

DATA MINING CLUSTERING PENYEBARAN VIRUS COVID-19 BERDASARKAN PROVINSI DI INDONESIA METODE K-MEANS

Yusuf Anggiat¹, Jehan Rizal Hermawan², Tri Cahyo Widiyanto³, Sita Anggraeni⁴
Sandra Dewi Saraswati⁵

Program Studi Informatika, Fakultas Teknologi Informasi, Universitas Nusa Mandiri,
Jalan Jatiwaringin Raya No 02 RT 08/13, Cipinang Melayu, Jakarta Timur, 021-28534236

e-mail: yusufanggiat@gmail.com¹, jehan70567@gmail.com², tricahyo229@gmail.com³,
sita.sia@nusamandiri.ac.id⁴, sandra.sww@nusamandiri.ac.id⁵

ABSTRACT

In February 2020, the earth was in an uproar with the circulation of the latest type of virus, namely COVID-19 or also known as the novel coronavirus. Artificial Intelligence (AI) can practice detecting outbreaks and speculate on the shoot effects of widespread disease. The data we use is secondary data obtained from the Ministry of Health of the Republic of Indonesia as of December 312020 on the Kemkes.go.id website. The method we use is K-Means clustering using the RapidMiner application with K-Means operator, cluster operator distance performance and validation. using the Davies Bouldin Index operator to find a value close to 0. the attributes used are positive cases, positive cases, and deaths by province. 30 provinces, cluster_1 (C2 Alert Zone) there are 3 provinces and cluster_2 (C3 Hazard Zone) there is only 1 province, namely DKI Jakarta. From a health perspective, it is very helpful to see which provinces should be prioritized in terms of health to reduce the level of health. positive cases and efforts to increase the recovery rate by implementing a prototype strict health column The results of the clusters that are formed can be developed by other methods, to obtain other information that has not been known or tried.

Keywords: *Clustering, K-Means, COVID-19*

ABSTRAK

Pada bulan Februari tahun 2020, bumi gempar dengan pedaran virus tipe terkini ialah COVID-19 ataupun dapat dikenal dengan novel coronavirus. *Artificial Intelligence* (AI) bisa berlatih mengetahui wabah serta berspekulasi efek pusat dari penyakit meluas. data yang kami gunakan adalah data sekunder yang diperoleh dari Kementrian Kesehatan Republik Indonesia per 31 Desember 2020 pada website Kemkes.go.id. metode yang kami gunakan adalah *Clustering* menggunakan *K-Means* menggunakan tools *RapidMiner* dengan operator *K-Means*, operator *cluster distance performance* dan validasinya menggunakan operator *Davies Bouldin*

Index untuk mencari nilai yang mendekati 0. atribut yang digunakan adalah kasus positif, kasus positif, dan meninggal berdasarkan provinsi. Berdasarkan pengujian hasil dari validasi *Davies Bouldin Index*, terdapat nilai 0.139 (optimal) dengan kategori *cluster_0* (C1=Zona Aman) terdapat 30 provinsi, *cluster_1* (C2=Zona Waspada) terdapat 3 provinsi dan *cluster_2* (C3=Zona Bahaya) terdapat hanya 1 provinsi, yaitu DKI Jakarta. Dampak dari segi kesehatan, ini sangat membantu untuk melihat provinsi mana saja yang harus di prioritaskan dalam hal kesehatan dalam upaya menurunkan tingkat kasus positif dan upaya meningkatkan angka sembuh dengan menerapkan protokol kesehatan yang ketat Hasil *cluster* yang terbentuk dapat dikembangkan dengan metode lain, supaya mendapatkan informasi lain yang belum diketahui atau dicoba.

