Bagian 1: Kombinasi Metode Klastering dan Klasifikasi (Kasus Pandemi Covid-19 di Indonesia)

Agus Perdana Windarto^{1,*}, Ulfah Indriani², Mokhammad Ramdhani Raharjo³, Linda Sari Dewi⁴

¹Program Studi Sistem Informasi, STIKOM Tunas Bangsa, Pematangsiantar, Sumatera Utara, Indonesia

²Mahasiswa program studi Doktor (S3) Teknologi Informasi, Universitas Putra Indonesia "YPTK" Padang, Indonesia

³Program Studi Sistem Informasi, Universitas Potensi Utama, Medan – Sumatera Utara, Indonesia

⁴Program Studi Sistem Informasi, Universitas Islam Kalimantan Muhammad Arsyad Al Banjari Banjarmasin, Indonesia

⁵ Program Studi Sistem Informasi, STMIK Nusa Mandiri, Jakarta Pusat, Indonesia

Email: ^{1,2,*}agus.perdana@amiktunasbangsa.ac.id, ³batakeraton@gmail.com, ⁴ramdhani@uniska-bjm.ac.id,

⁵linda.lrw@nusamandiri.ac.id

Email Penulis Korespondensi: agus.perdana@amiktunasbangsa.ac.id

Abstrak—Tujuan dari penelitian untuk mengkombinasi metode klastering dan klasifikasi yang merupakan bagian dari data mining. Kasus yang diangkat adalah jumlah persebaran pandemi Covid-19 di Indonesia per tanggal 7 juli 2020 yang terdiri 34 *record*. Sumber data diperoleh dari Data Kementerian Kesehatan, dicuplik dan diolah dari covid19.go.id dan bnpb.go.id. Variabel yang digunakan pada penelitian adalah jumlah kasus positif (x_1), jumlah kasus sembuh (x_2) dan jumlah kasus meninggal dunia (x_3) berdasarkan provinsi. Metode klastering dan klasifikasi yang digunakan adalah k-medoids dan C4.5. Metode k-medoids berfungsi untuk melakukan pemetaan berupa kluster terhadap wilayah di Indonesia berdasarkan provinsi. Label pemetaan yang digunakan sebanyak 3 kluster yakni: kluster tinggi (C_1 = zona merah), kluster waspada (C_2 = zona kuning), kluster rendah (C_3 = zona hijau). Hasil dari pemetaan diteruskan dengan menggunakan metode C4.5 untuk melihat aturan berupa pohon keputusan. Proses analisa dibantu dengan *software* RapidMiner. Penentuan jumlah kluster (k) ditentukan dengan menggunakan parameter *Davies Bouldin Index* (DBI) untuk mengoptimalkan hasil kluster yang diperoleh. Untuk $k=3$ memiliki nilai optimal yakni 0.740. Hasil pemetaan diperoleh 9 provinsi berada di kluster tinggi (C_1 = zona merah), 3 provinsi berada di kluster waspada (C_2 = zona kuning) dan 22 provinsi berada di kluster rendah (C_3 = zona hijau). Nilai yang diperoleh dari pohon keputusan untuk kluster tinggi (C_1 = zona merah) berdasarkan C4.5 adalah jika jumlah kasus positif lebih kecil dari 9524 dan lebih besar dari 4329 ($4329 > x_1 < 9524$). Sembilan provinsi yang masuk kluster tinggi (C_1 = zona merah) adalah Aceh, Bali, DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur, Kalimantan Selatan, Sumatera Selatan dan Sulawesi Selatan. Hasil dari kombinasi metode tersebut dapat diterapkan dan memberikan pengetahuan berupa informasi baru tentang pemetaan berupa kluster terhadap jumlah persebaran pandemi Covid-19 di Indonesia.

