster 0, cluster 1, cluster 2 dan cluster 3, berikut merupakan nilai centroid baru:

Tabel 4. Titik Pusat Cluster Iterasi ke-2

C0	1	12	2	5109	4627	325
C1	10,111	11	3,2222	328,33	299	16,333
C2	14	28	38	2217	2020	117
C3	3,5	18,25	4	934,5	852,75	42,5

Proses penghitungan nilai centroid baru telah dilakukan, langkah selanjutnya adalah menghitung ulang data dengan nilai centroid baru. Dalam studi kasus ini, penulis melakukan perhitungan hingga iterasi ke-2. Pada iterasi 1 dan 2 data tidak merubah keanggotaan cluster

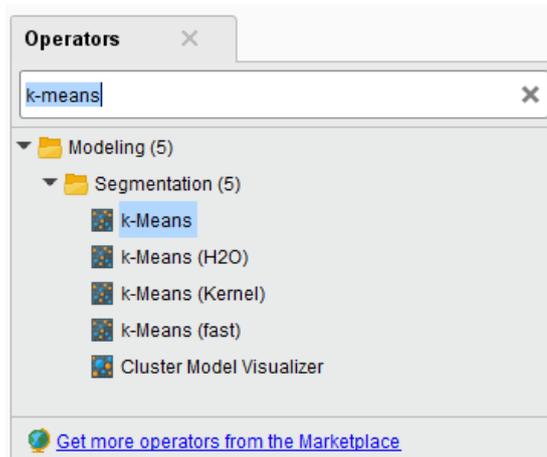
sehingga iterasi dihentikan, berikut tabel perhitungan untuk iterasi ke-2:

Tabel 5. Hasil Iterasi ke-2

Kab/Kota	Jarak ke Centroid				Kelompok Data
	C0	C1	C2	C3	
Bandar Lampung	0	6456,1	3899,3	5634,8	C0
Lampung Selatan	5743,1	713,31	1844,5	109,17	C3
Lampung Timur	5481,9	974,23	1583,5	153,11	C3
Lampung Utara	5354,9	1102,3	1456,0	281,71	C3
Metro	5959,3	496,95	2060,9	325,39	C3
Tulang Bawang	6630,5	174,85	2731,9	996,03	C1
Tulang Bawang Barat	6636,4	181,02	2737,9	1002,1	C1
Mesuji	6761,8	305,99	2863,3	1127,3	C1
Tanggamus	6142,6	314,11	2244,0	508,34	C1
Pringsewu	6286,7	171,10	2388,1	652,81	C1
Pesawaran	6102,9	353,81	2204,0	468,46	C1
Pesisir Barat	6540,1	85,350	2641,4	905,47	C1
Lampung Barat	6259,9	199,35	2361,1	625,65	C1
Lampung Tengah	3899,3	2557,4	0	1736,1	C2
Way Kanan	6744,1	288,32	2845,5	1109,6	C1

4.3 Implementasi Algoritma K-Means Menggunakan RapidMiner

Penerapan Algoritma K-Means menggunakan RapidMiner bertujuan untuk mengelompokkan data kasus Covid-19 di Provinsi Lampung. Terdapat empat pilihan K-Means pada RapidMiner yaitu k-Means, k-Means (H2O), kMeans (Kernel), k-Means (fast) seperti pada gambar dibawah ini.



Gambar 3. Pilihan K-Means

Untuk menghasilkan k-Means dari dataset yang telah disiapkan, dataset terlebih dahulu di import ke dalam RapidMiner. Selanjutnya pembuatan lembar proses, dimana didalam lembar proses ini selanjutnya berisi proses penerapan algoritma maupun proses pengujian menggunakan metode-metode pengujian yang ada pada operators RapidMiner. Langkah-langkah tersebut dapat dilihat pada penjelasan-penjelasan dibawah ini.

