

ANALISIS CLUSTERING KASUS COVID 19 DI INDONESIA MENGUNAKAN ALGORITMA K-MEANS

Tikaridha Hardiani¹

¹Teknologi Informasi, Universitas 'Aisyiyah Yogyakarta

email: tikaridha@unisayogya.ac.id¹

Abstrak

Kasus Covid 19 semakin meningkat di Indonesia. Penelitian ini bermaksud untuk menerapkan teknik penambangan data (*data mining*) dengan algoritma clustering K-Means. Penelitian ini menganalisis pola penyebaran Covid 19 dengan mengelompokkan kasus Covid 19 di Indonesia yang didapatkan dari website *dataset* Kaggle. Metode penambangan data (*data mining*) yang digunakan yaitu *Cross Industry Standard Process for Data-Mining* (CRISP-DM). Penelitian ini mengembangkan dari penelitian yang terdahulu dengan data dan atribut yang lebih banyak. Data yang digunakan sebanyak 16.284 dari tanggal 1 Maret 2020 hingga 9 Juli 2021. Pengelompokan provinsi di Indonesia ke beberapa *cluster* tertentu sehingga dapat mengetahui daerah dengan jumlah kasus yang banyak dan yang sedikit serta karakteristiknya. Hasil klasterisasi provinsi diharapkan dapat memberikan saran kepada pemerintah dalam membuat aturan atau kebijakan terkait pembatasan kegiatan masyarakat atau kebijakan lainnya dalam mengatasi penyebaran COVID-19. Penentuan jumlah *cluster* yang optimal atau validasi *cluster* menggunakan *David Boulden index* (DBI). *Cluster* yang terbaik ditentukan dari nilai David Boulden Index yang terendah. Hasil penelitian ini diperoleh 3 *cluster* yang terbaik dengan nilai DBI terendah, yaitu sebesar 0,47. *Cluster* 1 terdiri dari 30 provinsi, *Cluster* 2 dan 3 masing-masing 2 provinsi. Saran untuk peneliti selanjutnya yaitu menambahkan algoritma clustering yang lain dan membandingkan beberapa algoritma untuk memperoleh hasil yang terbaik.

Kata kunci: clustering, K-Means, data mining, kasus Covid 19, Indonesia

Abstract

Covid-19 cases are increasing in Indonesia. This study intends to apply data mining techniques with the K-Means clustering algorithm. This study analyzes the pattern of the spread of Covid 19 by grouping Covid 19 cases in Indonesia obtained from the Kaggle dataset website. The data mining method used is Cross Industry Standard Process for Data-Mining (CRISP-DM). This study develops from previous research with more data and attributes. The data used is 16,284 from March 1, 2020 to July 9, 2021. The grouping of provinces in Indonesia into certain clusters so as to identify areas with a large number of cases and few and their characteristics. The results of the provincial clustering are expected to provide advice to the government in making rules or policies regarding restrictions on community activities or other policies in overcoming the spread of COVID-19. Determination of the optimal number of clusters or cluster validation using the David Boulden index (DBI). The best cluster is determined from the lowest David Boulden Index value. The results of this study obtained the best 3 clusters with the lowest DBI value, which was 0.47. Cluster 1 consists of 30 provinces, Cluster 2 and 3 consist of 2 provinces each. Suggestions for future work are adding another clustering algorithm and comparing several algorithms to obtain the best results.

Keywords : clustering, K-Means, data mining, Covid 19, Indonesia

Diterima Redaksi: 08-03-2022 | Selesai Revisi: 30-07-2022 | Diterbitkan Online: 31-07-2022

DOI: <https://doi.org/10.23887/janapati.v11i2.45376>

PENDAHULUAN

Covid 19 ialah penyakit yang ditimbulkan virus corona. Penyakit ini berasal dari Wuhan, Provinsi Hubei, China pada bulan Desember 2019. Penyakit ini bergerak sangat cepat dari Wuhan ke penjuru dunia. Tanggal 11 Maret 2020 Organisasi kesehatan dunia memutuskan Covid 19 sebagai pandemi[1]. Virus corona membahayakan manusia karena mempunyai gejala mematikan seperti nyeri dada, sesak nafas, kesulitan bernafas dan ketidakmampuan berbicara[2].

Indonesia termasuk negara penduduk terpadat di dunia. Indonesia menempati peringkat ke empat penduduk terpadat di dunia, oleh karena itu virus Covid 19 ini diprediksi akan menyebar dalam waktu yang lama[3]. Kasus Covid 19 di Indonesia pada tanggal 9 November 2021, menurut website covid19.go.id, terkonfirmasi positif sebanyak 4.248.843, sembuh sebanyak 4.095.663 dan yang meninggal sebanyak 143.578 jiwa.