Kata Kunci: Data Mining, Klastering, Klasifikasi, Covid-19, Indonesia

Abstract—The purpose of this research is to combine the classification and classification methods that are part of data mining. The case raised was the number of the spread of the Covid-19 pandemic in Indonesia as of July 7, 2020 with 34 records. Data sources were obtained from Ministry of Health Data, sampled and processed from covid19.go.id and bnpb.go.id. The variables used in the study are the number of positive cases (x_1), number of cases cured (x_2) and number of deaths (x_3) by province. The classification and classification methods used are k-medoids and C4.5. The k-medoids method works to map clusters of regions in Indonesia by province. The mapping labels used are 3 clusters: high cluster (C_1 = red zone), alert cluster (C_2 = yellow zone), low cluster (C_3 = green zone). The results of the mapping are continued using the C4.5 method to see the rules in the form of a decision tree. The analysis process is assisted with the RapidMiner software. Determination of the number of clusters (k) is determined by using the Davies Bouldin Index (DBI) parameter to optimize the cluster results obtained. For $k = 3$ has an optimal value of 0.740. The mapping results obtained 9 provinces are in the high cluster (C_1 = red zone), 3 provinces are in the alert cluster (C_2 = yellow zone) and 22 provinces are in the low cluster (C_3 = green zone). The value obtained from the decision tree for cluster height (C_1 = red zone) based on C4.5 is if the number of positive cases is smaller than 9524 and greater than 4329 ($4329 > x_1 < 9524$). The nine provinces included in the high cluster (C_1 = red zone) are Aceh, Bali, DKI Jakarta, West Java, Central Java, East Java, South Kalimantan, South Sumatra and South Sulawesi. The results of the combination of these methods can be applied and provide knowledge in the form of new information about mapping in the form of clusters to the distribution of the Covid-19 pandemic in Indonesia.

Keywords: Data Mining, Classification, Classification, Covid 19, Indonesia

1. PENDAHULUAN

Data mining merupakan salah satu cabang *artificial intelligence* yang bekerja dengan cara mencari pola atau informasi menarik dengan menguraikan penemuan pengetahuan di dalam database dengan menggunakan teknik tertentu [1]–[4]. Ada banyak teknik yang terkenal dalam data mining diantaranya klastering, klasifikasi, estimasi, asosiasi dan peramalan [5]. Diantara teknik tersebut, klastering dan klasifikasi merupakan metode yang sering digunakan [6]–[8]. Beberapa metode terkenal dari klastering adalah k-medoids dan k-means [5]. Berbicara metode k-means dan k-medoids, menurut [6] bahwa metode k-means efisien untuk kumpulan data yang lebih kecil dan k-medoids berperforma lebih baik untuk dataset besar. Selain itu menurut [5] k-medoids muncul sebagai penanggulangan kelemahan metode k-means yang sensitif terhadap *outlier*. Sementara metode klasifikasi yang sering digunakan adalah C4.5 [9]–[11].

Banyak penelitian memanfaatkan kedua metode tersebut untuk menyelesaikan permasalahan. Diantaranya [12] yang memanfaatkan kelebihan dari klastering (k-medoids). Pada penelitian tersebut metode k-medoids dapat diterapkan pada kasus ekspor bahan minyak mentah ke negara tujuan. Hasil penelitian menyatakan bahwa

cluster tinggi (C1) terdiri dari 3 negara dan *cluster* rendah (C2) terdiri dari 6 negara. Selanjutnya penelitian [10] juga memanfaatkan kelebihan dari C4.5. pada penelitian tersebut model klasifikasi C4.5 dapat diterapkan pada penentuan kualitas layanan pelanggan di bank BTN cabang pematangsiantar dengan tingkat akurasi 77,78%.

Berdasarkan hal tersebut mengkombinasikan kedua metode dalam suatu penelitian menjadi hal menarik. Mengingat kedua metode memiliki kesamaan dalam hal pemetaan (klastering dan klasifikasi). Kasus yang digunakan untuk menguji kombinasi kedua metode adalah jumlah persebaran pandemi Covid-19 di Indonesia per tanggal 7 juli 2020 yang bersumber dari Data Kementerian Kesehatan, dicuplik dan diolah dari covid19.go.id dan bnpb.go.id. Berikut merupakan data keseluruhan peta kepadatan kasus positif covid-19 di Indonesia per tanggal 7 juli 2020.