Kata Kunci: Clustering, K-Means, COVID-19

1. Pendahuluan

Karakter dan konsep *data science* di Informatika 4.0 menjelang pertengahan tahun 2022 sangat diperlukan dan mewariskan efek yang absah dalam berkehidupan keanekaragaman di Indonesia maupun universal. Teknik informatika yang sekarang adalah penerapan perubahan mutlak dari pemikiran para ahli di lingkup elektronika dan kemajuan informatika. Prinsip Teknik informatika adalah kenyataan yang ditemukan dari perniagaan modernisasi, karena terobosan dan inovatif kemajuan informasi mengangkat peranan penting dalam setiap organisasi maupun individu. Dalam hal ini terdapat jenis plural teknologi yang akan menjadi peran istimewa dalam perubahan kemajuan informatika 4.0, yaitu *Big Data*, *Internet of Things*, *Cloud Computing*, *Addictive Manufacturing*, dan juga yang terpenting yaitu *Artificial Intelligence*.

Artificial Intelligence ialah kemampuan yang terbuat kedalam suatu *system* yang dapat diolah dalam peran objektif tetapi dapat di pula *Artificial Intelligence* ataupun cuma disingkat AI, didefinisikan selaku intelek entitas objektif. Andreas Kaplan serta Michael Haenlein memaknakan kemampuan ciptaan selaku “kehandalan sistem buat memaknakan keluaran informasi dengan betul, buat berlatih

dari informasi itu, serta memakai penataran itu untuk menggapai tujuan serta kewajiban khusus lewat menyesuaikan diri yang tidak kompleks”[1].

Machine Learning adalah bentuk salah satu cabang dari ilmu keahlian buatan, uniknya yang terfokus meninjau cara kinerja komputer mampu berkembang dari data untuk meningkatkan keahlian dan mempraktikannya[2].

Pada bulan Februari tahun 2020, bumi gempar dengan pedaran *virus* tipe terkini ialah COVID-19 ataupun dapat dikenal dengan *novel coronavirus*. *Coronavirus* berjalan meluas lewat droplets ataupun tetes larutan yang mulanya diawali dari batuk berdarah serta bersin, kontak individu semacam memegang serta bersalaman tangan, memegang barang ataupun elemen yang ada *virus* setelah itu memegang ataupun bersentuhan dengan mulut, hidung, ataupun mata saat sebelum mensterilkan zona tangan[3].

BlueDot ialah salah satu ilustrasi serta sesuatu startup yang berawal dari Kanada yang sediakan layanan berplatform *Artificial Intelligence* (AI) buat berikan peringatan bahaya penyebaran wabah virus COVID-19 sebagian hari saat sebelum pusat pengaturan penangkalan penyakit ataupun badan

kesehatan dunia menghasilkan peringatan pada warga di semua penduduk dunia.

Berdasarkan latar belakang diatas maka penulis mengambil judul sebagai berikut “Clustering Menggunakan K-Means Studi Kasus Penyebaran Virus COVID-19 Berdasarkan Provinsi Di Indonesia”.

2. Metodologi

Pengumpulan data dilakukan untuk mendukung penelitian berdasarkan latar belakang masalah dan kebutuhan penelitian. Adapun jenis data yang digunakan yakni :

1. Data Primer (Private)

Data yang kami gunakan dalam penelitian ilmiah ini adalah data-data kutipan yang kami dapatkan dari Google.Scholar.com dan repository.nusamandiri.ac.id, untuk sumber pengertian atau istilah-istilah dalam penelitian ilmiah kami.

2. Data Sekunder (Public)

Data yang kami gunakan untuk penelitian ilmiah ini adalah data sekunder (public), penulis mendapatkan data tersebut dari website kemkes.go.id terdapat 34 Record yaitu seluruh jumlah provinsi yang ada di Indonesia. Adapun pengumpulan data melalui kegiatan dibawah ini:

1. Observasi

Metode Observasi adalah metode pengumpulan data maupun informasi melalui beberapa sumber terpercaya dengan cara menelaah maupun mempelajari dari buku, website, maupun terjun langsung kelapangan untuk mengetahui informasi-informasi yang lebih akurat. Observasi yang telah dilakukan pada penulisan ini adalah observasi melalui mempelajari artikel-artikel terkait dengan pengertian metode K-Means Clustering, selain itu juga ada pengetahuan tentang COVID-19 di Indonesia, dan pengetahuan tentang RapidMiner.