Data mining atau penambangan data merupakan proses pencarian pola yang menarik dan tersembunyi dari kumpulan data yang berukuran besar. Data tersebut tersimpan di basis data, gudang data, atau tempat penyimpanan data lainnya[4]. Penambangan data merupakan salah satu bagian dari kecerdasan buatan. Teknik dalam data mining yaitu asosiasi, clustering, estimasi, klustering dan prediksi[5]. Teknik penambangan data yang dipakai dalam penelitian ini yaitu *clustering*. Teknik clustering dan klasifikasi memiliki perbedaan, pada clustering tidak mempunyai variable target dalam melakukan pengelompokan[6]. Analisis *clustering* memilah kumpulan data menjadi beberapa kategori dengan kesamaan keunikan atau perhitungan jarak [4]. *Clustering* adalah proses pengelompokan dan pembagian pola data menjadi beberapa jumlah data set. Pola serupa yang terbentuk dikelompokkan pada cluster yang sama dan memisahkan diri dengan membentuk pola yang berbeda ke *cluster* yang berbeda.[7].

Algoritma K-Means merupakan algoritma yang mengelompokkan data dengan mencoba memisahkan data menjadi beberapa kelompok sehingga data yang memiliki beberapa kemiripan dalam kelompok yang sama. Data yang berbeda dikelompokkan dalam kelompok yang lain[2][8]. Tujuan dari algoritma K-Means adalah meminimalkan fungsi dengan meminimalkan variasi antar data dalam suatu

cluster dan memaksimalkan variasi data pada *cluster* lainnya[9].

Penelitian yang dilakukan oleh Rahmad Kurniawan menganalisis metode *clustering* dan prediksi resiko covid 19 di negara pandemi. Penelitian ini menerapkan dua metode *unsupervised learning* yaitu K-Means dan korelasi pada suatu set.Data yang digunakan dari 27 Maret sampai 16 Agustus 2020. Hasil akurasi cluster sebesar 97% dengan 5 cluster dan prediksi hubungan terkait antara total kematian dan atribut pasien kritis sebesar 0,85 [10]. Indraputra dan Fitriana menggunakan K-Means untuk *clustering* data covid 19 dengan mengelompokkan wilayah-wilayah di dunia. Penelitian ini menghasilkan 2 *cluster* dengan *cluster* ke dua mempunyai jumlah terjangkit dan meninggal yang lebih tinggi daripada *cluster* 1[7]. Mohammad Reza dan kawan-kawan meneliti tingkat penyebaran Covid 19 di negara yang beresiko tinggi menggunakan metode fuzzy *clustering*. Penelitian ini menggunakan sebaran penyebaran Covid-19 di Amerika Serikat, Spanyol, Italia, Jerman, Inggris, Prancis, dan Iran dibandingkan dan diklaster menggunakan teknik fuzzy clustering. Rangkaian waktu dari dataset Covid-19 di negara-negara tertentu dipertimbangkan. Kemudian, hubungan antara penyebaran Covid-19 dan jumlah penduduk dipelajari dengan menggunakan korelasi Pearson[11]. Hasil *clustering* memperlihatkan bahwa penyebaran di Spanyol dan Italia memperlihatkan persamaan dengan negara lain. Wargijono dalam penelitian yang dilakukan memadankan algoritma *clustering* yaitu K Means dan K medoid. Data yang digunakan bersumber dari kementerian kesehatan dengan 4 atribut, antara lain kasus terkonfirmasi, pengobatan, pemulihan, dan kasus kematian. Dalam penelitian ini, hanya 2 atribut yang digunakan: kasus terkonfirmasi dan kasus kematian. Dari analisis data dan hasil pengolahan melalui perbandingan antara metode K-Medoids dan K-means untuk klasterisasi penyebaran virus corona di Indonesia, diperoleh suatu kesimpulan. Dengan nilai indeks Davies Boulden dari nilai K2 hingga K9 ternyata metode K-Means mendapatkan nilai terkecil pada K-5 sebesar 0,064, sedangkan K-Medoids pada nilai k-2 sebesar 0,411. Dengan demikian, dari kedua metode yang digunakan dapat disimpulkan bahwa metode yang paling baik untuk klasterisasi penyebaran wabah virus corona di Indonesia adalah metode K-Means[2]. Algoritma K-Means juga digunakan oleh Ali dalam mengelompokkan kabupaten atau kota di Jawa

Tengah dengan data Covid 19. Hasil *clustering* menunjukkan kelompok 2 dan kelompok 3 ialah daerah yang lebih diperhatikan oleh pemerintah. Daerah tersebut mempunyai besaran kasus aktif dan kasus kematian COVID-19 yang tinggi di Jawa Tengah[12].

Penelitian yang dilakukan oleh Abdullah dkk bertujuan untuk mengelompokkan provinsi-provinsi di Indonesia dari risiko pandemi COVID-19 berdasarkan data penyakit virus corona 2019 (COVID-19). Pengelompokan ini diperoleh dari data Satuan Tugas Percepatan Penanganan Covid-19 Indonesia (SATGAS COVID-19) pada 19 April 2020. Algoritma yang digunakan yaitu K Means dengan perhitungan optimasi *cluster* menggunakan *elbow method*. *Clustering* menghasilkan 3 grup provinsi[13]. Zulfa dan kawan-kawan melakukan pembagian kasus covid 19 di provinsi Lampung menggunakan algoritma K-Means. Data yang digunakan meliputi atribut Suspek, Kabupaten/Kota, Probable, Selesai isolasi, Konfirmasi Positif, dan Kematian. Pembagian menjadi empat kelompok berdasarkan zona merah, orange, kuning dan hijau. Menggunakan *Davies Bouldin Index* untuk validasi *cluster* [14].