Gambar 1. Data Nasional - 7 Juli 2020

Yang menjadi pertanyaan pada penelitian adalah bagaimana membangun model kombinasi metode klastering dan klasifikasi pada kasus jumlah persebaran pandemi Covid-19 di Indonesia dan menguji model kombinasi tersebut dengan menggunakan bantuan software. Sehingga hasil dari penelitian dapat memberikan hasil pemetaan berupa klaster dan nilai aturan berupa pohon keputusan terhadap hasil klaster yang terbentuk. Sehingga dari hasil tersebut akan diperoleh pengetahuan berupa informasi.

2. METODOLOGI PENELITIAN

Data yang digunakan penelitian adalah data sekunder yang diperoleh dari Kementerian Kesehatan, dicuplik dan diolah dari covid19.go.id dan bnpb.go.id. data tersebut adalah data jumlah persebaran pandemi covid-19 di Indonesia pada tanggal 7 juli 2020 yang terdiri dari 34 *record*. Kemudian data di *preprocessing* sesuai dengan kebutuhan dengan memanfaatkan *software microsoft excel* dan *RapidMiner*. Variabel yang digunakan adalah jumlah kasus positif (x_1), jumlah kasus sembuh (x_2) dan jumlah kasus meninggal dunia (x_3) berdasarkan provinsi. Berikut data penelitian seperti yang ditunjukkan pada tabel 1.

Tabel 1. Data Penelitian

No	Provinsi	Jumlah Kasus Positif	Jumlah Kasus Sembuh	Jumlah Kasus Meninggal
1	Aceh	88	42	3
2	Bali	1.94	1.034	25
3	Banten	1.531	952	80
4	Bangka Belitung	171	136	2
5	Bengkulu	144	96	13
6	DI Yogyakarta	346	282	8
7	DKI Jakarta	12.86	8.277	661
8	Jambi	121	81	1
9	Jawa Barat	3.779	1.763	180
10	Jawa Tengah	4.878	1.617	220
11	Jawa Timur	14.6	5.114	1.079
12	Kalimantan Barat	344	305	4
13	Kalimantan Timur	603	452	11
14	Kalimantan Tengah	1.058	464	62
15	Kalimantan Selatan	3.695	1.064	205
16	Kalimantan Utara	206	176	2
17	Kepulauan Riau	313	263	16
18	Nusa Tenggara Barat	1.392	889	65
19	Sumatera Selatan	2.356	1.21	112
20	Sumatera Barat	780	637	31
21	Sulawesi Utara	1.252	317	92
22	Sumatera Utara	1.821	494	108
23	Sulawesi Tenggara	487	285	8
24	Sulawesi Selatan	6.192	2.242	202
25	Sulawesi Tengah	191	164	6
26	Lampung	201	159	12

No	Provinsi	Jumlah Kasus Positif	Jumlah Kasus Sembuh	Jumlah Kasus Meninggal
27	Riau	236	214	11
28	Maluku Utara	967	124	32
29	Maluku	830	414	17
30	Papua Barat	266	185	4
31	Papua	2.057	963	19
32	Sulawesi Barat	127	89	2
33	Nusa Tenggara Timur	118	54	1
34	Gorontalo	276	227	15

Sumber: Data olahan – 7 Juli 2020

Berdasarkan pendahuluan yang telah dijabarkan, penelitian ini melalui beberapa tahap. Mulai dari studi literatur baik yang bersumber dari jurnal (nasional dan internasional) dan prosiding (nasional dan internasional). Setelah mencari sumber literatur, proses pengumpulan data dilakukan dengan memanfaatkan data Kementerian Kesehatan, dicuplik dan diolah dari covid19.go.id dan bnpb.go.id tentang jumlah persebaran pandemi covid-19 di Indonesia pada tanggal 7 juli 2020 yang terdiri dari 34 *record*. Data yang diperoleh kemudian di *preprocessing* sebelum diolah. Data yang telah di *preprocessing* kemudian diproses dengan metode k-medoids dan hasilnya dilanjutkan diproses dengan metode C4.5. Hasil kemudian di uji validitas untuk melihat apakah memperoleh hasil yang maksimal. Misalnya menggunakan parameter *Davies Bouldin Index* (DBI) untuk mengevaluasi hasil pemetaan klaster.