2. Studi Pustaka

Studi Pustaka yaitu dasar teori yang digunakan dalam penelitian untuk menyelesaikan masalah permasalahan dan merupakan referensi yang kuat untuk bahan penilitian.

2.1 Data Mining

Data mining ialah suatu susunan cara buat mencari angka imbuh dari kumpulan-

kumpulan informasi yang berbentuk wawasan serta data yang sepanjang ini tidak dikenal dengan cara buku petunjuk [4]

Data mining buat menjabarkan wawasan apa yang dikira sesuai dengan detail dimensi serta batasan, memakai *database* bersama dengan *preprocessing* yang dibutuhkan, pengumpulan ilustrasi serta alih bentuk dari *database*. KDD ialah salah satu dari metode yang amat populer yang berpusat pada cara dengan cara totalitas cara temuan ilmu atau *knowledge* dari informasi, tercantum cara penyimpanan serta pengaksesan informasi, algoritma yang efisien serta sempurna bila berkaitan dengan informasi yang besar, pemahaman serta cerminan dari hasil penemuan wawasan [5].

2.2 Clustering

Clustering merupakan alur proses tempat satu set subjek informasi ke dalam himpunan bagian yang disebut dengan klaster. Ikon ataupun subjek yang di dalam klaster mempunyai kesamaan karakter antar satu serupa yang lain serta berlainan dengan klaster yang lain. Partisi tidak dicoba dengan cara buku petunjuk melainkan dengan sesuatu algoritma klaster. Oleh sebab itu, klastering amat berjasa serta dapat menciptakan tim ataupun golongan yang tidak diketahui serta belum dikenal dalam informasi. klastering banyak dipakai dalam bermacam aplikasi ataupun *web* semacam misalnya pada *business intelligence*, identifikasi pola pandangan, mesin pencarian, aspek ilmu kesehatan, serta buat keamanan (*security*). Di dalam *business intelligence*, klaster dapat menata banyak informasi atau data ke dalam jumlah golongan[4].

2.3 K-Means

Dalam menyokong cara pengerjaan informasi klastering, kita selaku pengarang memakai *K-Means*. *K-means* ialah tata cara analisis golongan yang membidik pada pemartisipasian N subjek observasi kedalam K, golongan (*cluster*) dimana tiap subjek observasi dipunyai oleh suatu golongan dengan *means* (rata-rata) terdekat, mendekati

dengan algoritma *expectation maximization* buat *Gausstion Mixture*. Dimana keduanya berupaya buat menciptakan pusat dari golongan dalam informasi sebesar listerasi koreksi yang dicoba kedua algoritma [4].

K- Means merupakan salah satu tata cara pengelompokan informasi non jenjang (penyekat) yang berupaya memilah informasi yang terdapat ke dalam wujud 2 ataupun lebih golongan. Tata cara ini memilah informasi ke dalam golongan sambungan, alhasil informasi berkarakteristik serupa dimasukkan ke dalam satu golongan yang serupa serta informasi yang berkarakteristik berlainan dikelompokkan kedalam golongan yang berlainan ataupun sambungan [6].

2.4 RapidMiner

RapidMiner ialah perangkat lunak yang bertabiat terbuka buat *public* yang bisa di unduh dengan cara *free* pula. *RapidMiner* merupakan suatu pemecahan buat melaksanakan analisa kepada *data mining*, *text mining* serta aplikasi perkiraan. *RapidMiner* memakai bermacam metode deskriptif serta perkiraan dalam membagikan pengetahuan pada konsumennya alhasil bisa membuat ketetapan yang sangat pas serta cermat[7].

2.5 Davis Bouldin Index

Davies bouldin index (DBI) merupakan metode buat menilai ataupun memikirkan hasil algoritma klastering. Awal kali dipublikasikan oleh David L. Davies serta Donald W. Bouldin pada tahun 1979. Dengan memakai DBI sesuatu klaster hendak dikira mempunyai desain klastering yang maksimal merupakan yang mempunyai nilai mendekati nol (0)[8].

