

Visualisasi Data Mining Cluster Covid-19 Menggunakan K-means dan Aplikasi Google Colaboratory

Diana Laily Fithri¹, Fajar Nugraha², Noor Latifah³

¹⁾ Program Studi Sistem Informasi, Universitas Muria Kudus
Jl. Lingkar Utara, Gondangmanis, Bae, Kudus
Email: Diana.laily@umk.ac.id

ABSTRAK

Pada tahun 2020 secara merata penduduk dunia digemparkan dengan virus covid-19, tidak terkecuali masyarakat Indonesia yang secara langsung juga mulai terdampak dengan kasus penderita sembuh dan meninggal. Indonesia memiliki data dari beberapa provinsi yang terdampak dengan beberapa kriteria cluster, yaitu cluster 1 sampai dengan cluster 4. Dalam pengelompokan cluster tersebut banyak masyarakat yang mengalami masalah dengan gejala ringan sampai dengan gejala berat. Oleh karena itu, kami melakukan penelitian untuk melakukan pengelompokan kasus covid-19 dengan menggunakan data mining. Data mining tersebut digunakan untuk menentukan pola sesuai dengan algoritma yang digunakan. Algoritma K-means sangat membantu dalam mengatasi permasalahan dalam melakukan pengelompokan kasus covid-19 ini karena dapat menghasilkan cluster sesuai dengan data yang ada. Hasil cluster yang telah terlihat adalah 2 provinsi cluster zona merah untuk penderita kasus Covid-19 terdiri dari provinsi DKI Jakarta dan Jawa Barat, 2 provinsi cluster zona orange terdiri dari provinsi Jawa Tengah dan dan Jawa Timur, 8 provinsi cluster zona kuning terdiri dari provinsi Bali, Banten, DI Yogyakarta, Sumatera Barat, Kalimantan Timur, Sumatera Utara, Riau, Sulawesi Selatan, dan 22 provinsi lainnya termasuk cluster zona hijau

Kata kunci: Cluster, Data, Analisa, covid-19.

ABSTRACT

In 2020, the world's population was evenly shocked by the Covid-19 virus, including the Indonesian people who were also directly affected by cases of patients recovering and dying. Indonesia has data from several affected provinces with several cluster criteria, namely cluster 1 to cluster 4. In this cluster grouping, many people experience problems with mild to severe symptoms. Therefore, we conducted a study to group Covid-19 cases using data mining. The data mining is used to determine the pattern according to the algorithm used. The K-means algorithm is very helpful in overcoming the problem of grouping Covid-19 cases because it can produce clusters according to the existing data. The cluster results that have been seen are 2 red zone cluster provinces for Covid-19 case sufferers consisting of DKI Jakarta and West Java provinces, 2 orange zone cluster provinces consisting of Central Java and East Java provinces, 8 yellow zone cluster provinces consisting of Bali province. , Banten, DI Yogyakarta, East Kalimantan, West Sumatra, North Sumatra, South Sulawesi, and Riau, and 22 other provinces including green zone clusters

Keywords: Cluster, Data, Analysis, Covid-19

Pendahuluan

Pada tahun 2020, dunia dikejutkan oleh virus yang berasal dari Wuhan, Hubei, China yang diakibatkan virus *Severe Acute Respiratory Syndrome Coronavirus – 2* (SARS – CoV – 2) (Damanik *et al.*, 2021, Febrina, 2021). Munculnya COVID-19 yang merupakan virus menyerang sistem pernapasan, yang dapat memberikan dampak buruk bagi penderitanya disertai gejala yang ringan maupun yang berat, *Middle East Respiratory Syndrome* (MERS) dan *Severe Acute Respiratory Syndrome* (SARS) . Virus covid-19 ini dapat menular melalui kontak fisik, dan menggunakan barang secara bergantian dengan penderita covid-19, selain itu penyebaran juga bisa dari ketika berdekatan dengan seseorang yang menderita batuk, bersin, berbicara, hingga bernapas dengan penderita covid-19. Penyebaran COVID-19 juga bisa terjadi melalui udara, permukaan yang terkontaminasi virus tersebut, dan limbah manusia sendiri, tidak memakai masker pada saat berbicara dengan penderita Covid-19, dsb(Elsa Ramadanti and Muhamad Muslih, 2021). Virus ini tidak dapat diprediksi sebelumnya, karena virus tersebut berjalan begitu cepat menjangar ke tubuh apalagi jika penderita tersebut memiliki riwayat penyakit yang kompleks.