Setelah itu diperoleh keputusan tentang penelitian yang dilakukan. Berikut adalah gambar metodologi penelitian seperti yang ditunjukkan berikut:

Gambar 2. Metodologi Penelitian

3. HASIL DAN PEMBAHASAN

Berdasarkan metodologi penelitian yang ditunjukkan pada gambar 1, proses dilakukan menggunakan bantuan *software* RapidMiner. Kombinasi metode klastering dan klasifikasi memiliki peran masing-masing. Metode klastering melakukan pemetaan berupa klaster berdasarkan data tabel 1 dan hasil pemetaan akan diproses kembali menggunakan metode klasifikasi untuk melihat nilai aturan berupa pohon keputusan. Sebelum melakukan pemetaan, penentuan jumlah klaster dilakukan dengan menggunakan parameter *Davies Bouldin Index* (DBI). Berikut grafik perbandingan jumlah klaster (k) dengan menggunakan *Davies Bouldin Index*.

Tabel 2. Hasil Perbandingan DBI Klaster

klaster	Davies Bouldin Index
k=2	0.828
k=3	0.74
k=4	0.867

Gambar 3. Perbandingan DBI pada RapidMiner

Berdasarkan perbandingan *nilai Davies-Bouldin Index* pada gambar 3, jumlah kluster ($k=3$) menjadi kluster terbaik karena memiliki nilai terkecil yakni 0.74 (optimal). Berdasarkan hal tersebut jumlah kluster yang digunakan pada penelitian ini adalah 3 label yakni kluster tinggi (C1= zona merah), kluster waspada (C2= zona kuning), kluster rendah (C3= zona hijau). Berikut adalah model kombinasi metode klastering dan klasifikasi dengan menggunakan software RapidMiner seperti yang ditunjukkan pada gambar berikut:

Gambar 4. View Proses pada RapidMiner

Pada gambar 4 dijelaskan rancangan kombinasi metode klastering dan klasifikasi dengan menggunakan bantuan *software* RapidMiner. Beberapa parameter digunakan untuk menghasilkan *output* yang sesuai keinginan seperti *change attribute role*, *applay model*, *performance*, *clustering* (k-medoids) dan *decision tree*. Setiap parameter memiliki tugas masing-masing. *Read excel* (covid training) yang digunakan untuk *input* data (**tabel 1**) akan diproses menuju *clustering* (k-medoids) dan *change attribute role*. *change attribute role* akan mengecek apakah label kluster yang digunakan pada data sudah benar. Jika oke, *apply model* akan menampilkan visual data (grafik dan tabel) dengan menggunakan data yang bersumber dari *read excel* (covid testing = covid training). Data tersebut akan diproses ke *decision tree* yang selanjutnya akan diukur *performance* dari hasil yang diperoleh. Untuk parameter k-medoids diubah sesuai dengan keinginan. *Measure types* yang digunakan adalah *mixedmeasures*. Hal ini dilakukan karena parameter ini tersedia ketika tipe pengukuran dilakukan secara *random* pada saat penentuan nilai *centroid*. Sedangkan parameter C4.5, nilai *criterion* yang dipilih adalah *information gain*. Berikut gambar pengaturan parameter pada setiap metode seperti yang ditunjukkan pada gambar berikut:

Gambar 5. Pengaturan parameter pada setiap metode (K-medoids dan C4.5)

Untuk melihat hasil analisis, klik ikon Play. Tunggu beberapa saat, komputer membutuhkan waktu untuk menyelesaikan perhitungan. Berikut hasil pemetaan berupa kluster pada jumlah persebaran pandemi Covid-19 di Indonesia seperti yang ditunjukkan pada gambar berikut:

Gambar 6. Hasil pemetaan kluster

Pada gambar 6 hasil dari pemetaan dapat dilihat dimana klaster tinggi (C1= zona merah) terdiri 9 provinsi (*cluster_0*), klaster waspada (C2= zona kuning) terdiri 3 provinsi (*cluster_1*) dan klaster rendah (C3= zona hijau) terdiri 22 provinsi (*cluster_2*). Berikut hasil lengkap pemetaan klaster terhadap jumlah persebaran pandemi Covid-19 di Indonesia:

- klaster tinggi (C1= zona merah) adalah Aceh, Bali, DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur, Kalimantan Selatan, Sumatera Selatan dan Sulawesi Selatan
- klaster waspada (C2= zona kuning) adalah Banten, Nusa Tenggara Barat dan Papua
- klaster rendah (C3= zona hijau) adalah Bangka Belitung, Bengkulu, DI Yogyakarta, Jambi, Kalimantan Barat, Kalimantan Timur, Kalimantan Tengah, Kalimantan Utara, Kepulauan Riau, Sumatera Barat, Sulawesi Utara, Sumatera Utara, Sulawesi Tenggara, Sulawesi Tengah, Lampung, Riau, Maluku Utara, Maluku, Papua Barat, Sulawesi Barat, Nusa Tenggara Timur, Gorontalo.

Dalam penentuan klaster (*cluster_0*, 1 dan 2) juga didasarkan atas hasil *centroid* akhir yang ditunjukkan oleh gambar berikut:

Attribute	cluster_0	cluster_1	cluster_2
Positif	6.192	2.067	276
Sembuh	2.242	963	227
Meninggal	202	19	15

Gambar 7. Hasil *centroid* akhir (k-medoids)

Pada gambar 7, *cluster_0* adalah klaster tinggi (C1= zona merah), *cluster_1* adalah klaster waspada (C2= zona kuning) dan *cluster_2* adalah klaster rendah (C3= zona hijau). Hasil dari pemetaan klaster dapat di visualisasikan dengan diagram *plot scatter* seperti yang ditunjukkan pada gambar berikut:

Gambar 8. Grafik hasil pemetaan dengan RapidMiner (*plot scatter*)

Tabel 3. Hasil pemetaan kluster dengan RapiMiner (data *export*)

Provinsi	Positif	Sembuh	Meninggal	label
Aceh	88.0	42.0	3.0	cluster_0
Bali	1.9	1.0	25.0	cluster_0
Banten	1.5	952.0	80.0	cluster_1
Bangka Belitung	171.0	136.0	2.0	cluster_2
Bengkulu	144.0	96.0	13.0	cluster_2
DI Yogyakarta	346.0	282.0	8.0	cluster_2
DKI Jakarta	12.9	8.3	661.0	cluster_0
Jambi	121.0	81.0	1.0	cluster_2
Jawa Barat	3.8	1.8	180.0	cluster_0
Jawa Tengah	4.9	1.6	220.0	cluster_0
Jawa Timur	14.6	5.1	1.1	cluster_0
Kalimantan Barat	344.0	305.0	4.0	cluster_2
Kalimantan Timur	603.0	452.0	11.0	cluster_2
Kalimantan Tengah	1.1	464.0	62.0	cluster_2
Kalimantan Selatan	3.7	1.1	205.0	cluster_0
Kalimantan Utara	206.0	176.0	2.0	cluster_2
Kepulauan Riau	313.0	263.0	16.0	cluster_2
Nusa Tenggara Barat	1.4	889.0	65.0	cluster_1
Sumatera Selatan	2.4	1.2	112.0	cluster_0
Sumatera Barat	780.0	637.0	31.0	cluster_2
Sulawesi Utara	1.3	317.0	92.0	cluster_2
Sumatera Utara	1.8	494.0	108.0	cluster_2
Sulawesi Tenggara	487.0	285.0	8.0	cluster_2
Sulawesi Selatan	6.2	2.2	202.0	cluster_0
Sulawesi Tengah	191.0	164.0	6.0	cluster_2
Lampung	201.0	159.0	12.0	cluster_2
Riau	236.0	214.0	11.0	cluster_2
Maluku Utara	967.0	124.0	32.0	cluster_2
Maluku	830.0	414.0	17.0	cluster_2
Papua Barat	266.0	185.0	4.0	cluster_2
Papua	2.1	963.0	19.0	cluster_1
Sulawesi Barat	127.0	89.0	2.0	cluster_2
Nusa Tenggara Timur	118.0	54.0	1.0	cluster_2
Gorontalo	276.0	227.0	15.0	cluster_2