3. Hasil dan Pembahasan

Pada tahap ini, setelah pengumpulan data telah dilakukan dan didapatkan 34 *record* yaitu jumlah provinsi di Indonesia dengan atribut positif, sembuh dan meninggal. Dataset didapatkan di Pusat Data Kesehatan Indonesia tahun 2020 (per 31 Desember 2020).

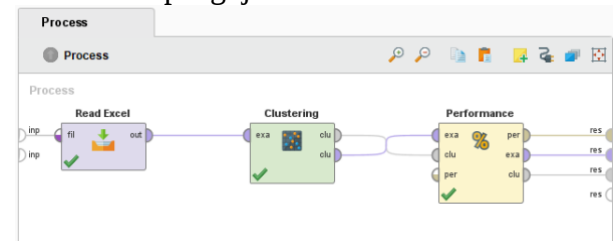
Tabel II.2
TABEL DATASET COVID-19 SETELAH DI PREPROCESSING

Provinsi	Positif	Sembuh	Meninggal
Aceh	8.746	7.149	358
Sumatera Utara	18.149	15.410	679
Sumatera Barat	23.464	20.364	504
Riau	24.966	23.104	583
Jambi	3.227	2.417	55
Sumatera Selatan	11.826	9.515	611
Bengkulu	3.603	2.597	117
Lampung	6.276	4.317	274
Kepulauan Bangka Belitung	2.337	1.717	35
Kepulauan Riau	6.995	6.075	172
DKI Jakarta	183.735	164.787	3.270
Jawa Barat	83.579	70.788	1.172
Jawa Tengah	81.716	54.409	3.562
Daerah Istimewa Yogyakarta	12.155	8.175	260
Jawa Timur	84.152	72.135	5.827
Banten	18.170	10.397	425
Bali	17.593	16.031	519
Nusa Tenggara Barat	5.664	4.551	273
Nusa Tenggara Timur	2.167	1.202	50
Kalimantan Barat	3.118	2.742	27
Kalimantan Tengah	9.740	7.476	269
Kalimantan Selatan	15.303	13.806	585
Kalimantan Timur	27.076	22.691	743
Kalimantan Utara	3.794	2.290	57
Sulawesi Utara	9.671	7.059	310
Sulawesi Tengah	3.552	1.865	106
Sulawesi Selatan	31.047	26.816	594
Sulawesi Tenggara	7.097	6.986	149
Gorontalo	3.841	3.308	104
Sulawesi Barat	1.941	1.552	34
Maluku	5.722	4.481	79
Maluku Utara	2.771	2.353	89
Papua Barat	5.979	5.444	99
Papua	13.216	7.088	147

Gambar 1. Dataset

Data yang telah di dapatkan lalu di preprocessing dengan tidak memasukan *recovery case*, *case fatality rate*, dan nomor supaya pengolahan menggunakan *tools RapidMiner* berjalan dengan sesuai keinginan penulis.

Data yang telah tersusun dimasukkan ke dalam aplikasi *RapidMiner* untuk melakukan pengujian.



Gambar 2. Clustering K-Means dengan tools *RapidMiner*

Berdasarkan evaluasi hasil dari pengujian dan eksperimen model, didapatkan perbandingan nilai Davies Bouldin Index yakni 0,139 (optimal). Berdasarkan hal tersebut jumlah cluster yang digunakan pada penelitian ilmiah ini adalah 3 Parameter cluster.