Gejala covid-19 yang tergolong berat dapat terjadi ketika pasien mengalami gejala yang fatal dan berakibat kematian. Gejala ringan yang bisa terjadi yaitu demam, bersin, sakit tenggorokan dan lain sebagainya. Dari hal tersebut kami mencoba untuk menganalisis data Covid-19 yang ada untuk dijadikan informasi, dengan cara mengelompokkannya agar dapat menemukan jarak status dan zonasi Covid-19(Damanik *et al.*, 2021). Beberapa daerah memiliki jumlah penderita yang cukup banyak, sehingga mempersulit didalam pengelompokan atau clustering terhadap covid-19 tersebut. Dengan adanya permasalahan tersebut, kami ingin melakukan sebuah penelitian dengan menggunakan data terdampak covid-19 dengan jumlah kasus, pasien sembuh dan pasien meninggal. Clustering atau pengelompokan amat berjasa karena dapat mempermudah serta dapat menciptakan kelompok atau golongan yang tidak diketahui serta belum dikenal dalam membentuk suatu informasi(Anggiat *et al.*, 2022). Untuk mengatasi permasalahan tersebut, kami ingin melakukan sebuah penelitian yang berhubungan dengan pengelompokan kasus covid-19 di beberapa daerah dengan menggunakan algoritma data mining K-means. Data mining serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual(Desyanti, 2021). K-Means adalah algoritma data clustering non hirarki yang digunakan untuk membagi data yang ada dalam bentuk satu atau lebih cluster/kelompok.(Damanik *et al.*, 2021). Algoritma K-means adalah algoritma cluster yang dapat mengelompokan model dari pola untuk memisahkan data menjadi beberapa kelompok yang sama. Data yang berbeda nantinya dapat dikelompokan menjadi beberapa kelompok data cluster yang lain (Hardiani, 2022). Penggunaan teknik *data mining* diharapkan dapat memberikan pengetahuan yang sebelumnya tersembunyi di dalam gudang data se-hingga menjadi informasi yang berharga, dari beberapa algoritma yang ada pada data mining akan dipilih algoritma Apriori sebagai langkah dalam menghasilkan knowledge (Wahyuni *et al.*, 2021).

Dilihat dari masalah diatas teknik penyelesaian yang digunakan dengan analisa cluster K-means. Pengelompokan analisis cluster merupakan suatu teknik *multivariate* dengan tujuan agar bisa mengahasilka beberapa cluster dengan data kasus covid-19 yang ada di beberapa provinsi. Dimana hasilnya nanti dapat digunakan untuk memberikan bantuan dari masyarakat yang terdampak, dan nantinya dapat tepat, terarah dan sesuai sasaran.

Metode Penelitian

1. Metode Pengumpulan Data

Metode pengumpulan data yang digunakan dalam pebelitian ini menggunakan metode pustaka dengan memahami, menganalisa serta membuat hipotesa sederhana terlebih dahulu. Data pengelompokan kasus virus Covid-19 didapatkan dari dokumen dengan keterangan jumlah kasus yang jelas.

2. Metode Analisis Data

Analisa data dilakukan dengan melakukan proses preprocessing data dengan memisahkan antara data yang satu dengan data yang lain, dengan konsep algoritma K-means nantinya dapat menghasilkan cluster pengelompokan untuk kasus covid-19 berdasarkan provinsi dan zona yang ada, zona cluster K-means yaitu zona hijau, zona kuning, zona orange, zona merah. Analisa yang dilakukan melalui observasi dan pengumpulan data serta dokumentasi dimana data akan dipelajari sebelumnya dan membuat pola dari data tersebut dan nantinya dapat memberikan kesimpulan berupa cluster covid-19. Tahapan dalam pengelompokan analisa cluster yaitu, menentukan ukuran kemiripan objek data yang diteliti yaitu data penderita covid-19, Melakukan proses standarisasi data covid-19, Melakukan pengclusteran sesuai dengan cluster yang telah ditentukan, dan Melakukan validasi cluster (*Analisis Cluster untuk Pemetaan Data Kasus Covid-19 di Indonesia Menggunakan K-Means*, no date)

3. Studi Literatur

Studi literature berguna agar dapat menemukan jawaban dari setiap masalah yaitu untuk data covid-19, dan nantinya dapat menemukan model cluster yang dihasilkan dari pola yang telah dibuat dengan menggunakan algoritma K-means.

Hasil dan Pembahasan

Mengenakan masker untuk melawan Covid-19 bermanfaat dalam menekan penyebaran pandemi, bukan dengan mencegah pemakainya agar tidak terinfeksi tetapi dengan mencegah pemakainya menulari orang lain (Masriani, Alam and Arham, 2022). Data mining atau sering juga disebut dengan penambangan suatu data yang nantinya dapat menentukan pola yang sesuai dengan algoritma yang digunakan dan terdiri dari kumpulan data yang memiliki pola pengelompokan yang berbeda-beda. Data tersebut tersimpan di suatu database misalnya gudang data, atau tempat penyimpanan data (Hardiani, 2022). Konsep database data mining adalah salah satu cara yang berpusat dengan cara totalitas cara temuan ilmu atau

knowledge dari informasi, tercantum cara penyimpanan serta pengaksesan informasi (Anggiat *et al.*, 2022).

Model pengelompokan untuk data kasus penyakit covid-19 ini dengan menggunakan sampel data yang ada pada tahun 2021, kemudian akan dilakukan preprocessing dan pembentukan pola dengan cluster K-Means. Data hasil covid 2019 akan ditampilkan pada tabel berikut.

