

Penerapan Data Mining Menggunakan Metode Clustering untuk Menentukan Status Provinsi di Indonesia 2020

Ahmad Husain Ardiansyah¹, Wisnu Nugroho², Nurul Hanifatul Alfiyah³, Rahmat Aji Handoko⁴,
Muhammad Arfan Bakhtiar⁵

^{1,2,3,4,5}Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Magelang
E-mail: ¹ahmad.husain.ardiansyah11@gmail.com, ²wisnun0153@gmail.com,
³nurulhanifa4724@gmail.com, ⁴ajihndk@gmail.com, ⁵agr.melati@gmail.com

Abstrak – *Coronavirus is a group of viruses that can cause disease in animals or humans. Several types of coronaviruses are known to cause respiratory infections in humans ranging from cold coughs to more serious ones such as Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS). A new type of Coronavirus has been found to cause COVID-19. In Indonesia alone, this virus began to spread from the beginning of March until now. Many areas affected by this virus. There are also many regions where the level of outbreaks of this virus tends to be small. The choice of clustering method in the classification process to divide regions in Indonesia to be divided into red or yellow or green zones is based on calculations using the K-Mean clustering formula. K-Means Clustering algorithm is an algorithm that groups the same data in a certain group and different data in another group. The results of the grouping of regions included in the red zone are Jakarta with a number of more than 7,000 positive patients infected, for those entering the yellow zone namely Banten with more than 800 positive patients infected. And as many as less than 600 infected patients can be categorized as green zones.*

Kata Kunci — *tuliskan 3 Algorithm, Covid-19, Disease, K-Means Clustering*

1. PENDAHULUAN

World Health Organization (WHO) menjelaskan bahwa *Coronaviruses* (Cov) adalah sistem pernafasan yang terinfeksi oleh virus[1]. Infeksi virus ini disebut Covid19. Virus Corona menyebabkan penyakit flu biasa sampai penyakit yang lebih parah seperti Sindrom Pernafasan Timur Tengah (MERS-CoV) dan Sindrom Pernafasan Akut Parah (SARS-CoV). Virus Corona adalah *zoonotic* yang artinya ditularkan antara hewan dan manusia. Berdasarkan Kementerian Kesehatan Indonesia, perkembangan kasus *Covid-19* di Wuhan berawal pada tanggal 30 Desember 2019 dimana Wuhan *Municipal Health Committee* mengeluarkan pernyataan "*urgent notice on the treatment of pneumonia of unknown cause*". Penyebaran virus Corona ini sangat cepat bahkan sampai ke lintas negara. Sampai saat ini terdapat 212 negara yang mengkonfirmasi terkena virus Corona. Salah satunya di negara kita sendiri, Indonesia. Di Indonesia sendiri jumlah virus corona sudah mencapai 32 ribu lebih terjangkit virus ini. Penyebaran virus corona ini telah meluas ke berbagai daerah yang ada di Indonesia.

Covid-19 pertama masuk dilaporkan masuk ke Indonesia pada tanggal 2 Maret 2020 dengan jumlah 2 kasus. Kemudian dari data yang ada pada tanggal 31 Maret 2020 tercatat kasus virus ini meningkat sampai berjumlah 1.528 kasus terjangkit Covid-19. Tingkat mortalitas di Indonesia mencapai angka

tertinggi di Asia Tenggara dengan angka sebesar 8.9 %. Dengan angka tingkat mortalitas tertinggi di Asia Tenggara ini disebabkan karena angka kematian di Indonesia mencapai 136 jiwa. Selain itu jika ada yang terjangkit dan meninggal dunia pasti ada juga yang sembuh dari virus covid-19 ini, jumlah yang sembuh tercatat mencapai 81 orang.

Maksud kami mengambil judul ini karena melihat data- data banyaknya orang yang terjangkit, meninggal dunia, sampai sembuh yang terjadi di berbagai penjuru daerah di Indonesia. Untuk itu akan dibuat klasifikasi daerah dari zona merah, kuning, dan hijau yang berarti untuk zona merah sendiri berarti daerah itu adalah daerah bahaya dan daerah dengan jumlah terjangkit virus dengan jumlah cukup banyak. Untuk zona kuning berarti zona itu masuk di daerah dengan zona siaga, maksudnya adalah di daerah tersebut masuk daerah bisa dibilang dengan korban terjangkit virus Covid-19 dengan jumlah yang tidak banyak namun tidak juga sedikit. Yang terakhir zona hijau yang berarti zona aman, maksudnya di zona ini daerah yang terjangkit virus ini datanya paling sedikit terjangkit. Pengklasifikasian tersebut menggunakan metode *Clustering (K-Means)*.