Row No.	Id	Provinsi	cluster	Positif	Sembuh	Meninggal
1	1	Aceh	cluster_0	8746	7149	358
2	2	Sumatera Ut...	cluster_0	18149	15410	679
3	3	Sumatera Ba...	cluster_0	23464	20364	504
4	4	Riau	cluster_0	24966	23104	583
5	5	Jambi	cluster_0	3227	2417	55
6	6	Sumatera Sel...	cluster_0	11826	9515	611
7	7	Bengkulu	cluster_0	3603	2597	117
8	8	Lampung	cluster_0	6276	4317	274
9	9	Kepulauan B...	cluster_0	2337	1717	35
10	10	Kepulauan Ri...	cluster_0	6995	6075	172

Gambar 3. ExampleSet Result

Pada gambar 3. menampilkan hasil dari clustering data, label cluster terbagi menjadi 3 kelompok yaitu cluster_0, cluster_1, cluster_2. Pembagian ini berdasarkan hasil dari centroid tiap masing-masing data dengan jarak terdekat.

Name	Type	Missing	Statistics
Provinsi	Polynomial	0	Label: Sumatera Utara (1), Aceh (1), Aceh (1), Bali (1), ... (32 more)
cluster	Normal	0	Label: cluster_0 (1), cluster_0 (30), cluster_0 (30), cluster_1 (3), ... (1)
Positif	Integer	0	Min: 1841, Max: 183735, Average: 21834.941
Sembuh	Integer	0	Min: 1202, Max: 164787, Average: 17973.441
Meninggal	Integer	0	Min: 27, Max: 5827, Average: 651.118

Gambar 4. ExampleSet Static

ExampleSet Static menampilkan hasil statistic dari proses *clustering K-Means*, terdapat juga pada atribut yang menampilkan hasil maksimal dan minimal dan rata-rata nilai dari masing-masing atribut tersebut.

Cluster Model

```
Cluster 0: 2 items
Cluster 1: 2 items
Cluster 2: 1 items
Cluster 3: 1 items
Cluster 4: 4 items
Total number of items: 10
```

Gambar 5. Cluster Model

Pada gambar 5. bisa dilihat hasil pembagian data terhadap tiap cluster, *cluster 0* terdapat 30 provinsi, *cluster 1* terdapat 3 provinsi dan *cluster 2* terdapat 1 provinsi. Total dari *dataset* yang diuji adalah 34 provinsi

Atribut	cluster_0	cluster_1	cluster_2
Positif	10306.867	83149	183735
Sembuh	8299.267	65777.333	164787
Meninggal	276.900	3520.333	3279

Gambar 6. Centroid Tabel

Centroid Table menampilkan nilai centroid dari masing-masing atribut pada tiap cluster. Nilai tersebut menjadi pusat

perhitungan dataset dengan cara menghitung dan mengukur kedekatan nilai dengan masing-masing titik pusat cluster.