Tabel 1. Data kasus covid-19 berdasarkan provinsi

No	Provinsi Asal	Kasus	Sembuh	Kematian
1	Aceh	37.733	34.027	1.927
2	Bali	112.582	107.627	3.943
3	Banten	131.475	128.002	2.670
4	Bangka Belitung	51.209	48.924	1.397
5	Bengkulu	23.021	22.463	466
6	DI Yogyakarta	154.829	148.125	5.190
7	DKI Jakarta	857.765	842.510	13.539
8	Jambi	29.547	28.366	767
9	Jawa Barat	702.722	685.219	14.624
10	Jawa Tengah	482.009	448.597	29.894
11	Jawa Timur	395.475	364.251	29.413
12	Kalimantan Barat	39.961	38.711	1.045
13	Kalimantan Timur	156.811	150.282	5.381
14	Kalimantan Tengah	45.162	42.338	1.367
15	Kalimantan Selatan	69.425	66.531	2.359
16	Kalimantan Utara	31.202	31.202	773
17	Kepulauan Riau	53.633	51.435	1.737
18	Nusa Tenggara Barat	27.423	26.380	799
19	Sumatera Selatan	59.615	56.184	3.046
20	Sumatera Barat	89.276	86.244	2.121
21	Sulawesi Utara	34.179	32.317	1.028
22	Sumatera Utara	104.706	100.076	2.837
23	Sulawesi Tenggara	20.030	19.216	523
24	Sulawesi Selatan	108.584	104.498	2.206
25	Sulawesi Tengah	46.341	43.858	1.561
26	Lampung	49.064	43.964	3.764
27	Riau	127.735	122.629	4.064
28	Maluku Utara	11.956	11.567	302
29	Maluku	14.524	14.113	258
30	Papua Barat	22.899	22.461	352
31	Papua	33.746	31.441	496
32	Sulawesi Barat	12.159	11.567	336
33	Nusa Tenggara Timur	62.716	60.747	1.296
34	Gorontalo	11.735	11.152	458

Langkah-langkah dalam pengelompokan menggunakan K-means :

1. Penentuan manual jumlah cluster

Penentuan cluster dibagi menjadi empat cluster yaitu zona kuning, zona merah, zona hijau, zona orange.

Tabel 2 . Penentuan 4 cluster awal

PENENTUAN CLUSTER BARU			
Cluster	Kasus	Sembuh	Kematian
1 - Zona Hijau	31343,368	29763,263	1018,737
2 - Zona Kuning	107068,545	102813,182	3192,091
3 - Zona Orange	438742,000	406424,000	29653,500
4 - Zona Merah	780243,500	763864,500	14081,500

2. Clustering data

Dari tabel data Kasus Penyakit Covid-19 pada tahun 2021 berdasarkan data Provinsi didapatkan pengelompokan suatu data pada iterasi 1 untuk 4 cluster tersebut. Cluster zona hijau (C1) ada 21 Provinsi, cluster zona kuning (C2) ada 9 Provinsi, cluster zona orange (C3) sebanyak 2 kasus yakni Provinsi Jawa Tengah dan Jawa Timur, serta cluster zona merah (C4) ada 2 provinsi yakni DKI Jakarta dan Jawa Barat. Proses pencarian dan penentuan jarak terdekat, pengelompokan data pada iterasi 1 dan clustering ditampilkan pada tabel berikut:

Tabel 3. Perhitungan jarak pusat cluster iterasi 1

No	Provinsi Asal	Kasus	Sembuh	Kematian	C1	C2	C3	C4	Jarak Terpendek	Kelompok
1	Aceh	37.733	34.027	1.927	7735,100874	97675,77554	547956,6611	1041216,765	7735,100874	1
2	Bali	112.582	107.627	3.943	112565,6612	7357,641672	443081,2822	936227,7096	7357,641672	2
3	Banten	131.475	128.002	2.670	140285,1372	35077,40247	415523,68	908488,8045	35077,40247	2
4	Bangka Belitung	51.209	48.924	1.397	27602,9027	77637,3305	528002,373	1021171,933	27602,9027	1
5	Bengkulu	23.021	22.463	466	11084,27608	116308,096	566658,5476	1059833,704	11084,27608	1
6	DI Yogyakarta	154.829	148.125	5.190	171101,1437	65865,11615	384608,1484	87700,1125	65865,11615	2
7	DKI Jakarta	857.765	842.510	13.539	1159175,021	1053946,682	604988,3892	110430,9374	110430,9374	4
8	Jambi	29.547	28.366	767	2289,684536	107507,46	557855,5631	1051037,846	2289,684536	1
9	Jawa Barat	702.722	685.219	14.624	938379,7702	833144,8034	384236,8778	110430,9374	110430,9374	4
10	Jawa Tengah	482.009	448.597	29.894	615918,0071	510744,3444	60420,63437	434266,5641	60420,63437	3
11	Jawa Timur	395.475	364.251	29.413	495257,6354	390148,1153	60420,63437	554952,9736	60420,63437	3
12	Kalimantan Barat	39.961	38.711	1.045	12422,81202	92828,45668	543192,0311	1036356,95	12422,81202	1
13	Kalimantan Timur	156.811	150.282	5.381	174028,5087	68792,3819	381684,3816	874772,6934	68792,3819	2
14	Kalimantan Tengah	45.162	42.338	1.367	18686,89039	86562,11032	536902,1865	1030100,461	18686,89039	1
15	Kalimantan Selatan	69.425	66.531	2.359	5291,61415	52288,88288	502660,2109	995828,4828	52288,88288	2