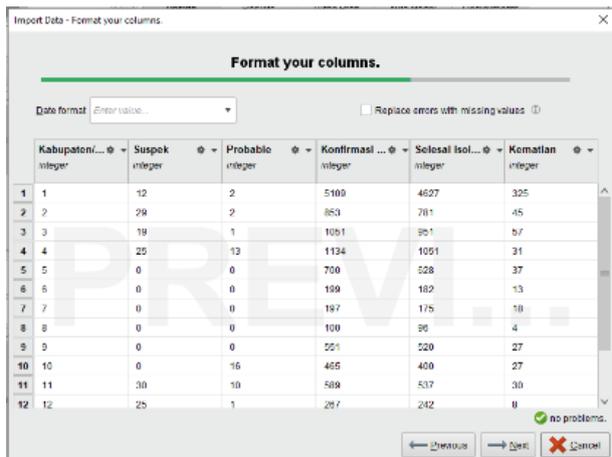
4.3.1 Import Data

Sebelum menerapkan algoritma k-Means maka data dengan format *.xls di import terlebih dahulu ke dalam RapidMiner dengan klik “Add Data”. Selanjutnya pilih lokasi penyimpanan dataset yang akan digunakan. Karena dataset penulis berbentuk file yang memiliki format *.xls dan bukan berasal dari database seperti sql maka pilih lokasi dari “My Computer”.



Gambar 4. Import Data

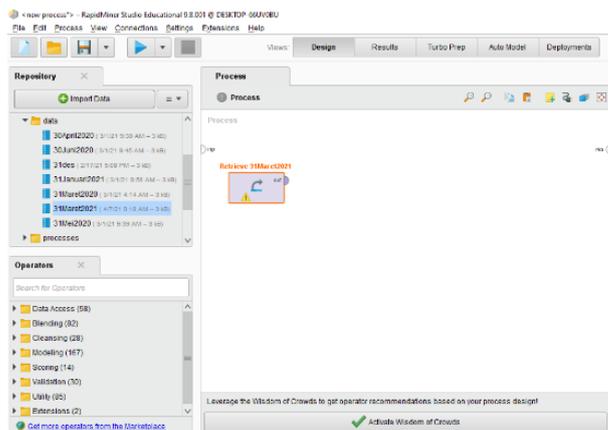
Selanjutnya pilih data import ke dalam RapidMiner. Apabila data yang dipilih telah terindikasi tidak ada masalah maka akan muncul “no problems” pada proses import data. Lalu lanjutkan hingga finish, sehingga data berhasil di import ke dalam RapidMiner.



Gambar 5. Proses Import Data

4.3.2 Membuat Lembar Proses

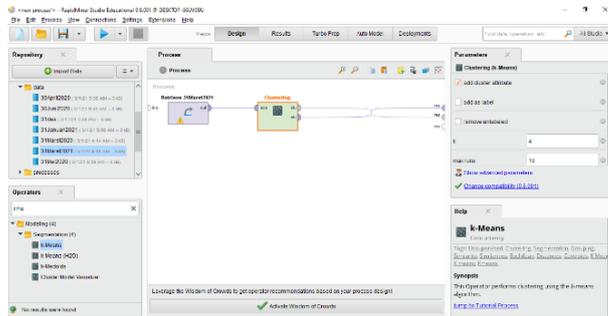
Setelah data diimport, data dimasukkan ke halaman Proses. Dengan menyeret data yang diimport. Contoh file data dalam penelitian yang digunakan disini adalah 31 Maret 2021. Kemudian Drag 31 Maret 2021 ke halaman proses. Gambar 6 di bawah ini merupakan implementasi dari RapidMiner.