2. METODE PENELITIAN

2.1 Data Mining

Menganalisis data yang prespektif dan menyimpulkan menjadi informasi penting yang dapat digunakan untuk meningkatkan

keuntungan, biaya pengeluaran diperkecil, atau juga bisa keduanya merupakan pengertian dari *data mining*[2]. *Data mining* adalah analisis pada data untuk mencari hubungan yang jelas dan menyimpulkan belum diketahui sebelumnya dengan cara terkini agar mudah dipahami dan bermanfaat[2].

2.2 Tahap-Tahapan dalam Data Mining

Ada tahap-tahapan dalam data mining, yang bersifat interaktif, pengguna langsung terlibat atau dengan perantara *knowledge base*. Berikut tahap-tahap dari *data mining*[2]:

- a. *Pembersihan data (Data Cleaning)*
Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan.
- b. *Integrasi data (Data Integration)*
Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru.
- c. *Seleksi data (Data Selection)*
Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database.
- d. *Transformasi data (Data Transformation)*
Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam Data Mining.
- e. *Proses Mining*
Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data. Beberapa metode yang dapat digunakan berdasarkan pengelompokan Data Mining.
- f. *Evaluasi pola (Pattern Evaluation)*
Untuk mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan.
- g. *Presentasi pengetahuan (Knowledge Presentation)*
Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

2.3 Metode Clustering

Clustering adalah memisahkan data ke dalam kelompok-kelompok yang memiliki obyek dengan karakteristik yang serupa[3]. Clustering ini merupakan awalan dari sebuah data mining.

Dengan menggunakan clustering ini, dapat mengklasifikasikan pasien Covid-19 dari tiap provinsi di Indonesia. Adapun tujuan dari cluster ini, yaitu dapat menentukan status pada setiap provinsi.

HASIL DAN PEMBAHASAN

3.1 Data Pengujian

Penggunaan data uji berupa table yang memiliki komponen table penyusun yang memiliki 4 atribut, yaitu nama provinsi di Indonesia dan status pasien Covid-19, diantaranya positif, sembuh, dan meninggal. Dimana dapat dilihat pada table 1.

Tabel 1. Jumlah Pasien Covid-19 Di setiap Provinsi Pada Bulan Mei

No	Provinsi	Positif	Sembuh	Meninggal
1	Aceh	19	13	1
2	Bali	440	304	4
3	Banten	807	243	65
4	Bangka Belitung	46	23	1
5	Bengkulu	80	16	2
6	DI Yogyakarta	222	143	8
7	DKI Jakarta	7072	1917	489
8	Jambi	92	11	0
9	Jawa Barat	2155	589	140
10	Jawa Tengah	1314	340	65
11	Jawa Timur	4651	574	365
12	Kalimantan Barat	182	50	4
13	Kalimantan Timur	250	162	3
14	Kalimantan Tengah	370	88	11
15	Kalimantan Selatan	827	98	77
16	Kalimantan Utara	161	63	2
17	Kepulauan Riau	190	98	15
18	NTB	562	215	10
19	Sumatra Selatan	946	184	33
20	Sumatra Barat	525	253	24
21	Sulawesi Utara	297	51	36
22	Sumatra Utara	382	118	40
23	Sulawesi Tenggara	218	110	3
24	Sulawesi Selatan	1415	564	71
25	Sulawesi Tengah	125	61	4
26	Lampung	126	68	9
27	Riau	112	81	6
28	Maluku Utara	147	26	11
29	Maluku	192	36	6
30	Papua Barat	159	51	2

31	Papua	632	68	6
32	Sulawesi Barat	91	38	2
33	NTT	92	14	1
34	Gorontalo	90	21	3
Jumlah		24989	6691	1519

3.2 Proses Clustering

Proses *clustering* untuk mengklasifikasikan status provinsi di Indonesia. Status provinsi di Indonesia dapat dilihat pada table Cluster.

- a. Cluster pertama diambil dari provinsi Aceh, Kepulauan Riau, dan Gorontalo. Maka didapat hasil pada table 2.