16	Kalimantan Utara	31.202	31.202	773	1466,402171	104353,9462	554720,1855	1047871,473	1466,402171	1
17	Kepulauan Riau	53.633	51.435	1.737	31096,74831	74143,05345	524504,9694	1017679,221	31096,74831	1
18	Nusa Tenggara Barat	27.423	26.380	799	5183,053389	110413,6363	560757,83	1053944,267	5183,053389	1
19	Sumatera Selatan	59.615	56.184	3.046	38748,55202	66529,24857	516830,0483	1010068,764	38748,55202	1
20	Sumatera Barat	89.276	86.244	2.121	80916,49041	24336,38625	474762,842	967857,7797	24336,38625	2
21	Sulawesi Utara	34.179	32.317	1.028	3816,079679	101426,2332	551766,878	1044961,424	3816,079679	1
22	Sumatera Utara	104.706	100.076	2.837	101632,9806	3633,163206	454035,5394	947149,6848	3633,163206	2
23	Sulawesi Tenggara	20.030	19.216	523	15475,23248	120711,7279	571050,2257	1064241,414	15475,23248	1
24	Sulawesi Selatan	108.584	104.498	2.206	107483,9786	2471,353893	448237,6375	941292,5649	2471,353893	2
25	Sulawesi Tengah	46.341	43.858	1.561	20588,45815	84653,46241	534997,0446	1028192,104	20588,45815	1
26	Lampung	49.064	43.964	3.764	22873,96293	82632,20059	532819,3517	1026151,374	22873,96293	1
27	Riau	127.735	122.629	4.064	133883,0286	28644,8812	421805,38	914904,7269	28644,8812	2
28	Maluku Utara	11.956	11.567	302	26598,64205	131835,559	582168,2324	1075363,699	26598,64205	1
29	Maluku	14.524	14.113	258	22986,96612	128221,7768	578561,0838	1071748,456	22986,96612	1
30	Papua Barat	22.899	22.461	352	11183,69095	116400,3936	566755,2964	1059923,743	11183,69095	1
31	Papua	33.746	31.441	496	2976,691657	102359,43	552706,2241	1045890,657	2976,691657	1
32	Sulawesi Barat	12.159	11.567	336	26450,14394	131688,4383	582017,7163	1075218,241	26450,14394	1
33	Nusa Tenggara Timur	62.716	60.747	1.296	44094,34018	61158,05018	511558,6856	1004680,746	44094,34018	1
34	Gorontalo	11.735	11.152	458	27040,37082	132278,9945	582603,9043	1075809,938	27040,37082	1

**Tabel 4. Hasil pengelompokan iterasi 1
 ATAU DENGAN KESIMPULAN**

Cluster	Total Kasus
C1	21
C2	9
C3	2
C4	2

Proses K-means akan melakukan iterasi sampai pengelompokan data sama dengan pengelompokan data iterasi yang dulu (Sroyer, Mandowen and Reba, 2022). Dengan kata lain, proses iterasi akan terus dilakukan sampai dengan data pada iterasi berikutnya, dan akan dilakukan proses mencari nilai rata-rata dengan cara menambahkan semua nilai pada tiap kelompok (cluster), lalu dibagi dengan total kasus pada cluster tersebut. Nilai rata-rata tersebut akan digunakan sebagai penentuan cluster terbaru. Lalu akan dilakukan dengan proses mencari jarak terpendek dari suatu data kasus covid-19, cluster iterasi 2 dapat dilihat pada hasil table berikut.

Tabel 5. Penentuan cluster 2

CLUSTER BARU 2			
Cluster	Kasus	Sembuh	Kematian
1	34183,571	32496,810	1128,476
2	117269,222	112668,222	3419,000