Penelitian ini bertujuan untuk menerapkan algoritma K-Means guna mengelompokkan tingkat penyebaran covid 19 di Indonesia yang diperoleh dari website *dataset* Kaggle. Pengelompokan provinsi di Indonesia ke beberapa klater tertentu sehingga dapat mengetahui daerah dengan jumlah kasus yang banyak dan yang sedikit. Hasil klasterisasi provinsi diharapkan dapat memberikan saran kepada pemerintah dalam menentukan aturan atau kebijakan terkait pembatasan kegiatan masyarakat atau kebijakan lainnya dalam mengatasi penyebaran COVID-19. Penentuan jumlah *cluster* yang optimal menggunakan *David boulden index*. Penelitian ini mengembangkan dari penelitian yang terdahulu dengan data dan atribut yang lebih banyak. Data yang digunakan sebanyak 16.284 dari tanggal 1 Maret 2020 hingga 9 Juli 2021.

METODE

Penelitian ini menggunakan data Covid 19 yang terdapat pada website *dataset* Kaggle. *Dataset* terdiri dari 16.284 data. Penelitian ini menggunakan variable yang terdiri dari total kasus, total meninggal, total sembuh, total kasus aktif, peluang kematian dan peluang kesembuhan. Metode yang dipakai dalam penelitian ini yaitu *Cross Industry Standard Process for Data-Mining* (CRISP-DM). CRISP-DM merupakan model proses penambangan data yang digunakan untuk memecahkan

masalah oleh para ahli. Penelitian berpedoman pada enam tahap CRISP-DM, atau siklus hidup pengembangan data mining sebagai *framework* dari proyek *data mining*. Metodologi CRISP-DM dijelaskan dalam model proses hierarkis, yang terdiri dari empat tingkat abstraksi (dari umum ke khusus): fase, tugas umum, tugas khusus, dan contoh proses[15]. Tahapan CRISP-DM yaitu[16][17][18]:

a. Pemahaman Bisnis (*Business Understanding*).

Fase awal CRISP-DM yang berpusat mengidentifikasi faktor penting seperti kriteria keberhasilan, tujuan dan persyaratan kebutuhan bisnis[19]. Tahap ini menentukan tujuan penelitian Kemudian dilakukan penerjemahan dan analisis pada batasan dan tujuan yang menjadi masalah penambangan data.

b. Pemahaman Data (*Data Understanding*)

Fase kedua pada CRISP-DM. yang mempunyai fokus pengumpulan data, pengecekan kualitas data, dan eksplorasi data untuk mendapatkan informasi yang tersembunyi[19]. Melakukan pengumpulan data untuk dianalisis dan diselidiki sehingga diketahui struktur pola data awal agar diperoleh gambaran pengetahuan awal yang akan dihasilkan dari data mining. Kemudian data tersebut dievaluasi kualitasnya untuk menghilangkan data-data *missing value*, duplikasi, dan tipografi. Hubungan antara pemahaman bisnis dan pemahaman data saling berkaitan.

c. Pengolahan Data (*Data Preparation*).

Fase ketiga dalam CRISP DM, yang berfokus pada pemilihan dan persiapan akhir data. Tahap ini dilakukan persiapan data dengan menentukan kasus dan variabel yang ingin dianalisis. Tahap ini mencakup pemilihan variable, tabel, pembersihan dan transformasi data[19]. Pada tahapan ini dapat dilakukan berulang kali. Setelah pengolahan data dilakukan diharapkan data siap dan memenuhi kriteria dilakukan pemodelan.

d. Pemodelan (*Modelling*).

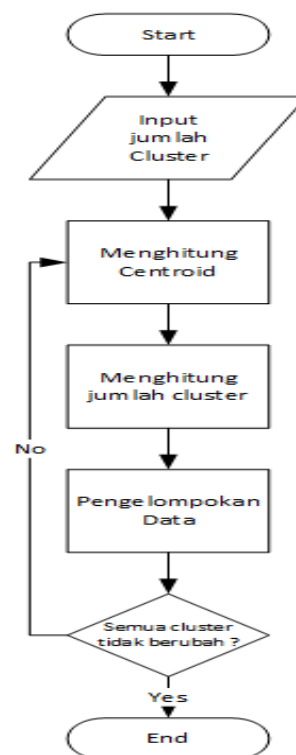
Fase keempat di CRISP-DM dan penerapan berbagai teknik pemodelan. Pemodelan meliputi pengaturan situasi dan spesifikasi sehingga memungkinkan data dapat diproses menggunakan metode *data mining* yang direncanakan. Tahap pemodelan ini diperlukan *tools* atau pengkodean menggunakan bahasa pemrograman tertentu. Penelitian ini menggunakan perangkat lunak Rapidminer. Pemodelan menggunakan algoritma K-Means. Proses CRISP DM dapat kembali

ke tahap sebelumnya untuk menyesuaikan data dengan kebutuhan metode penambangan data yang digunakan. Langkah-langkah algoritma K-Means yaitu[14]:

1. Memutuskan jumlah kluster k
2. Menetapkan nilai-nilai awal k pusat *cluster*. Inisiasi dapat dilakukan dengan cara acak. Pusat-pusat *cluster* diberi nilai awal dengan angka-angka acak.
3. Menempatkan tiap-tiap item dataset yang jaraknya paling dekat dengan nilai

centroid ke dalam centroid kluster tersebut.

4. Menghitung rata-rata nilai item dalam setiap kluster untuk dijadikan sebagai centroid yang baru.
5. Mengulang langkah 2 dan langkah 3 sehingga nilai centroid sama dengan nilai rata-rata item dalam kluster. Menggunakan *euclidean distance* untuk menghitung jarak antar titik. Alur dari algoritma K-Means dijelaskan pada Gambar 1.



Gambar 1 Alur algoritma K-Means

e. Tahap Evaluasi (*Evaluation*).

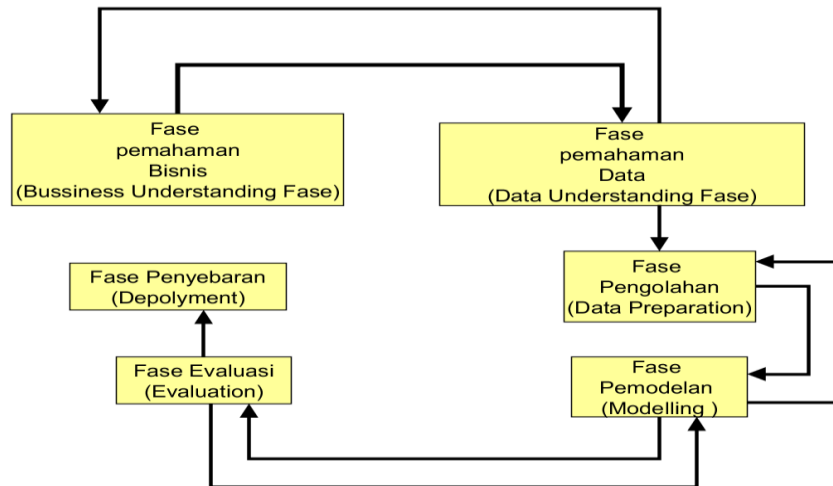
Fase kelima dari proses DM yang berpusat pada evaluasi dari hasil modeling. Interpretasi model tergantung pada algoritma dan model dapat dievaluasi untuk meninjau ketercapaian tujuan[19]. Tahapan ini mengevaluasi model yang digunakan. Menetapkan apakah terdapat model yang memenuhi tujuan pada fase awal. Evaluasi dan validasi untuk performa kluster dengan mengoptimalkan jumlah *cluster* dengan teknik *Davies Bouldin Index* (DBI). *Davies Bouldin Index* adalah ukuran untuk mengevaluasi suatu pengelompokan.

Davies Bouldin Index merupakan salah satu metode yang diperkenalkan oleh David L. Davies dan Donald W. Bouldin. *Davies Bouldin Index* digunakan untuk mengevaluasi *cluster* secara umum berdasarkan kuantitas dan kedekatan antar anggota *cluster*. memiliki korelasi positif untuk kasus “dalam kelas” dan korelasi negatif untuk kasus “antar kelas”.[20]. Perhitungan nilai *Davies Bouldin Index* didasarkan pada rasio rasio *cluster* ke-i terhadap *cluster* ke-j. Semakin kecil nilai *Davies Bouldin Index* maka *cluster* yang dihasilkan semakin baik[9][21].

f. Penyebaran (*Deployment*)

Fase keenam dan fase terakhir dari proses CRISP-DM, yang berfokus pada penentuan penggunaan pengetahuan, pengorganisasian, pelaporan dan penyajian pengetahuan yang diperoleh saat dibutuhkan [19]. Penyebaran disini adalah penggunaan model yang dihasilkan yakni rangkaian metode dan data-data representatif yang sudah terolah sehingga memberikan informasi optimal ketika

proses *data mining*. Penyebaran dalam lingkup sederhana adalah penggunaan hasil akhir dari data *mining* contohnya: pelaporan hasil proses yang menggunakan *data mining* untuk analisis *cluster* data Covid 19 di Indonesia. Gambar 2 menunjukkan tahapan CRISP-Data Mining.