Tabel 2. Hasil dari Cluster pertama

No	Provinsi	C1	C2	C3
1	Aceh	1		
2	Bali		2	
3	Banten		3	
4	Bangka Belitung	4		
5	Bengkulu			5
6	DI Yogyakarta		6	
7	DKI Jakarta		7	
8	Jambi			8
9	Jawa Barat		9	
10	Jawa Tengah		10	
11	Jawa Timur		11	
12	Kalimantan Barat		12	
13	Kalimantan Timur		13	
14	Kalimantan Tengah		14	
15	Kalimantan Selatan		15	
16	Kalimantan Utara		16	
17	Kepulauan Riau		17	
18	NTB		18	
19	Sumatra Selatan		19	
20	Sumatra Barat		20	
21	Sulawesi Utara		21	
22	Sumatra Utara		22	
23	Sulawesi Tenggara		23	
24	Sulawesi Selatan		24	
25	Sulawesi Tengah			25
26	Lampung			26
27	Riau			27
28	Maluku Utara			28
29	Maluku		29	
30	Papua Barat		30	
31	Papua		31	

32	Sulawesi Barat			32
33	NTT			33
34	Gorontalo			34

3.3 Cluster kedua diambil dari hasil cluster pertama, dan hasil yang didapat seperti berikut,

- a. Cluster yang di dapat dari *clustering* pertama, dapat dilihat pada table 3.

Tabel 3. Cluster dari Clustering Pertama

C1	C2	C3	Kondisi
32,50	609,26	106,11	Positif
18,00	274,74	37,33	Sembuh
1,00	64,30	4,22	Meninggal

- b. Hasil dapat dilihat pada tabel 4.

Tabel 4. Hasil dari Cluster

No	Provinsi	C1	C2	C3
1	Aceh	1		
2	Bali		2	
3	Banten		3	
4	Bangka Belitung	4		
5	Bengkulu			5
6	DI Yogyakarta			6
7	DKI Jakarta		7	
8	Jambi			8
9	Jawa Barat		9	
10	Jawa Tengah		10	
11	Jawa Timur		11	
12	Kalimantan Barat			12
13	Kalimantan Timur			13
14	Kalimantan Tengah			14
15	Kalimantan Selatan		15	
16	Kalimantan Utara			16
17	Kepulauan Riau			17
18	NTB		18	
19	Sumatra Selatan		19	
20	Sumatra Barat		20	
21	Sulawesi Utara			21
22	Sumatra Utara		22	
23	Sulawesi Tenggara			23
24	Sulawesi Selatan		24	
25	Sulawesi Tengah			25
26	Lampung			26
27	Riau			27
28	Maluku Utara			28
29	Maluku			29
30	Papua Barat			30

31	Papua		31
32	Sulawesi Barat		32
33	NTT		33
34	Gorontalo		34

3.4 Cluster ketiga diambil dari hasil cluster kedua, dan hasil yang didapat seperti berikut,

a. Cluster yang di dapat dari *clustering* ketiga, dapat dilihat pada table 4.

Tabel 4. Cluster dari *Clustering* Pertama

c1	c2	c3	kondisi
32,50	1671,38	168,21	positif
18,00	420,54	62,53	sembuh
1,00	106,85	6,74	meninggal

b. Hasil dari cluster 3 dapat dilihat dari table 5

no	provinsi	c1	c2	c3
1	Aceh	1		
2	Bali			2
3	Banten			3
4	Bangka Belitung	4		
5	Bengkulu	5		
6	DI Yogyakarta			6
7	DKI Jakarta		7	
8	Jambi	8		
9	Jawa Barat		9	
10	Jawa Tengah		10	
11	Jawa Timur		11	
12	Kalimantan Barat			12
13	Kalimantan Timur			13
14	Kalimantan Tengah			14
15	Kalimantan Selatan			15
16	Kalimantan Utara			16
17	Kepulauan Riau			17
18	NTB			18
19	Sumatra Selatan		19	
20	Sumatra Barat			20
21	Sulawesi Utara			21
22	Sumatra Utara			22
23	Sulawesi Tenggara			23
24	Sulawesi Selatan		24	
25	Sulawesi Tengah			25
26	Lampung			26
27	Riau			27
28	Maluku Utara			28
29	Maluku			29
30	Papua Barat			30
31	Papua			31
32	Sulawesi Barat	32		
33	NTT	33		
34	Gorontalo	34		

3.5 Proses *clustering* dilakukan hingga mendapatkan hasil cluster yang tidak berubah. Dalam kasus ini,

didapat data cluster yang tidak berubah terjadi pada cluster ke 8 dan 9.

a. Cluster yang didapat dari data cluster 8 pada tabel 6.

Tabel 6. Cluster dari data cluster 8

C1	C2	C3	Kondisi
223,15	5861,50	1244,00	Positif
83,92	1245,50	336,33	Sembuh
8,23	427,00	75,17	Meninggal

b. Hasil dari cluster 8 dapat dilihat dari table 7.