Gambar 6. Lembar Proses

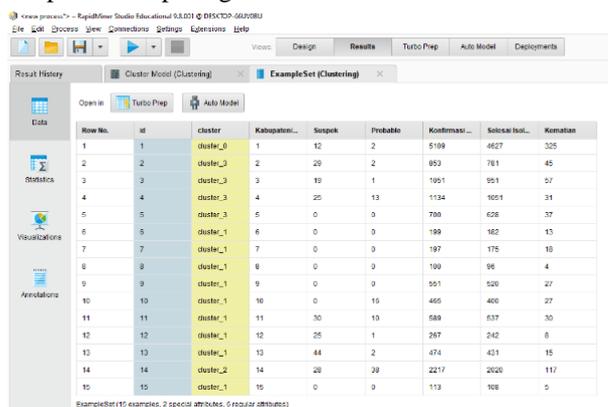
4.3.3 Menerapkan Algoritma K-Means

Langkah selanjutnya adalah memasukkan Algoritma k-Means pada proses yang telah terbentuk diatas untuk mengetahui pengelompokkan dari proses yang telah dibuat. Cari “k-Means” pada Operator lalu tarik ke sebelah “data” lalu hubungkan ke “Clustering” dan “res”. Gambar 7 dibawah ini merupakan implementasi pada Rapid Miner.



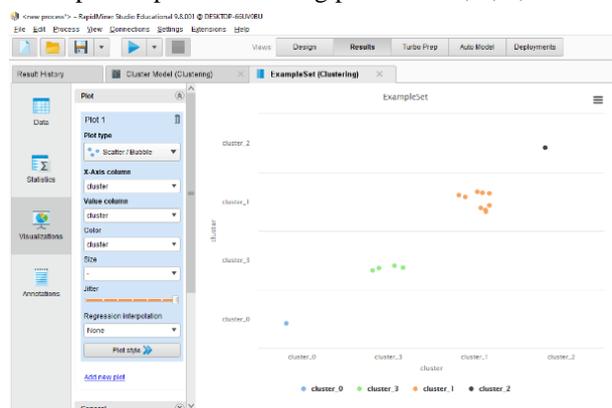
Gambar 7. Menerapkan K-Means

Masukan nilai k yang diinginkan, disini peneliti memasukan nilai 4 karena pada kasus ini peneliti mengelompokkannya menjadi 4 cluster. Setelah itu lakukan running pada proses untuk mendapatkan hasil clustering dari Algoritma k-Means yang diterapkan pada RapidMiner. Hasil Clustering dari Algoritma k-Means dapat dilihat pada gambar 8.



Gambar 8. Hasil Clustering

Proses run akan berjalan dengan baik jika muncul tampilan data pada tampilan Data View yang berhasil dijalankan seperti gambar diatas. Pada bagian visualisasi, kita dapat melihat garis dimana ditentukan jumlah clustering group di setiap kabupaten/kota di Provinsi Lampung. Untuk visualisasi di sisi kiri untuk sumbu x, sumbu y, dan warna custom pilih (cluster) semuanya akan terlihat pada gambar di bawah ini. Dalam visualisasi tersebut, kita dapat mengetahui bagian grup mana yang mendapatkan posisi clustering pada level 0, 1, 2, dan 3.



Gambar 9. Visualisasi per Tanggal 31 Maret 2021

Pada gambar diatas merupakan tampilan akhir dari aplikasi k-Means untuk menentukan pengelompokan data clustering. Dari pengujian di atas dapat disimpulkan bahwa Cluster 0 berjumlah 1 dilambangkan dengan warna biru, cluster 1 berjumlah 9 dilambangkan dengan orange, cluster 2 berjumlah 1 dilambangkan dengan warna hitam, dan cluster 3 berjumlah 4 dilambangkan dengan warna hijau. [18][19].

Setelah mengimplementasikan alat *RapidMiner* hasil yang didapat sama dengan hasil perhitungan manual. Untuk hasil klaster Kasus Covid-19 akhir bulan Maret 2020 sampai Maret 2021 di Provinsi Lampung dapat dilihat pada Tabel 6.