Gambar 2 Tahapan CRISP DM

HASIL DAN PEMBAHASAN

a. Pemahaman Bisnis (*Business Understanding*)

Pemahaman bisnis dari penelitian ini ialah mengelompokkan dan menganalisis hasil cluster data Covid 19 yang ada di Indonesia. Data covid 19 ini meliputi 34 provinsi. Hasil pengelompokan ini dapat memberikan saran kepada pengambil keputusan tentang tingkat penyebaran covid di Indonesia.

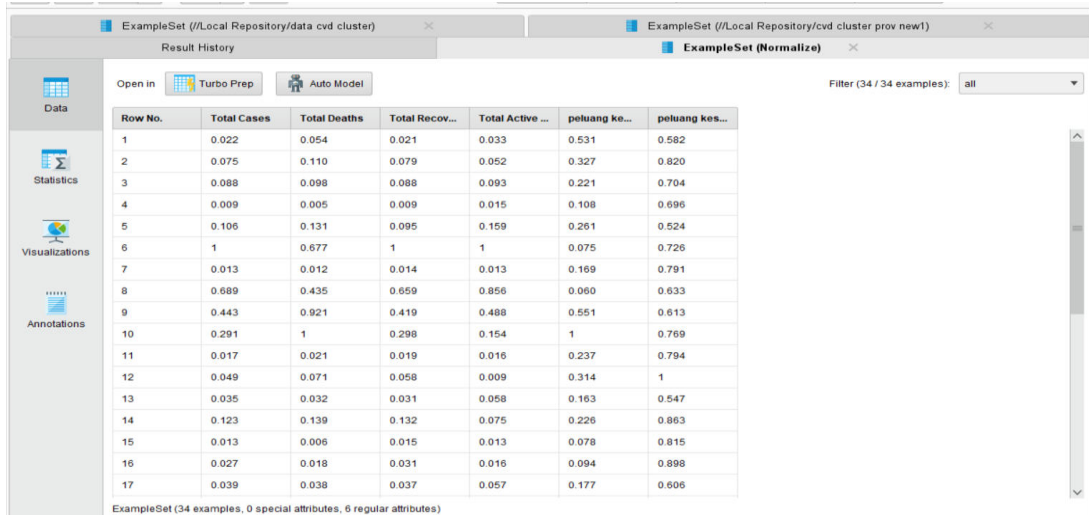
b. Pemahaman data (*Data Understanding*)

Dataset Covid 19 meliputi 34 provinsi di Indonesia yang meliputi data: *Location ISO Code, Date, Location New Deaths, New Cases, New Recovered, Location Level, Total Cases, Total Deaths, Total Active Cases, City, Province, Country, Continent, Island, Total Regencies, Total Districts, Total Rural Villages, Total Urban, Total Recovered, Population Density Villages, New Active Cases, Time Zone,*

Total Cities, Special Statu, Population, Area (km²), Longitude, Latitude, Total Cases per Million, New Deaths per Million, Case Fatality Rate, Case Recovered Rate, Total Deaths per Million, Growth Factor of New Cases dan Growth Factor of New Deaths, New Cases per Million. Jumlah data sebanyak 16.284 dari tanggal 1 Maret 2020 hingga 9 Juli 2021.

c. Pengolahan data (*Data Preparation*)

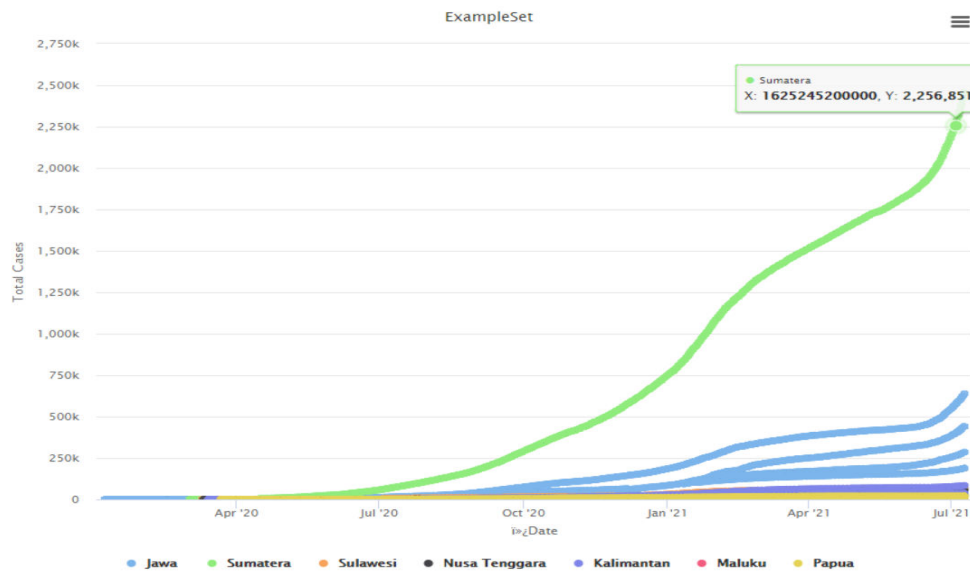
Dataset memiliki 495 data provinsi yang missing value. Data tersebut dihapus dari dataset, kemudian memilih atribut yang sesuai dengan pemodelan. Dataset juga dilakukan normalisasi. Data yang digunakan yaitu total kasus, total meninggal, total sembuh, total kasus aktif. Menambahkan peluang kematian (%) dan peluang kesembuhan (%). Tampilan data yang sudah diolah pada rapidminer diperlihatkan pada Gambar 3.