Pada tabel 3, hasil pemetaan kluster dapat diexport ke excel sehingga hasil dari pemetaan dapat digunakan sesuai dengan kebutuhan. Hasil pemetaan tersebut selanjutnya diproses dengan metode C4.5 untuk melihat informasi berupa aturan pohon keputusan. Berikut hasil pohon keputusan yang diperoleh dengan menggunakan bantuan *software* RapidMiner seperti yang ditunjukkan pada gambar berikut:

Gambar 9. Hasil pohon keputusan dengan RapidMiner (*plot scatter*)

Dari pohon keputusan dapat diambil suatu informasi berdasarkan kluster bahwa label *cluster_0* (kluster tinggi (C1= zona merah)) terjadi jika jumlah kasus positif lebih kecil dari 9524 dan lebih besar dari 4329 ($4329 > x1 < 9524$). Berikut aturan pohon keputusan lengkap yang diperoleh dengan menggunakan bantuan *software* RapidMiner.

Tree

```
label = cluster_0
| Positif > 4.329
| | Positif > 9.524: Aceh {Aceh=1, Bali=0, Banten=0, Bangka Belitung=0,
Bengkulu=0, DI Yogyakarta=0, DKI Jakarta=1, Jambi=0, Jawa Barat=0, Jawa Tengah=0,
Jawa Timur=1, Kalimantan Barat=0, Kalimantan Timur=0, Kalimantan Tengah=0,
Kalimantan Selatan=0, Kalimantan Utara=0, Kepulauan Riau=0, Nusa Tenggara Barat=0,
Sumatera Selatan=0, Sumatera Barat=0, Sulawesi Utara=0, Sumatera Utara=0, Sulawesi
Tenggara=0, Sulawesi Selatan=0, Sulawesi Tengah=0, Lampung=0, Riau=0, Maluku
Utara=0, Maluku=0, Papua Barat=0, Papua=0, Sulawesi Barat=0, Nusa Tenggara Timur=0,
Gorontalo=0}
| | Positif ≤ 9.524: Jawa Tengah {Aceh=0, Bali=0, Banten=0, Bangka Belitung=0,
Bengkulu=0, DI Yogyakarta=0, DKI Jakarta=0, Jambi=0, Jawa Barat=0, Jawa Tengah=1,
Jawa Timur=0, Kalimantan Barat=0, Kalimantan Timur=0, Kalimantan Tengah=0,
Kalimantan Selatan=0, Kalimantan Utara=0, Kepulauan Riau=0, Nusa Tenggara Barat=0,
Sumatera Selatan=0, Sumatera Barat=0, Sulawesi Utara=0, Sumatera Utara=0, Sulawesi
Tenggara=0, Sulawesi Selatan=1, Sulawesi Tengah=0, Lampung=0, Riau=0, Maluku
Utara=0, Maluku=0, Papua Barat=0, Papua=0, Sulawesi Barat=0, Nusa Tenggara Timur=0,
Gorontalo=0}
| | Positif ≤ 4.329
| | | Positif > 3.026: Jawa Barat {Aceh=0, Bali=0, Banten=0, Bangka Belitung=0,
Bengkulu=0, DI Yogyakarta=0, DKI Jakarta=0, Jambi=0, Jawa Barat=1, Jawa Tengah=0,
Jawa Timur=0, Kalimantan Barat=0, Kalimantan Timur=0, Kalimantan Tengah=0,
Kalimantan Selatan=1, Kalimantan Utara=0, Kepulauan Riau=0, Nusa Tenggara Barat=0,
Sumatera Selatan=0, Sumatera Barat=0, Sulawesi Utara=0, Sumatera Utara=0, Sulawesi
Tenggara=0, Sulawesi Selatan=0, Sulawesi Tengah=0, Lampung=0, Riau=0, Maluku
Utara=0, Maluku=0, Papua Barat=0, Papua=0, Sulawesi Barat=0, Nusa Tenggara Timur=0,
Gorontalo=0}
| | | Positif ≤ 3.026: Bali {Aceh=0, Bali=1, Banten=0, Bangka Belitung=0,
Bengkulu=0, DI Yogyakarta=0, DKI Jakarta=0, Jambi=0, Jawa Barat=0, Jawa Tengah=0,
Jawa Timur=0, Kalimantan Barat=0, Kalimantan Timur=0, Kalimantan Tengah=0,
Kalimantan Selatan=0, Kalimantan Utara=0, Kepulauan Riau=0, Nusa Tenggara Barat=0,
Sumatera Selatan=1, Sumatera Barat=0, Sulawesi Utara=0, Sumatera Utara=0, Sulawesi
Tenggara=0, Sulawesi Selatan=0, Sulawesi Tengah=0, Lampung=0, Riau=0, Maluku
Utara=0, Maluku=0, Papua Barat=0, Papua=0, Sulawesi Barat=0, Nusa Tenggara Timur=0,
Gorontalo=0}
label = cluster_1: Banten {Aceh=0, Bali=0, Banten=1, Bangka Belitung=0, Bengkulu=0,
DI Yogyakarta=0, DKI Jakarta=0, Jambi=0, Jawa Barat=0, Jawa Tengah=0, Jawa Timur=0,
Kalimantan Barat=0, Kalimantan Timur=0, Kalimantan Tengah=0, Kalimantan Selatan=0,
Kalimantan Utara=0, Kepulauan Riau=0, Nusa Tenggara Barat=1, Sumatera Selatan=0,
Sumatera Barat=0, Sulawesi Utara=0, Sumatera Utara=0, Sulawesi Tenggara=0, Sulawesi
Selatan=0, Sulawesi Tengah=0, Lampung=0, Riau=0, Maluku Utara=0, Maluku=0, Papua
Barat=0, Papua=1, Sulawesi Barat=0, Nusa Tenggara Timur=0, Gorontalo=0}
```