3	438742,000	406424,000	29653,500
4	780243,500	763864,500	14081,500

Tabel 6. Perhitungan Jarak Pusat Cluster Iterasi 2

No	Provinsi Asal	Kasus	Sembuh	Kematian	C1	C2	C3	C4	Jarak Terpendek	Kelompok
1	Aceh	37.733	34.027	1.927	3946,84258	111860,0847	547956,6611	1041216,765	3946,84258	1
2	Bali	112.582	107.627	3.943	108622,1923	6903,517194	443081,2822	936227,7096	6903,517194	2
3	Banten	131.475	128.002	2.670	136342,3624	20916,25837	415523,68	908488,8045	20916,25837	2
4	Bangka Belitung	51.209	48.924	1.397	23659,87975	91822,47715	528002,373	1021171,933	23659,87975	1
5	Bengkulu	23.021	22.463	466	15023,95451	130493,0256	566658,5476	1059833,704	15023,95451	1
6	DI Yogyakarta	154.829	148.125	5.190	167157,6915	51682,26425	384608,1484	877700,1125	51682,26425	2
7	DKI Jakarta	857.765	842.510	13.539	1155232,34	1039762,2	604988,3892	110430,9374	110430,9374	4
8	Jambi	29.547	28.366	767	6220,293158	121692,5883	557855,5631	1051037,846	6220,293158	1
9	Jawa Barat	702.722	685.219	14.624	934436,7381	818959,645	384236,8778	110430,9374	110430,9374	4
10	Jawa Tengah	482.009	448.597	29.894	611975,8478	496572,4266	60420,63437	434266,5641	60420,63437	3
11	Jawa Timur	395.475	364.251	29.413	491316,9583	375989,4106	60420,63437	554952,9736	60420,63437	3
12	Kalimantan Barat	39.961	38.711	1.045	8485,388173	107013,4002	543192,0311	1036356,95	8485,388173	1
13	Kalimantan Timur	156.811	150.282	5.381	170085,0414	54609,50387	381684,3816	874772,6934	54609,50387	2
14	Kalimantan Tengah	45.162	42.338	1.367	14745,56942	100747,2201	536902,1865	1030100,461	14745,56942	1
15	Kalimantan Selatan	69.425	66.531	2.359	49008,14828	66474,32944	502660,2109	995828,4828	49008,14828	1
16	Kalimantan Utara	31.202	31.202	773	3269,963793	118538,2361	554720,1855	1047871,473	3269,963793	1
17	Kepulauan Riau	53.633	51.435	1.737	27153,37238	88328,39524	524504,9694	1017679,221	27153,37238	1
18	Nusa Tenggara Barat	27.423	26.380	799	9123,006048	124598,8176	560757,83	1053944,267	9123,006048	1
19	Sumatera Selatan	59.615	56.184	3.046	34806,85923	80713,17011	516830,0483	1010068,764	34806,85923	1
20	Sumatera Barat	89.276	86.244	2.121	76973,51021	38516,8121	474762,842	967857,7797	38516,8121	2
21	Sulawesi Utara	34.179	32.317	1.028	206,0287059	115611,5082	551766,878	1044961,424	206,0287059	1
22	Sumatera Utara	104.706	100.076	2.837	97689,7076	17797,11598	454035,5394	947149,6848	17797,11598	2
23	Sulawesi Tenggara	20.030	19.216	523	19418,29259	134896,8902	571050,2257	1064241,414	19418,29259	1
24	Sulawesi Selatan	108.584	104.498	2.206	103541,0849	11985,6992	448237,6375	941292,5649	11985,6992	2
25	Sulawesi Tengah	46.341	43.858	1.561	16645,32353	98838,81603	534997,0446	1028192,104	16645,32353	1
26	Lampung	49.064	43.964	3.764	18970,22925	96810,85432	532819,3517	1026151,374	18970,22925	1
27	Riau	127.735	122.629	4.064	129939,5969	14462,55937	421805,38	914904,7269	14462,55937	2
28	Maluku Utara	11.956	11.567	302	30541,85524	146020,7095	582168,2324	1075363,699	30541,85524	1
29	Maluku	14.524	14.113	258	26929,92629	142406,8272	578561,0838	1071748,456	26929,92629	1
30	Papua Barat	22.899	22.461	352	15121,57202	130585,1763	566755,2946	1059923,743	15121,57202	1
31	Papua	33.746	31.441	496	1306,22687	116544,1299	552706,2241	1045890,657	1306,22687	1
32	Sulawesi Barat	12.159	11.567	336	30393,53043	145873,6468	582017,7163	1075218,241	30393,53043	1
33	Nusa Tenggara Timur	62.716	60.747	1.296	40152,22044	75341,7182	511558,6856	1004680,746	40152,22044	1
34	Gorontalo	11.735	11.152	458	30983,68589	146464,2719	582603,9043	1075809,938	30983,68589	1

Tabel 7. Pengelompokan kelompok cluster

No	Provinsi Asal	Kasus	Sembuh	Kematian	Kelompok
1	Aceh	37.733	34.027	1.927	1
2	Bali	112.582	107.627	3.943	2
3	Banten	131.475	128.002	2.670	2
4	Bangka Belitung	51.209	48.924	1.397	1
5	Bengkulu	23.021	22.463	466	1
6	DI Yogyakarta	154.829	148.125	5.190	2
7	DKI Jakarta	857.765	842.510	13.539	4
8	Jambi	29.547	28.366	767	1
9	Jawa Barat	702.722	685.219	14.624	4
10	Jawa Tengah	482.009	448.597	29.894	3
11	Jawa Timur	395.475	364.251	29.413	3
12	Kalimantan Barat	39.961	38.711	1.045	1
13	Kalimantan Timur	156.811	150.282	5.381	2
14	Kalimantan Tengah	45.162	42.338	1.367	1
15	Kalimantan Selatan	69.425	66.531	2.359	1
16	Kalimantan Utara	31.202	31.202	773	1
17	Kepulauan Riau	53.633	51.435	1.737	1
18	Nusa Tenggara Barat	27.423	26.380	799	1
19	Sumatera Selatan	59.615	56.184	3.046	1
20	Sumatera Barat	89.276	86.244	2.121	2
21	Sulawesi Utara	34.179	32.317	1.028	1
22	Sumatera Utara	104.706	100.076	2.837	2
23	Sulawesi Tenggara	20.030	19.216	523	1
24	Sulawesi Selatan	108.584	104.498	2.206	2
25	Sulawesi Tengah	46.341	43.858	1.561	1
26	Lampung	49.064	43.964	3.764	1
27	Riau	127.735	122.629	4.064	2
28	Maluku Utara	11.956	11.567	302	1
29	Maluku	14.524	14.113	258	1
30	Papua Barat	22.899	22.461	352	1
31	Papua	33.746	31.441	496	1
32	Sulawesi Barat	12.159	11.567	336	1
33	Nusa Tenggara Timur	62.716	60.747	1.296	1
34	Gorontalo	11.735	11.152	458	1

Tabel 8. Nilai rata-rata cluster

Nilai Rata-rata Cluster			
C1	C2	C3	C4
35785,45455	123249,75	438742	780243,5
34043,81818	118435,375	406424	763864,5
1184,409091	3551,5	29653,5	14081,5