Row No.	Total Cases	Total Deaths	Total Recov...	Total Active ...	peluang ke...	peluang kes...
1	0.022	0.054	0.021	0.033	0.531	0.582
2	0.075	0.110	0.079	0.052	0.327	0.820
3	0.088	0.098	0.088	0.093	0.221	0.704
4	0.009	0.005	0.009	0.015	0.108	0.696
5	0.106	0.131	0.095	0.159	0.261	0.524
6	1	0.677	1	1	0.075	0.726
7	0.013	0.012	0.014	0.013	0.169	0.791
8	0.689	0.435	0.659	0.856	0.060	0.633
9	0.443	0.921	0.419	0.488	0.551	0.613
10	0.291	1	0.298	0.154	1	0.769
11	0.017	0.021	0.019	0.016	0.237	0.794
12	0.049	0.071	0.058	0.009	0.314	1
13	0.035	0.032	0.031	0.058	0.163	0.547
14	0.123	0.139	0.132	0.075	0.226	0.863
15	0.013	0.006	0.015	0.013	0.078	0.815
16	0.027	0.018	0.031	0.016	0.094	0.898
17	0.039	0.038	0.037	0.057	0.177	0.606

Gambar 3. Data yang sudah diolah

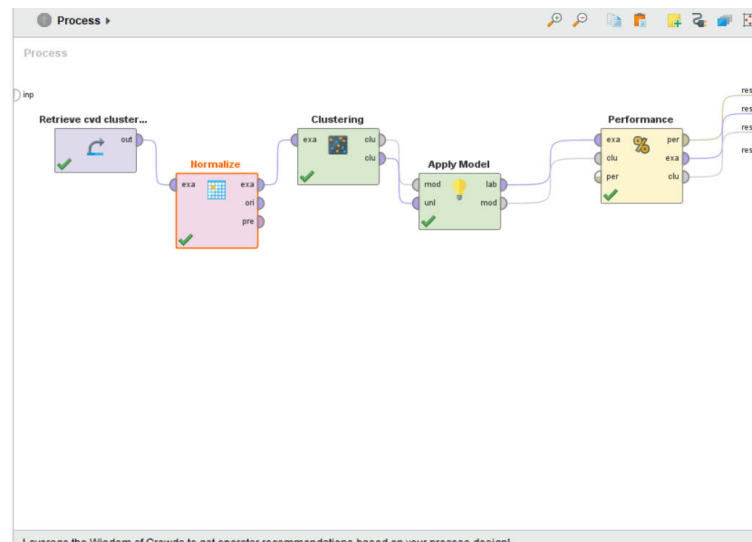
Visualisasi data total kasus provinsi di Indonesia diperlihatkan pada Gambar 4. Pada Gambar 4 terlihat total kasus di pulau Sumatera meningkat pada bulan Juli 2021. Papua dan Sulawesi menjadi pulau yang mempunyai kasus terendah pada bulan Juli 2021.



Gambar 4. Visualisasi data total kasus provinsi di Indonesia

d. Pemodelan

Pemodelan dengan algoritma K-Means menggunakan perangkat lunak Rapidminer, ditunjukkan pada Gambar 5.



Gambar 5 Pemodelan *clustering* di Rapidminer

e. Evaluasi

Evaluasi menggunakan metode *Davies Bouldin Index*. Hasil perhitungan *Davies Bouldin Index* ditunjukkan pada Tabel 1.

Tabel 1 Perhitungan Nilai DBI

Cluster	Nilai DBI
2	0,72
3	0,47
4	0,68
5	0,61
6	0,58
7	0,50

Dilihat dari tabel di atas, maka *cluster* yang optimal adalah di *cluster* 3 dengan nilai DBI 0,474. Hasil dari pengelompokan 3 *cluster*