Tabel 7. Hasil dari Cluster 8

No	Provinsi	C1	C2	C3
1	Aceh	1		
2	Bali	2		
3	Banten			3
4	Bangka Belitung	4		
5	Bengkulu	5		
6	DI Yogyakarta	6		
7	DKI Jakarta		7	
8	Jambi	8		
9	Jawa Barat			9
10	Jawa Tengah			10
11	Jawa Timur		11	
12	Kalimantan Barat	12		
13	Kalimantan Timur	13		
14	Kalimantan Tengah	14		
15	Kalimantan Selatan			15
16	Kalimantan Utara	16		
17	Kepulauan Riau	17		
18	NTB	18		
19	Sumatra Selatan			19
20	Sumatra Barat	20		
21	Sulawesi Utara	21		
22	Sumatra Utara	22		
23	Sulawesi Tenggara	23		
24	Sulawesi Selatan			24
25	Sulawesi Tengah	25		
26	Lampung	26		
27	Riau	27		
28	Maluku Utara	28		
29	Maluku	29		
30	Papua Barat	30		
31	Papua	31		
32	Sulawesi Barat	32		
33	NTT	33		

34	Gorontalo	34		
----	-----------	----	--	--

3.6 Dari semua proses clustering maka didapatkan bahwa : zona siaga (kuning), zona aman (hijau), dan zona berbahaya (merah). Tabel pembagian zona berdasarkan hasil *clustering* pada table 7.

Table 7. Hasil yang diperoleh dari *clustering*

No	Provinsi	Positif	Sembuh	Meninggal
1	Aceh	19	13	1
2	Bali	440	304	4
3	Banten	807	243	65
4	Bangka Belitung	46	23	1
5	Bengkulu	80	16	2
6	DI Yogyakarta	222	143	8
7	DKI Jakarta	7072	1917	489
8	Jambi	92	11	0
9	Jawa Barat	2155	589	140
10	Jawa Tengah	1314	340	65
11	Jawa Timur	4651	574	365
12	Kalimantan Barat	182	50	4
13	Kalimantan Timur	250	162	3
14	Kalimantan Tengah	370	88	11
15	Kalimantan Selatan	827	98	77
16	Kalimantan Utara	161	63	2
17	Kepulauan Riau	190	98	15
18	NTB	562	215	10
19	Sumatra Selatan	946	184	33
20	Sumatra Barat	525	253	24
21	Sulawesi Utara	297	51	36
22	Sumatra Utara	382	118	40
23	Sulawesi Tenggara	218	110	3
24	Sulawesi Selatan	1415	564	71
25	Sulawesi Tengah	125	61	4
26	Lampung	126	68	9
27	Riau	112	81	6
28	Maluku Utara	147	26	11
29	Maluku	192	36	6
30	Papua Barat	159	51	2
31	Papua	632	68	6
32	Sulawesi Barat	91	38	2

33	NTT	92	14	1
34	Gorontalo	90	21	3

3. SIMPULAN

Berdasarkan pengujian yang dilakukan menggunakan *Algoritma K-Means Clustering*, menyatakan bahwa:

1. Pusat cluster yang diperoleh untuk cluster pertama atau zona hijau berada pada provinsi Aceh, Bali, Bangka Belitung, Bengkulu, DIY, Jambi, Kalimantan Barat, Kalimantan Timur, Kalimantan Tengah, Kalimantan Utara, Kepulauan Riau, NTB, Sumatera Barat, Sumatera Utara, Sulawesi Tenggara, Sulawesi Tengah, Lampung, Riau, Maluku, Maluku Utara, Papua, Papua Barat, NTT, Gorontalo.
2. Pusat cluster yang diperoleh dari cluster kedua atau zona merah berada pada provinsi DKI Jakarta dan Jawa Timur.
3. Pusat cluster yang diperoleh dari cluster ketiga atau zona kuning berada pada provinsi Banten, Jawa Barat, Jawa Tengah, Kalimantan Selatan, Sumatera Selatan, Sulawesi Selatan.

4. SARAN

Adapun saran yang dapat diberikan mengenai penelitian tentang menentukan status provinsi di Indinonesia, dapat menentukan status provinsi dengan berbagai metode tidak hanya dengan *clustering*.

DAFTAR PUSTAKA

- [1] Pertanyaan jawaban terkait COVID-19 untuk publik. (n.d.).
- [2] Saleh, A. (2015). Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. 2(3), 207–217.
- [3] Method, I., Based, K. C., Value, S., Interface, W., Study, C., & Magelang, I. U. M. M. (2015). Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang. 18(1), 76–82.