Tabel 9. Hasil pengelompokan iterasi 3

ATAU DENGAN KESIMPULAN	
Cluster	Total Kasus
C1	22
C2	8
C3	2
C4	2

3. Analisa Data

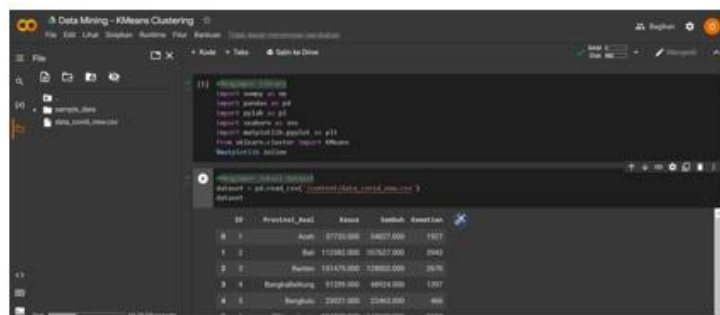
Pemrosesan data mining juga diperlukan suatu analisa data yang merupakan cara untuk mengtasi permasalahan penelitian melalui prosedur penegeloaan data terutama masalah yang berkaitan dengan penelitian(*Analisis Cluster untuk Pemetaan Data Kasus Covid-19 di Indonesia Menggunakan K-Means*, no date)

Proses data mining dilakukan secara berulang-ulang dengan meghasilkan suatu iterasi. Pada iterasi 3, cluster data yang dilakukan terhadap 4 cluster dengan iterasi 2 didapatkan hasil yang sama. Dari 34 data jumlah kasus penderita Covid-19 berdasarkan daerah dapat diketahui, 2 provinsi berada dalam pengelompokan/cluster zona merah yaitu provinsi DKI Jakarta dan Jawa Barat, 2 provinsi atau daerah berada dalam pengelompokan cluster zona orange terdiri dari provinsi Jawa Tengah dan dan Jawa Timur, 8 provinsi atau daerah berada dalam kelompok atau cluster zona kuning terdiri dari provinsi Bali, Banten, DI Yogyakarta, Kalimantan Timur, Sumatera Barat, Sumatera Utara, Sulawesi Selatan, dan Riau, dan 22 provinsi lainnya termasuk cluster kelompok zona hijau.

4. Implementasi K-Means Clustering dengan Google Colaboratory

- a. Mengimpor library dan Mengimpor lokasi dataset

Dalam proses import library dan import lokasi dataset dengan memasukkan data provinsi sesuai data yang ada dengan total kasus beserta sembuh dan data kematian.



Gambar 1. Proses library import dataset

- b. Proses data set K-means yaitu dengan memasukkan semua data ke aplikasi tersebut berdasarkan kasus, pasien sembuh dan meninggal.

Kode	Teks	1	2	3
13	KalimantanTengah	45162.000	42338.000	1367
14	KalimantanSelatan	69425.000	66531.000	2359
15	KalimantanUtara	31202.000	31202.000	773
16	KepulauanRiau	53633.000	51435.000	1737
17	NusaTenggaraBarat	27423.000	26380.000	799
18	SumateraSelatan	59615.000	56184.000	3046
19	SumateraBarat	89276.000	86244.000	2121
20	SulawesiUtara	34179.000	32317.000	1028
21	SumateraUtara	104706.000	100076.000	2837
22	SulawesiTenggara	20030.000	19216.000	523
23	SulawesiSelatan	108584.000	104498.000	2206
24	SulawesiTengah	46341.000	43858.000	1561
25	Lampung	49064.000	43964.000	3764
26	Riau	127735.000	122629.000	4064
27	MalukuUtara	11956.000	11567.000	302
28	Maluku	14524.000	14113.000	258
29	PapuaBarat	22899.000	22461.000	352
30	Papua	33746.000	31441.000	496