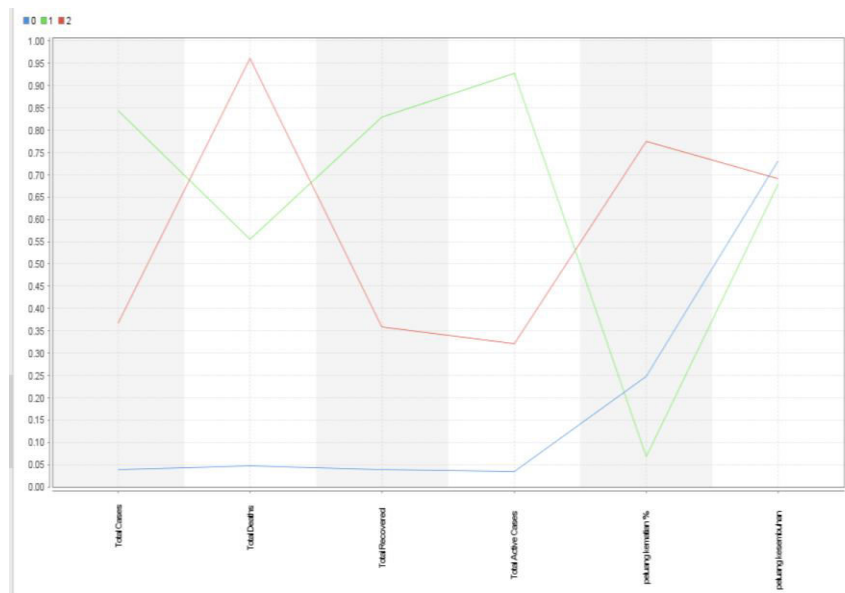
ditunjukkan pada Tabel 2. *Cluster* atau kelompok 1 meliputi 30 provinsi, *Cluster* atau kelompok 2 meliputi 2 provinsi dan *cluster* atau kelompok 3 meliputi 2 provinsi.

Tabel 2 Hasil

Cluster	Provinsi
1	Aceh, Bali, Banten, Bengkulu, DIY, Jambi, Lampung, Kepulauan bangka Belitung, Kepulauan Riau, Maluku, Maluku Utara, Nusa Tenggara Timur, Papua, Riau, Sulawesi Barat, Sulawesi Selatan, Sulawesi Tenggara, Sumatera Utara, Sulawesi Tengah, Sumatera Selatan, Gorontalo, Nusa Tenggara Barat, Kalimantan Barat, Kalimantan Timur, Kalimantan Selatan, Kalimantan utara, Papua Barat
2	Jawa Barat, DKI Jakarta
3	Jawa Timur, Jawa Tengah

pengelompokan 3 cluster

Visualisasi hasil *clustering* diperlihatkan pada Gambar 6.



Gambar 6. Visualisasi hasil *clustering*

Grafik visualisasi hasil pengelompokan menunjukkan *cluster* 1 total kasus, total kematian, total kesembuhan, total kasus aktif rendah; peluang kematian sedang dan peluang kesembuhan tinggi. *Cluster* 2 menunjukkan total kasus, total kesembuhan, total kasus aktif tinggi; total kematian sedang, peluang kematian rendah dan peluang kesembuhan tinggi. *Cluster*

3 menunjukkan total kasus, total kesembuhan, total kasus aktif sedang; total kematian, peluang kematian dan peluang kesembuhan tinggi. Cluster 2 dan 3 merupakan provinsi yang mempunyai mobilitas tinggi dan tingkat resiko Covid 19 tinggi.

f. Penyebaran

Tahap akhir metode CRISP DM yaitu pembuatan laporan. Laporan akhir menjelaskan tentang pengetahuan yang diperoleh pada data dalam proses penambangan data. Penelitian yang dilakukan menghasilkan pola, informasi, dan pengetahuan dalam mengelompokkan kasus Covid 19 di Indonesia pada periode Maret 2020 hingga Juli 2021.

KESIMPULAN

Kesimpulan penelitian ini ialah penerapan algoritma K Means dihasilkan 3 cluster yang optimal berdasarkan perhitungan Davies Bouldin Index. Nilai perhitungan DBI terendah sebesar 0,474.

Cluster 1 meliputi 30 provinsi (Aceh, Jambi, Bali, Banten, Lampung, Bengkulu, DIY, Riau, Kalimantan Barat, Kalimantan Selatan, Kalimantan timur, Kalimantan utara, Kepulauan Bangka Belitung, Kepulauan Riau, Maluku, Maluku Utara, Nusa Tenggara Barat, Nusa Tenggara Timur, Papua, Papua Barat, Sulawesi Tenggara, Sulawesi Selatan, Sulawesi Barat, Sumatera Selatan, Sumatera Utara, Sulawesi Tengah, Sumatera Selatan dan Gorontalo); cluster 2 meliputi 2 provinsi (DKI Jakarta dan Jawa Barat); cluster 3 meliputi 2 provinsi (Jawa Tengah dan Jawa Timur).

Cluster 1 total kasus, total kematian, total kesembuhan, total kasus aktif rendah; peluang kematian sedang dan peluang kesembuhan tinggi. Cluster 2 menunjukkan total kasus, total kesembuhan, total kasus aktif tinggi; total kematian sedang, peluang kematian rendah dan peluang kesembuhan tinggi. Cluster 3 menunjukkan total kasus, total kesembuhan, total kasus aktif sedang; total kematian, peluang kematian dan peluang kesembuhan tinggi. Cluster 2 dan 3 merupakan provinsi yang mempunyai mobilitas tinggi dan tingkat resiko Covid 19 tinggi.

Saran untuk peneliti selanjutnya yaitu menambahkan algoritma clustering yang lain dan membandingkan beberapa algoritma untuk memperoleh hasil yang terbaik.