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 107676653.981
Avg. within centroid distance_cluster_0: 115071967.068
Avg. within centroid distance_cluster_1: 69615741.111
Avg. within centroid distance_cluster_2: 0.000
Davies Bouldin: 0.139
```

Gambar 7. Performance Vector

Performance Vector merupakan detail hasil dari proses *perhitungan cluster distance performance* pada *tools RapidMiner* yang menampilkan jarak titik pusat pada tiap-tiap *cluster* dan juga menampilkan nilai *Davies Bouldin*.

Hasil centroid pada masing-masing atribut terdapat pada table 1.

Tabel 1. Hasil Centroid Parameter 3 Cluster

Attribut	Cluster_0	Cluster_1	Cluster_2
Positif	10306,867	83149	183735
Sembuh	8299,267	65777,33	164787
Meninggal	276,900	3520,333	3270

Hasil dan Pembahasan

Berdasarkan hasil yang telah diuji pada proses *clustering K-Means*, *cluster_0* terdapat 30 provinsi yakni klaster rendah, *cluster_1* terdapat 3 provinsi yakni klaster waspada, *cluster_2* terdapat 1 provinsi yakni klaster tinggi. Berikut ExampleSet dari hasil pengujian Clustering K-Means.

Tabel 2. ExampleSet Hasil Pengujian Clustering

Provinsi	Positif	Sembuh	Meninggal	Cluster
Aceh	8746	7149	358	cluster_0
Sumatera Utara	18149	15410	679	cluster_0
Sumatera Barat	23464	20364	504	cluster_0
Riau	24966	23104	583	cluster_0

Jambi	3227	2417	55	cluster_0
Sumatera Selatan	11826	9515	611	cluster_0
Bengkulu	3603	2597	117	cluster_0
Lampung	6276	4317	274	cluster_0
Kepulauan Bangka Belitung	2337	1717	35	cluster_0
Kepulauan Riau	6995	6075	172	cluster_0
DKI Jakarta	183735	164787	3270	cluster_2
Jawa Barat	83579	70788	1172	cluster_1
Jawa Tengah	81716	54409	3562	cluster_1
Daerah Istimewa Yogyakarta	12155	8175	260	cluster_0
Jawa Timur	84152	72135	5827	cluster_1
Banten	18170	10397	425	cluster_0
Bali	17593	16031	519	cluster_0
Nusa Tenggara Barat	5664	4551	273	cluster_0
Nusa Tenggara Timur	2167	1202	50	cluster_0
Kalimantan Barat	3118	2742	27	cluster_0
Kalimantan Tengah	9740	7476	269	cluster_0
Kalimantan Selatan	15303	13806	585	cluster_0
Kalimantan Timur	27076	22691	743	cluster_0
Kalimantan Utara	3794	2290	57	cluster_0
Sulawesi Utara	9671	7059	310	cluster_0
Sulawesi Tengah	3552	1865	106	cluster_0
Sulawesi Selatan	31047	26816	594	cluster_0
Sulawesi Tenggara	7097	6986	149	cluster_0
Gorontalo	3841	3308	104	cluster_0
Sulawesi Barat	1941	1552	34	cluster_0
Maluku	5722	4481	79	cluster_0
Maluku Utara	2771	2353	89	cluster_0
Papua Barat	5979	5444	99	cluster_0
Papua	13216	7088	147	Cluster_0

Deskripsi Hasil Clustering

Tabel 3. Deskripsi Hasil Clustering

Klaster	Status	Zona	Jumlah
---------	--------	------	--------

	Klaster		
Cluster_0	Rendah	Aman	30 Provinsi
Cluster_1	Waspada	Waspada	3 Provinsi
Cluster_2	Tinggi	Bahaya	1 Provinsi

Proses *clustering K-Means* ini menghasilkan tiga *cluster* yakni, *cluster_0* terdapat 30 provinsi, *cluster_1* terdapat 3 provinsi dan *cluster_2* terdapat 1 provinsi. Dimana *cluster_0* (c1) mempunyai nilai *centroid* positif yakni 10306,867, sembuh 8299,267 dan meninggal 276,900 dengan status klaster rendah yang berzona aman. Sementara itu *cluster_1* (c2) mempunyai nilai *centroid* positif yakni 83149, sembuh 65777,333 dan meninggal 3520,333 dengan status klaster waspada yang berzona waspada. Dan *cluster_2* (c3) mempunyai nilai *centroid* positif yakni 183735, sembuh 164787, dan meninggal 3270 dengan status klaster tinggi yang berzona bahaya.

Cluster_0