Selain itu hasil uji performance dilakukan dengan menggunakan operator *cluster distance performance*. Operator ini digunakan untuk mengevaluasi kinerja metode pengelompokan berbasis *centroid*. Operator ini memberikan daftar nilai kriteria kinerja berdasarkan *centroid*. *Cluster distance performance* yang dimaksud penggunaan *Davies-Bouldin Index (DBI)*. Dimana pada penelitian ini nilai DBI cukup optimal yakni 0.740 seperti yang ditunjukkan pada gambar berikut:

<p>Avg. within centroid distance</p> <p>Avg. within centroid distance: -71708.965</p> <p>Davies Bouldin</p>	<p>PerformanceVector</p> <p>PerformanceVector:</p> <p>Avg. within centroid distance: -71708.965</p> <p>Avg. within centroid distance_cluster_0: -37709.045</p> <p>Avg. within centroid distance_cluster_1: -3811.573</p> <p>Avg. within centroid distance_cluster_2: -94876.759</p> <p>Davies Bouldin: -0.740</p>
--	--

Gambar 10. Hasil uji *performance*

4. KESIMPULAN

Berdasarkan hasil penelitian dapat disimpulkan bahwa kombinasi metode klastering dan klasifikasi dapat diterapkan pada kasus persebaran pandemi covid-19 di Indonesia dengan memanfaatkan *software* bantu RapidMiner. Dengan menggunakan kombinasi tersebut untuk hasil pemetaan wilayah berupa klaster diperoleh 9 provinsi berada di klaster tinggi (C1= zona merah), 3 provinsi berada di klaster waspada (C2= zona kuning) dan 22 provinsi berada di klaster rendah (C3= zona hijau). Nilai yang diperoleh dari pohon keputusan untuk klaster tinggi (C1= zona merah) berdasarkan C4.5 adalah jika jumlah kasus positif lebih kecil dari 9524 dan lebih besar dari 4329 ($4329 > x1 < 9524$). Sembilan provinsi yang masuk klaster tinggi (C1= zona merah) adalah Aceh, Bali, DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur, Kalimantan Selatan, Sumatera Selatan dan Sulawesi Selatan. Hasil dari penelitian dapat menjadi bahan referensi terhadap peta penyebaran covid-19 di Indonesia. Kedepannya, hasil dari pemetaan klaster akan dikembangkan dengan memanfaatkan metode *neural network* untuk melihat apakah *neural network* dalam melakukan prediksi dengan hasil pemetaan klaster berupa nilai akurasi.