Gambar 2. Proses dataset K-means

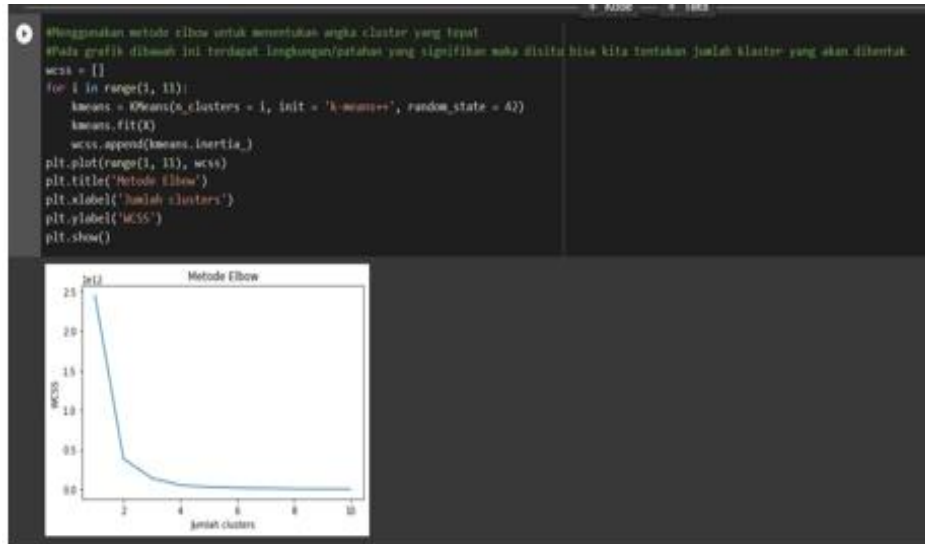
- c. Proses ketiga, yaitu dengan melakukan preprocessing data dengan langkah menghilangkan masalah yang bisa mengganggu saat pemrosesan data.

```

#Melakukan preprocessing
X = dataset.iloc[:,2:5].values
#Nilai values akan menghasilkan numpy array.
X[[3]]
array([[3.7730e+04, 3.4027e+04, 1.9270e+03],
       [1.1250e+05, 1.0727e+05, 3.9430e+03],
       [1.3175e+05, 1.2800e+05, 2.6700e+03],
       [5.1209e+04, 4.8920e+04, 1.3970e+03],
       [2.3021e+04, 2.2463e+04, 4.6600e+02],
       [1.5482e+05, 1.4812e+05, 5.1900e+03],
       [8.5770e+05, 8.4251e+05, 1.3539e+04],
       [2.9547e+04, 2.8366e+04, 7.6700e+02],
       [7.0272e+05, 6.8521e+05, 1.4624e+04],
       [4.8200e+05, 4.4859e+05, 2.9894e+04],
       [3.9547e+05, 3.6425e+05, 2.9433e+04],
       [3.9961e+04, 3.8711e+04, 1.0450e+03],
       [1.5681e+05, 1.5082e+05, 5.3810e+03],
       [4.5162e+04, 4.2338e+04, 1.3670e+03],
       [6.9425e+04, 6.6511e+04, 2.3590e+03],
       [3.1202e+04, 3.1020e+04, 7.7300e+02],
       [5.3633e+04, 5.1435e+04, 1.7370e+03],
       [2.7423e+04, 2.6380e+04, 7.9900e+02],
       [5.9615e+04, 5.6184e+04, 3.0400e+03],
       [8.9276e+04, 8.6244e+04, 2.1210e+03],
       [3.4179e+04, 3.2317e+04, 1.0380e+03],
       [1.0470e+05, 1.0007e+05, 2.8370e+03],
       [2.8030e+04, 1.9216e+04, 5.2300e+02],
       [1.8858e+05, 1.8449e+05, 2.2060e+03],
       [4.6341e+04, 4.3858e+04, 1.5610e+03],
       [4.9064e+04, 4.3964e+04, 3.7640e+03],
       [1.2773e+05, 1.2262e+05, 4.0640e+03],
       [1.1956e+04, 1.1567e+04, 3.0200e+02],
       [1.4524e+04, 1.4113e+04, 2.5800e+02],
       [2.2899e+04, 2.2461e+04, 3.5200e+02],
       [3.3746e+04, 3.1441e+04, 4.9600e+02])
    
```

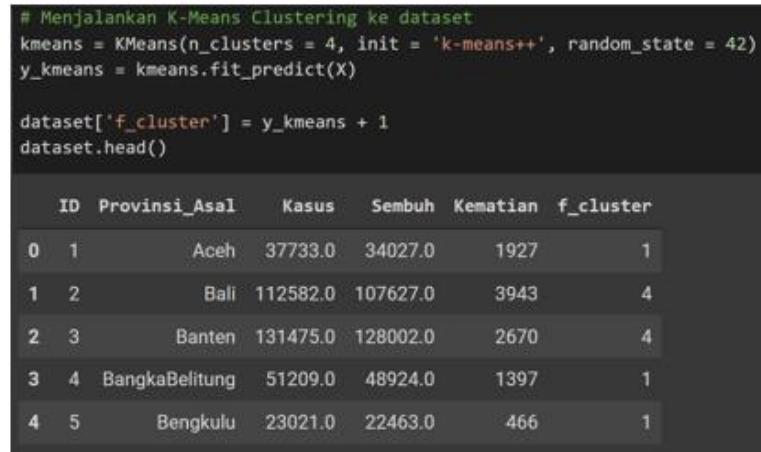
Gambar 3. Proses Preprocessing

- d. Dalam proses clustering ke data set akan merubah bentuk pola yang dihasilkan dari data tersebut.



Gambar 4. Menjalankan K-Means Clustering ke dataset

- e. Visualisasi hasil cluster adalah model custer yang akan dihasilkan dengan hasil f-cluster 1 sampai dengan 4.



