

Penerapan Data Mining Untuk Mengelompokkan Penyebaran Covid-19 di Indonesia Menggunakan Algoritma K-Means

Moh. Agus Efendi¹, Zaehol Fatah²,

^{1,2}Universitas Ibrahimy, Situbondo

Email: aguse6147@gmail.com¹, zaeholfatah@gmail.com²,

Abstrak

Pandemi COVID-19 yang dimulai pada akhir 2019 telah berdampak signifikan pada sektor kesehatan, ekonomi, dan sosial di seluruh dunia, termasuk di Indonesia. Kondisi geografis dan demografis yang beragam di Indonesia menghadirkan tantangan unik dalam memahami penyebaran kasus COVID-19. Variasi dalam distribusi kasus dipengaruhi oleh faktor-faktor seperti kepadatan penduduk, mobilitas antar wilayah, dan akses layanan kesehatan. Pendekatan berbasis data menjadi krusial untuk mengidentifikasi pola penyebaran dan mengembangkan strategi mitigasi yang efektif. Algoritma clustering K-Means digunakan sebagai metode data mining untuk mengelompokkan wilayah di Indonesia berdasarkan kesamaan karakteristik kasus COVID-19. Data sekunder, termasuk jumlah kasus, tingkat kesembuhan, dan angka kematian, diproses menggunakan aplikasi RapidMiner. Analisis menghasilkan tiga kluster: wilayah dengan prevalensi tinggi, sedang, dan rendah. Hasil tersebut konsisten dengan perhitungan manual, yang menunjukkan efektivitas algoritma serta keandalan RapidMiner dalam menangani data skala besar dan kompleks. Hasil clustering memberikan wawasan berharga tentang pola distribusi COVID-19, sehingga memungkinkan perumusan kebijakan mitigasi yang lebih spesifik dan terarah. Selain itu, penerapan algoritma K-Means menunjukkan potensi untuk membangun sistem peringatan dini dalam mengantisipasi lonjakan kasus di masa depan. Pendekatan ini menegaskan pentingnya Algorithms, COVID-19, Data Mining, Indonesia, K-Means, Regional Clustering, pengambilan keputusan berbasis data dalam pengelolaan pandemi dan strategi alokasi sumber daya.

Kata Kunci: Algoritma, COVID-19, Data Mining, Indonesia, K-Means, Pengelompokan Wilayah

PENDAHULUAN

Pandemi COVID-19 yang dimulai pada akhir tahun 2019 telah memberikan dampak yang signifikan terhadap berbagai aspek kehidupan manusia, baik dari segi kesehatan, ekonomi, dan sosial. Pesatnya penyebaran virus SARS-CoV-2 di Indonesia menghadirkan berbagai tantangan, terutama karena keragaman geografis dan demografi yang ada di negara kepulauan ini (R. Sitohang dan A. S. Dewi, 2016). Variasi penyebaran kasus COVID-19 yang terjadi di berbagai wilayah dipengaruhi oleh berbagai faktor seperti kepadatan penduduk, mobilitas antar wilayah, dan ketersediaan layanan kesehatan di setiap wilayah (S. Suyanto, 2018).

Dalam upaya mengurangi dampak pandemi COVID-19, pendekatan berbasis data berperan sangat penting dalam memahami pola penyebaran dan mendukung penerapan kebijakan mitigasi yang efektif (S.R. Girsang dan J. Simarmata, 2017). Salah satu metode yang relevan adalah data mining, dengan algoritma K-Means yang dapat digunakan untuk mengelompokkan wilayah berdasarkan kesamaan karakteristik penyebaran kasus COVID-19. Melalui penerapan algoritma ini, pola penyebaran dapat dipetakan berdasarkan faktor-faktor yang berperan, yang pada akhirnya dapat dijadikan landasan untuk merumuskan kebijakan yang lebih spesifik dan efektif.