REFERENCES

- [1] A. P. Windarto *et al.*, "Analysis of the K-Means Algorithm on Clean Water Customers Based on the Province," *J. Phys. Conf. Ser.*, vol. 1255, no. 1, 2019, doi: 10.1088/1742-6596/1255/1/012001.
- [2] A. P. Windarto, "Penerapan Data Mining Pada Ekspor Buah-Buahan Menurut Negara Tujuan Menggunakan K-Means Clustering," *Techno.COM*, vol. 16, no. 4, pp. 348–357, 2017.
- [3] A. P. Windarto, "Implementation of Data Mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method," *Int. J. Artif. Intell. Res.*, vol. 1, no. 2, pp. 26–33, 2017.
- [4] M. G. Sadewo, A. P. Windarto, and D. Hartama, "Penerapan datamining pada populasi daging ayam ras pedaging di indonesia berdasarkan provinsi menggunakan k-means clustering," *InfoTekJar (Jurnal Nas. Inform. dan Teknol. Jaringen)*, vol. 2, no. 1, pp. 60–67, 2017.
- [5] I. Kamila, U. Khairunnisa, and Mustakim, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 5, no. 1, pp. 119–125, 2019.
- [6] D. Marlina, N. Lina, A. Fernando, and A. Ramadhan, "Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak," *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 4, no. 2, p. 64, 2018, doi: 10.24014/coreit.v4i2.4498.
- [7] J. Jamal and D. Yanto, "Analisis RFM dan Algoritma K-Means untuk Clustering Loyalitas Customer," *Energy*, vol. 9, no. 1, pp. 0–8, 2019.
- [8] E. H. S. Atmaja, "Implementation of k-Medoids Clustering Algorithm to Cluster Crime Patterns in Yogyakarta," *Int. J. Appl. Sci. Smart Technol.*, vol. 1, no. 1, pp. 33–44, 2019, doi: 10.24071/ijasst.v1i1.1859.
- [9] D. Hartama, A. Perdana Windarto, and A. Wanto, "The Application of Data Mining in Determining Patterns of Interest of High School Graduates," *J. Phys. Conf. Ser.*, vol. 1339, no. 1, 2019, doi: 10.1088/1742-6596/1339/1/012042.
- [10] M. Widyastuti, A. G. Fepdiani Simanjuntak, D. Hartama, A. P. Windarto, and A. Wanto, "Classification Model C.45 on Determining the Quality of Customer Service in Bank BTN Pematangsiantar Branch," *J. Phys. Conf. Ser.*, vol. 1255, no. 012002, pp. 1–6, 2019, doi: 10.1088/1742-6596/1255/1/012002.
- [11] W. Katrina, H. J. Damanik, F. Parhusip, D. Hartama, A. P. Windarto, and A. Wanto, "C.45 Classification Rules Model for Determining Students Level of Understanding of the Subject," *J. Phys. Conf. Ser.*, vol. 1255, no. 012005, pp. 1–7, 2019, doi: 10.1088/1742-6596/1255/1/012005.
- [12] F. Rahman, I. I. Ridho, M. Muflih, S. Pratama, M. R. Raharjo, and A. P. Windarto, "Application of Data Mining Technique using K-Medoids in the case of Export of Crude Petroleum Materials to the Destination Country Application of Data Mining Technique using K-Medoids in the case of Export of Crude Petroleum Materials to the Destination C," 2020, doi: 10.1088/1757-899X/835/1/012058.