REFERENSI

- [1] H. Fransiska, "Clustering Provinces in Indonesia Based on Daily Covid-19 Cases," *J. Phys. Conf. Ser.*, vol. 1863, no. 1, 2021.
- [2] W. Utomo, "The comparison of k-means and k-medoids algorithms for clustering the spread of the covid-19 outbreak in Indonesia," *Ilk. J. Ilm.*, vol. 13, no. 1, pp. 31–35, 2021.
- [3] R. Djalante *et al.*, "Review and analysis

- of current responses to COVID-19 in Indonesia: Period of January to March 2020," *Prog. Disaster Sci.*, vol. 6, 2020.
- [4] T. Hardiani, S. Sulisty, and R. Hartanto, "pada Lembaga Keuangan Mikro," vol. 03, pp. 181–187, 2014.
- [5] Z. R. S. Elsi *et al.*, "Utilization of Data Mining Techniques in National Food Security during the Covid-19 Pandemic in Indonesia," *J. Phys. Conf. Ser.*, vol. 1594, no. 1, 2020.
- [6] S. Sindi, W. R. O. Ningse, I. A. Sihombing, F. Ilmi R.H.Zer, and D. Hartama, "Analisis algoritma K-Medoids clustering dalam pengelompokan penyebaran Covid-19 di Indonesia," *Jti (Jurnal Teknol. Informasi)*, vol. 4, no. 1, pp. 166–173, 2020.
- [7] R. A. Indraputra and R. Fitriana, "K-Means Clustering Data COVID-19," *J. Tek. Ind.*, vol. 10, no. 3, p. 3, 2020.
- [8] K. Hammouda, "A Comparative Study of Data Clustering Techniques," pp. 1–21.
- [9] N. Suarna, Y. A. Wijaya, Mulyawan, T. Hartati, and T. Suprpti, "Comparison K-Medoids Algorithm and K-Means Algorithm for Clustering Fish Cooking Menu from Fish Dataset," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1088, no. 1, p. 012034, 2021.
- [10] R. Kurniawan, S. N. H. S. Abdullah, F. Lestari, M. Z. A. Nazri, A. Mujahidin, and N. Adnan, "Clustering and Correlation Methods for Predicting Coronavirus COVID-19 Risk Analysis in Pandemic Countries," *2020 8th Int. Conf. Cyber IT Serv. Manag. CITSM 2020*, 2020.
- [11] M. R. Mahmoudi, D. Baleanu, Z. Mansor, B. A. Tuan, and K. H. Pho, "Fuzzy clustering method to compare the spread rate of Covid-19 in the high risks countries," *Chaos, Solitons and Fractals*, vol. 140, pp. 1–9, 2020.
- [12] A. Mahmudan, "Clustering of District or City in Central Java Based COVID-19 Case Using K-Means Clustering," *J. Mat. Stat. dan Komputasi*, vol. 17, no. 1, pp. 1–13, 2020.
- [13] D. Abdullah, S. Susilo, A. S. Ahmar, R. Rusli, and R. Hidayat, "The application of K-means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data," *Qual. Quant.*, no. 0123456789, 2021.
- [14] Z. Nabila, A. Rahman Isnain, and Z. Abidin, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means,"

- J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, p. 100, 2021.
- [15] R. Wirth, "CRISP-DM: Towards a Standard Process Model for Data Mining," no. 24959.
- [16] T. Hardiani, "Comparison of Naive Bayes Method, K-NN (K-Nearest Neighbor) and Decision Tree for Predicting the Graduation of 'Aisyiyah University Students of Yogyakarta," *Int. J. Heal. Sci. Technol.*, vol. 2, no. 1, pp. 75–85, 2021.
- [17] T. Hardiani, "Segmentasi Nasabah Simpanan Menggunakan Fuzzy C Means Dan Fuzzy Rfm (Recency , Frequency , Monetary) Pada Bmt Xyz," *J. Ilm. NERO*, vol. 3, no. 3, pp. 185–192, 2018.
- [18] A. Nadali and H. E. Nosratabadi, "Evaluating the Success Level of Data Mining Projects Based on CRISP-DM Methodology by a Fuzzy Expert System," *IEEE*, pp. 161–165, 2011.
- [19] U. Shafique and H. Qaiser, "A Comparative Study of Data Mining Process Models (KDD , CRISP-DM and SEMMA)," *Int. J. Innov. Sci. Res.*, vol. 12, no. 1, pp. 217–222, 2014.
- [20] Y. A. Wijaya, D. A. Kurniady, E. Setyanto, W. S. Tarihoran, D. Rusmana, and R. Rahim, "Davies Bouldin Index Algorithm for Optimizing Clustering Case Studies Mapping School Facilities," *TEM J.*, vol. 10, no. 3, pp. 1099–1103, 2021.
- [21] M. Mughnyanti, S. Efendi, and M. Zarlis, "Analysis of determining centroid clustering x-means algorithm with davies-bouldin index evaluation," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 725, no. 1, 2020.