Pada *cluster_0* (c1) yakni klaster rendah yang berzona aman didapati sebanyak 30 provinsi, yang bisa dianggap masih dalam keadaan aman dan tetap menerapkan protokol Kesehatan.

Tabel 4. Cluster_0 (c1)

Provinsi	Positif	Sembuh	Meninggal	Cluster_0
Aceh	8746	7149	358	cluster_0
Sumatera Utara	8149	15410	679	cluster_0
Sumatera Barat	23464	20364	504	cluster_0
Riau	24966	23104	583	cluster_0
Jambi	3227	2417	55	cluster_0
Sumatera Selatan	1826	9515	611	cluster_0
Bengkulu	3603	2597	117	cluster_0
Lampung	6276	4317	274	cluster_0
Kepulauan Bangka Belitung	2337	1717	35	cluster_0
Kepulauan Riau	6995	6075	172	cluster_0
Daerah Istimewa Yogyakarta	12155	8175	260	cluster_0
Banten	8170	10397	425	cluster_0
Bali	17593	16031	519	cluster_0

Nusa Tenggara Barat	5664	4551	273	cluster_0
Nusa Tenggara Timur	2167	1202	50	cluster_0
Kalimantan Barat	3118	2742	27	cluster_0
Kalimantan Tengah	9740	7476	269	cluster_0
Kalimantan Selatan	5303	3806 ¹	85	cluster_0
Kalimantan Timur	27076	22691	743	cluster_0
Kalimantan Utara	3794	2290	57	cluster_0
Sulawesi Utara	9671	7059	310	cluster_0
Sulawesi Tengah	3552	1865	106	cluster_0
Sulawesi Selatan	31047	26816	59	cluster_0
Sulawesi Tenggara	7097	6986	149	cluster_0
Gorontalo	3841	3308	104	cluster_0
Sulawesi Barat	1941	1552	34	cluster_0
Maluku	5722	4481	79	cluster_0
Maluku Utara	2771	2353	89	cluster_0
Papua Barat	597	5444	99	cluster_0
Papua	3216	7088	147	cluster_0

Cluster_1

Pada *cluster_1* (c2) yakni klaster waspada yang berzona waspada didapati sebanyak 3 provinsi. kepada masyarakat yang tinggal di provinsi Jawa Barat, Jawa Tengah dan Jawa Timur harus berwaspada karena telah memenuhi kriteria waspada yang harus meminimalisir kontak langsung dengan orang, terlebih sangat wajib menerapkan protokol Kesehatan.

Tabel 5.0 Cluster_1 (c2)

Provinsi	Positif	Sembuh	Meninggal	Cluster
Jawa Barat	83579	70788	1172	cluster_1
Jawa Tengah	81716	54409	3562	cluster_1
Jawa Timur	84152	72135	5827	cluster_1

Cluster_2

Pada *cluster_2* (c3) yakni klaster tinggi yang berzona bahaya didapati hanya 1 provinsi. DKI Jakarta jadi satu-satunya

provinsi dengan tingkat klaster yang berbahaya, semua elemen masyarakat DKI Jakarta harus bekerja sama untuk menurunkan tingkat kasus positif dan menaikkan angka jumlah kesembuhan dari *virus COVID-19* dengan menerapkan protokol kesehatan yang sangat ketat, seperti sekolah dan kuliah melalui daring (*Google Meet* dan *Zoom*) dan bekerja dari rumah (*Work From Home*).

Tabel 6.0 Cluster_2 (C3)

Provinsi	Positif	Sembuh	Meninggal	Cluster
DKI Jakarta	183735	164787	3270	cluster_2

Akurasi Davies Bouldin

Hasil uji *performance* dilakukan menggunakan operator *cluster distance performance*. Operator ini digunakan untuk mengevaluasi hasil kinerja metode clustering berbasis *centroid*. Operator ini juga memberikan daftar nilai kriteria kinerja berdasarkan *centroid*. *Cluster distance performance* yang dimaksud adalah *performance Davies Bouldin Index (DBI)*. Dimana pada penelitian ini nilai *Davies Bouldin Index* sangat optimal yakni 0.139.

4. Kesimpulan

Berdasarkan penelitian yang sudah dilakukan maka dapat menyimpulkan sebagai berikut:

1. Hasil *Clustering K-Means* Studi Kasus Penyebaran Virus COVID-19 Berdasarkan Provinsi Di Indonesia menggunakan tools *RapidMiner*, menghasilkan tiga *cluster* yaitu *cluster_0* (c1) terdapat 30 provinsi dengan *cluster* rendah dengan zona aman, *cluster_1* (c2) terdapat 3 provinsi dengan *cluster* waspada dengan zona waspada, dan *cluster_2* (c3) terdapat 1 provinsi dengan *cluster* tinggi dengan zona bahaya.