Tabel 6. Cluster Kasus Covid-19

Per Bulan	C0 (Zona Merah)	C1 (Zona Orange)	C2 (Zona Kuning)	C3 (Zona Hijau)
31 Maret 2020	2	6	3	4
30 April 2020	1	1	11	2
31 Mei 2020	12	1	1	1
30 Juni 2020	11	1	1	2
31 Juli 2020	4	1	7	3
31 Agustus 2020	1	3	1	10
30 September 2020	1	10	2	2
31 Oktober 2020	1	11	2	1
30 November 2020	1	9	1	4
31 Desember 2020	3	1	1	10
31 Januari 2021	1	7	1	6
28 Februari 2021	6	1	1	7
31 Maret 2021	1	9	1	4

4.4 Evaluasi Cluster

Pada pengujian mencari nilai *Davies-bouldin index* (DBI) menggunakan perhitungan manual dan mengguna-

kan bantuan *tools RapidMiner* terdapat perbedaan, hasil yang didapat sama-sama mendekati 0 yang artinya semakin kecil nilai DBI yang didapat (non-negatif > = 0), semakin baik clusternya. diperoleh dari pengelompokan K-means yang digunakan. Untuk hasil perhitungan secara manual dan menggunakan bantuan alat *RapidMiner* [24][27] dapat dilihat pada Tabel 7.

Tabel 7. Nilai *Davies-bouldin index* (DBI)

Per Bulan	Manual	RapidMiner
31 Maret 2020	0,3201	0,347
30 April 2020	0,1314	0,364
31 Mei 2020	0,1881	0,195
30 Juni 2020	0,1746	0,433
31 Juli 2020	0,1508	0,398
31 Agustus 2020	0,0924	0,208
30 September 2020	0,1231	0,353
31 Oktober 2020	0,0609	0,186
30 November 2020	0,18	0,357
31 Desember 2020	0,1676	0,413
31 Januari 2021	0,2305	0,296
28 Februari 2021	0,1229	0,301
31 Maret 2021	0,1362	0,313

5. Kesimpulan dan Saran

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan, diperoleh kesimpulan sebagai berikut:

- Berdasarkan hasil pengujian data menggunakan *RapidMiner* pada kasus Covid-19 di Provinsi Lampung per akhir bulan Maret 2020 sampai dengan Maret 2021 dapat dilihat pada Tabel 4.9 menghasilkan pengelompokan yang berbeda dikarenakan jumlah pada atribut Suspek, Probable, Konfirmasi Positif, Selesai Isolasi, dan Kematian pada setiap Kabupaten/Kota yang tidak sama. Untuk Zona Merah terdapat 2 bulan yaitu pada bulan Mei 2020 dan Juni 2020, Zona Orange terdapat 6 bulan yaitu pada bulan Maret 2020, September 2020, Oktober 2020, November 2020, Januari 2021, dan Maret 2021, Zona Kuning terdapat 2 bulan yaitu pada bulan April 2020 dan Juli 2020, Zona Kuning terdapat 3 bulan yaitu pada bulan Agustus 2020, Desember 2020 dan Februari 2021.
- Berdasarkan hasil pengujian mencari nilai *Davies-Bouldin Index* (DBI) menggunakan perhitungan secara manual dan menggunakan bantuan *tools RapidMiner* memiliki perbedaan. Pada kasus ini perhitungan manual memiliki hasil yang lebih baik jika dibandingkan dengan bantuan *tools RapidMiner*, tetapi hasil dari kedua perhitungan per akhir bulan samasama mendekati 0 yang berarti klaster yang dievaluasi menghasilkan klaster yang baik.

5.2 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 Kecamatan yang ada di Kabupaten/Kota dengan data kasus Covid-19 di Provinsi Lampung pada bulan-bulan berikutnya, dengan menggunakan algoritma yang berbeda misalnya *K-Medoids* atau *Fuzzy C-Means*. Uji validasi dapat dilakukan dengan metode lainnya seperti *Silhouette*.
2. Penelitian selanjutnya diharapkan untuk mencari informasi mengenai perbedaan hasil evaluasi DBI yang dilakukan secara perhitungan manual dan menggunakan bantuan *tools RapidMiner*.