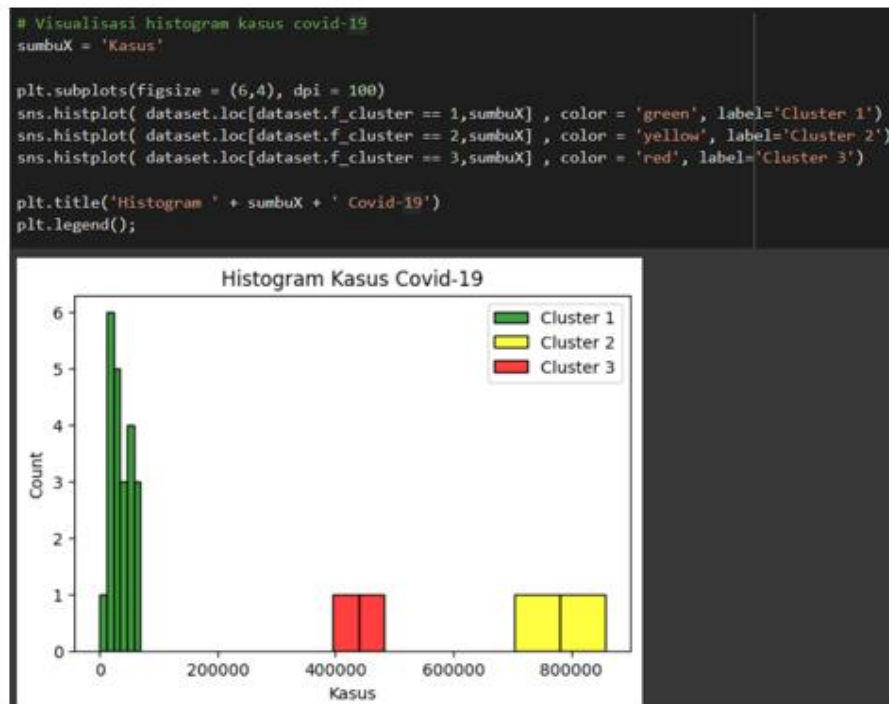
Gambar 5. Visualisasi hasil clusters

- f. Visualiasi histogram adalah hasil pola yang dihasilkan dengan algoritma K-Means dan bisa memberikan hasil pola cluster 1-cluster 4.



Gambar 6. Visualisasi histogram kasus Covid-19

- g. Eksport hasil output dihasilkan setelah mendapatkan hasil visualisasi histogram dengan beberapa cluster sebelumnya yang sudah terbentuk.



Gambar 7. Ekspor hasil output data Covid-19

Simpulan

Setelah melakukan pemrosesan menggunakan clustering K-MEANS dari data kasus covid tersebut, kami telah memperoleh data Pada iterasi 3, pengelompokan data yang dilakukan terhadap 4 cluster dengan iterasi 2 didapatkan hasil yang sama. Dari 34 data jumlah kasus penderita Covid-19 berdasarkan provinsi dapat diketahui, 2 provinsi cluster zona merah untuk penderita kasus Covid-19 terdiri dari provinsi DKI Jakarta dan Jawa Barat, 2 provinsi cluster zona orange terdiri dari provinsi Jawa Tengah dan dan Jawa Timur, 8 provinsi cluster zona kuning terdiri dari provinsi Bali, Banten, DI Yogyakarta, Kalimantan Timur, Sumatera Barat, Sumatera Utara, Sulawesi Selatan, dan Riau, dan 22 provinsi lainnya termasuk cluster zona hijau.

Daftar Pustaka

- Analisis Cluster untuk Pemetaan Data Kasus Covid-19 di Indonesia Menggunakan K-Means* (no date).
- Anggiat, Y. *et al.* (2022) 'Data Mining Clustering Penyebaran Virus Covid-19 Berdasarkan Provinsi di Indonesia Metode K-Means', *Jurnal Sistem Informasi Kaputama (JSIK)*, 6(2).
- Damanik, Y.F.S.Y. *et al.* (2021) 'Penerapan Data Mining Untuk Pengelompokan Penyebaran Covid-19 Di Sumatera Utara Menggunakan Algoritma K-Means', *Jurnal Ilmu Komputer dan Informatika*, 1(2). Available at: <https://doi.org/10.54082/jiki.13>.
- Desyanti, D. (2021) 'Implementasi Metode C.45 dalam Menganalisa Tingkat Kecemasan Mahasiswa Menyusun Tugas Akhir', *Jurnal Unitek*, 14(1), pp. 17–29. Available at: <https://doi.org/10.52072/unitek.v14i1.175>.
- Elsa Ramadanti and Muhamad Muslih (2021) 'Analisis Persebaran Kasus Covid-19 Di Jawa Barat Menggunakan Metode K-Means Clusterin | Seminar Nasional Sistem Informasi dan Manajemen Informatika Universitas Nusa Putra', *SISMATIK (Seminar Nasional Sistem Informasi dan Manajemen Informatika)*, pp. 319–326. Available at: <https://sismatik.nusaputra.ac.id/index.php/sismatik/article/view/41>.
- Febrina, Wetri (2021). Review: Perkembangan Teknologi Deteksi CoVid-19. *Jurnal UNITEK* 14(2) pp. 58-66. Available at: <https://doi.org/10.52072/unitek.v14i2.239>
- Hardiani, T. (2022) 'Analisis Clustering Kasus Covid 19 di Indonesia Menggunakan Algoritma K-Means', *Jurnal Nasional Pendidikan Teknik Informatika ...*, 11, pp. 156–165. Available at: <https://ejournal.undiksha.ac.id/index.php/janapati/article/view/45376%0Ahttps://ejournal.undiksha.ac.id/index.php/janapati/article/download/45376/22680>.
- Masriani, M., Alam, S. and Arham, A. (2022) 'Rancang Bangun Sistem Automatic Touchless Mask Machine Upaya Pengendalian Penggunaan Masker di Masa Pandemi Covid-19', *Jurnal Unitek*, 15(1), pp. 9–19. Available at: <https://doi.org/10.52072/unitek.v15i1.314>.

Sroyer, A.M., Mandowen, S.A. and Reba, F. (2022) 'Analisis Cluster Penyakit Malaria Provinsi Papua Menggunakan Metode Single Linkage Dan K-Means', *Jurnal Nasional Teknologi dan Sistem Informasi*, 7(3), pp. 147–154. Available at: <https://doi.org/10.25077/teknosi.v7i3.2021.147-154>.

Wahyuni, S. *et al.* (2021) 'Data Mining Prediksi Minat Customer Penjualan Handphone Dengan Algoritma Apriori', *Jurnal Unitek*, 14(2), pp. 10–19. Available at: <https://doi.org/10.52072/unitek.v14i2.243>.