2. *Cluster_0* (c1) terdapat 30 provinsi yaitu: Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan

Bangka Belitung, Kepulauan Riau, Daerah Istimewa Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, dan Papua dengan klaster rendah yang berzona aman.

3. *Cluster_1* (c2) terdapat 3 provinsi yaitu: Jawa Barat, Jawa Tengah dan Jawa Timur dengan klaster waspada yang berzona waspada.
4. DKI Jakarta menempati klaster tertinggi (c3) dengan zona bahaya dengan kasus positif sebanyak 183735 orang, sembuh sebanyak 164787 orang, dan meninggal terdapat 3270 orang.
5. Hasil pengujian *Davies Bouldin Index* mendapatkan nilai sebesar 0.139 yang berarti memiliki jarak antar provinsi *cluster* yang cukup baik karena mendekati angka nol (0).

Hasil dari penelitian ini dapat menjadi bahan referensi serta acuan terhadap *cluster* penyebaran COVID-19 di Indonesia. Kekurangan dalam penelitian ini adalah kasus penyebaran COVID-19 terus berlanjut sampai saat ini saat kami menulis penelitian ilmiah ini, bisa saja kasus penyebaran semakin meningkat atau bahkan menurun karena kekebalan manusia sudah teratasi oleh vaksin COVID-19, sehingga perlu penerapan atau implementasi pada aplikasi lain untuk mencari hasil yang sama dan juga yang terbaik.

5. Saran

Berdasarkan penelitian yang telah diuji dan dibuat, masih banyaknya hal yang belum dapat diterapkan dan di implementasikan dari penelitian ini, maka penulis mempertimbangkan beberapa saran sebagai berikut:

1. Hasil *cluster* yang terbentuk dapat dikembangkan dengan metode lain, supaya mendapatkan informasi lain yang belum diketahui atau dicoba.
2. Penelitian ini bisa dikembangkan lagi menjadi induk atau basis pengetahuan untuk sistem keputusan penerapan COVID-19 berdasarkan provinsi di Indonesia.

Daftar Pustaka

- [1] M. Siahaan, C. H. Jasa, K. Anderson, and M. Valentino, "Penerapan Artificial Intelligence (AI) Terhadap Seorang Penyandang Disabilitas Tunanetra," *Inf. Syst. Technol.*, vol. 01, no. 02, pp. 186–193, 2020.
- [2] T. Wahyono, "Fundamental of Python for Machine Learning: Dasar-Dasar Pemrograman Python untuk Machine Learning dan Kecerdasan Buatan," *Gava Media*, no. September 2018, p. 49, 2018.
- [3] P. C.- Di *et al.*, "Komunikasi Informasi Dan Edukasi (KIE) Vaksinasi Dan Pencegahan".
- [4] K. Annisa, B. S. Ginting, and M. A. Syari, "Penerapan Data Mining Pengelompokan Data Pengguna Air Bersih Berdasarkan Keluhannya Menggunakan Metode Clustering Pada PDAM Langkat," vol. 6341, no. April, 2022.
- [5] T. Kenikan, K. Di, S. D. N. Citamiang, and K. Sukabumi, "Penerapan algoritma c.45 untuk klasifikasi tingkat kenikan kelas di sdn citamiang 2 kota sukabumi," 2020, [Online]. Available: https://repository.bsi.ac.id/index.php/unduh/item/302730/19162343_Aris-Setiawan-new.pdf
- [6] et al Andy Febrianto, Achmadi Sentot, Agung Panji Sasmito, "Penerapan Metode K-Means Untuk Clustering Pengunjung Perpustakaan Itn Malang," *J. Mhs. Tek. Inform.*, vol. 5, no. 1, pp. 61–70, 2021.

- [7] D. Ardiansyah and W. Walim, "Algoritma c4.5 untuk klasifikasi calon peserta lomba cerdas cermat siswa smp dengan menggunakan aplikasi rapid miner," *J. Inkofar*, vol. 1, no. 2, pp. 5–12, 2018.
- [8] S. Butsianto and N. Saepudin, "Penerapan Data Mining Terhadap Minat Siswa Dalam Mata Pelajaran Matematika Dengan Metode K-Means," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 3, no. 1, pp. 51–59, 2020, doi: 10.32672/jnkti.v3i1.2008.