Daftar Pustaka

- [1] D. Safrizal, D. I. Putra, S. Sofyan, And Bimo, *Pedoman Umum Menghadapi Pandemi Covid-19 Bagi Pemerintah Daerah*. 2020.
- [2] Yuliana, "Corona Virus Diseases (Covid-19); Sebuah Tinjauan Literatur," Vol. 2, No. February, Pp. 187–192, 2020.
- [3] A. R. Setiawan, "Lembar Kegiatan Literasi Sainifik Untuk Pembelajaran Jarak Jauh Topik Penyakit Coronavirus 2019 (Covid-19)," Vol. 2, No. 1, Pp. 28–36, 2020.
- [4] Who, "Covid-19.Who.Int/Table," 2021. .
- [5] S. Sindi *Et Al.*, "Analisis Algoritma K-Medoids Clustering Dalam Pengelompokan Penyebaran Covid-19 Di Indonesia," Vol. 4, No. 1, Pp. 166–173, 2020.
- [6] N. Dwitri, J. A. Tampubolon, S. Prayoga, And P. P. P. A. N. W. F. I. R. H. Zer, "Penerapan Algoritma K-Means Dalam Menentukan Tingkat Penyebaran Pandemi Covid-19 Di Indonesia," Vol. 4, No. 1, Pp. 128–132, 2020.
- [7] R. Kemenkes, "Situasi Terkini Perkembangan Novel Coronavirus (Covid-19)," 2021.
- [8] D. Lampung, "Rekapitulasi Data Pemantauan Covid-19 Provinsi Lampung," 2021. .
- [9] G. T. Tim Pakar, "Menuju Tahapan Indonesia Produktif Dan Aman Covid-19," Pp. 1–29, 2020.
- [10] J. Han, M. Kamber, And J. Oei, *Data Mining Concepts And Techniques*. 2012.
- [11] D. T. Larose And C. D. Larose, *Discovering Knowledge In Data An Introduction To Data Mining*. 2014.
- [12] S. Susanto And D. Suryadi, *Pengantar Data Mining Menggali Pengetahuan Bongkahan Data*. 2010.
- [13] A. F. Muhammad, "Klasterisasi Proses Seleksi Pemain Menggunakan Algoritma K-Means," Pp. 1–5, 2015.
- [14] K. K. R. Indonesia, "Pertanyaan Dan Jawaban Terkait Coronavirus Disease 2019 (Covid-19) Update 6 Maret 2020," Vol. 2019, Pp. 1–9, 2020.
- [15] A. Wanto *Et Al.*, *Data Mining Algoritma Dan Implementasi*. 2020.
- [16] D. Aprilla, D. A. Baskoro, L. Ambarwati, And I. W. S. Wicaksana, *Belajar Data Mining Dengan Rapid Miner*. 2013.
- [17] P. L. Bpbd, *Laporan Situasi Harian Penanganan Covid-19 Di Wilayah Provinsi Lampung*. 2021.
- [18] I Ahmad, H Sulistiani, H Saputra 2018. *The Application Of Fuzzy K-Nearest Neighbour Methods For A Student Graduation Rate*. Indonesian Journal Of Artificial Intelligence And Data Mining 1 (1), Pp 47-52.
- [19] Z Abidin, Permata, I Ahmad, And Rusliyawati. 2020. *Effect Of Mono Corpus Quantity On Statistical Machine Translation Indonesian – Lampung Dialect Of Nyo*. *Journal Of Physics: Conference Series*. Icasmi 2020. Doi:10.1088/1742-6596/1751/1/012036
- [20] Isnain Ar, Ns Marga, D Alita. 2021. *Sentiment Analysis Of Government Policy On Corona Case Using Naive Bayes Algorithm*. Ar Ijccs (Indonesian Journal Of Computing And Cybernetics Systems) 15 (1), 55-64.
- [21] Isnain Ar, J Supriyanto, Mp Kharisma. 2020. *Implementation Of K-Nearest Neighbor (K-Nn) Algorithm For Public Sentiment Analysis Of Online Learning*. Ijccs (Indonesian Journal Of Computing And Cybernetics Systems) 15 (2), 121-130
- [22] P Prasetyawan, I Ahmad, Ri Borman, Ya Pahlevi, De Kurniawan. 2018. *Classification Of The Period Undergraduate Study Using Back-Propagation Neural Network*. International Conference On Applied Engineering (Icae), 1-5
- [23] D Alita, I Sari, Ar Isnain, S Styawati. 2021. *Penerapan Naive Bayes Classifier Untuk Pendukung Keputusan Penerima Beasiswa*. Jurnal Data Mining Dan Sistem Informasi 2 (1), 17-23
- [24] N Neneng, K Adi, R Isnanto. 2016. *Support Vector Machine Untuk Klasifikasi Citra Jenis Daging Berdasarkan Tekstur Menggunakan Ekstraksi Ciri Gray Level Co-Occurrence Matrices (GlcM)*. Jsinbis (Jurnal Sistem Informasi Bisnis) 6 (1), 1-10
- [25] N Shodik, N Neneng, I Ahmad. 2019. *Sistem Rekomendasi Pemilihan Smartphone Snapdragon 636 Menggunakan Metode Simple Multi Attribute Rating Technique (Smart)*. Jurnal Nasional Pendidikan Teknik Informatika: Janapati 7 (3), 219-228
- [26] Rd Kurniawati, I Ahmad. 2021. *Sistem Pendukung Keputusan Penentuan Kelayakan Usaha Mikro Kecil Menengah Dengan Menggunakan Metode Profile Matching Pada Uptd Plut Kumkm Provinsi Lampung*. Jurnal Teknologi Dan Sistem Informasi 2 (1), 74-79.

- [27] R Ruslaini, A Abizar, N Ramadhani, I Ahmad. 2021. *Peningkatan Manajemen Dan Teknologi Pemasaran Pada Umkm Ojesa (Ojek Sahabat Wanita) Dalam Mengatasi Less Contact Ekonomi Masa Covid-19*. Martabe: Jurnal Pengabdian Kepada Masyarakat 4 (1), 139-144.
- [28] S Pramono, I Ahmad, Ri Borman. 2020. *Analisis Potensi Dan Strategi Penembaan Ekowisata Daerah Penyanga Taman Nasional Way Kambas*. Jurnal Teknologi Dan Sistem Informasi 1 (1), 57-67.
- [29] H Sulistiani, I Darwanto, I Ahmad. 2020. *Penerapan Metode Case Based Reasoning Dan K-Nearest Neighbor Untuk Diagnosa Penyakit Dan Hama Pada Tanaman Karet*. Jepin (Jurnal Edukasi Dan Penelitian Informatika) 6 (1), 23-28
- [30] N Neneng, Y Fernando. 2017. *Klasifikasi Jenis Daging Berdasarkan Analisis Citra Tekstur Gray Level Co-Occurrence Matrices (GlcM) Dan Warna*. Prosiding Semnastek.
- [31] S Ningsih, S Saniati. 2018. *Eksperimen Pengenalan Ucapan Aksara Lampung Dengan Cmu Sphinx 4*. Jurnal Teknoinfo 12 (1), 33-37
- [32] M Bakri, R Wakhidah 2018. *Penerapan Klasterisasi K-Means Untuk Identifikasi Sebaran Budidaya Udang Vanname*. Seminar Nasional Penerapan Ilmu Pengetahuan Dan Teknologi 2018, 179-184