Selain mengelompokkan

penyebaran berdasarkan karakteristik yang sama, algoritma K-Means juga berpotensi mendukung terciptanya sistem peringatan dini terhadap lonjakan kasus di masa depan. Dengan kemampuannya dalam mengolah data berskala besar dan kompleks, data mining dengan menggunakan algoritma K-Means dapat membantu pemerintah dan pemangku kepentingan dalam merancang langkah mitigasi yang lebih tepat dan efisien (A. Kusri dan E. Luthfi, 2009). Oleh karena itu, pemahaman pola penyebaran COVID-19 melalui analisis data dengan pendekatan algoritma K-Means menjadi penting untuk mendukung pengambilan kebijakan yang lebih tepat dan berbasis data dalam menghadapi pandemi ini. Berdasarkan latar belakang tersebut, penting untuk menjawab pertanyaan penelitian terkait pola penyebaran kasus COVID-19 di Indonesia dan penerapan algoritma K-Means untuk mengidentifikasi karakteristik penyebarannya (Pradana & Sihombing, 2020).

METODE

1. Pengumpulan Data

penelitiannya menggunakan data sekunder yang diperoleh dari laporan resmi terkait COVID-19 di Indonesia. Data yang digunakan meliputi jumlah kasus, jumlah pasien sembuh, dan jumlah kematian akibat COVID-19 di berbagai provinsi di Indonesia. Informasi ini diambil dari sumber terpercaya, seperti Kementerian Kesehatan atau Satgas Penanganan COVID-19, dengan fokus pada data pada tahun tertentu. Data ini digunakan sebagai dasar analisis epidemiologi untuk penelitian ini.

2. Pengolahan Data

Definisi sederhana dari data mining adalah ekstraksi informasi atau pola penting atau menarik dari data dalam database besar. Dalam jurnal ilmiah, data mining dikenal juga dengan Knowledge Discovery in Database (KDD) (Amril Mutoi Siregar, 2020).

3. K-Means

K-Means adalah salah satu algoritma dalam *clustering* atau pengelompokan yang

digunakan untuk membagi sekumpulan data ke dalam beberapa kelompok berdasarkan kemiripan atau kedekatannya. Tujuan utama dari algoritma ini adalah untuk meminimalkan jarak antara data dengan pusat kelompok (*centroid*) yang ditentukan, serta mengelompokkan data yang memiliki karakteristik serupa ke dalam satu kelompok. Secara umum, algoritma K-Means bekerja dengan cara menginisialisasi sejumlah *centroid* secara acak, kemudian mengelompokkan data berdasarkan kedekatannya dengan *centroid* tersebut, dan terus memperbaharui posisi *centroid* hingga konvergen, yakni ketika tidak ada perubahan signifikan lagi pada posisi *centroid* (Tan, P.-N., Steinbach, M., & Kumar, V, 2006).

4. Clustering

Analisis clustering merupakan salah satu teknik dalam data mining yang digunakan untuk mengelompokkan objek atau individu yang memiliki kesamaan berdasarkan sejumlah kriteria tertentu. Tujuan dari analisis clustering adalah untuk mengelompokkan elemen-elemen yang sejenis ke dalam kelompok-kelompok (*cluster*) yang saling terpisah dan eksklusif. (Pradana, M. A., & Sihombing, R., 2020).

5. Covid 19

Menurut WHO (2020), virus COVID-19 ditularkan melalui kontak dekat melalui pernafasan (seperti batuk) dan muntah. Oleh karena itu, untuk membatasi penularan virus, WHO (2020) terus menganjurkan untuk sering mencuci tangan, menggunakan pelindung pernafasan, membersihkan dan mendisinfeksi permukaan secara teratur, menjaga jarak fisik dan menghindari orang dengan gejala pernafasan atau pernafasan (M.K. Rahayu Mayang Sari, 2024).

HASIL DAN PEMBAHASAN

1. Seleksi Data

Pada tahap awal, dataset yang tersedia mencakup sejumlah atribut seperti jumlah, nama provinsi, jumlah kasus, jumlah kesembuhan, dan jumlah kematian, dengan total 25 entri. Langkah pertama dalam

proses pengolahan data adalah mengimpor data terkait COVID-19 yang disimpan dalam format Excel (xlsx) ke dalam aplikasi RapidMiner. Untuk melakukan impor ini, pengguna dapat memilih opsi “impor data” atau menggunakan wizard konfigurasi pada operator Read Excel yang berfungsi untuk membaca data yang telah diunggah. Berikut ini adalah gambar yang menunjukkan pemilihan data.

Gambar 1. Tampilan Input Data Covid 19

2. Pemrosesan Sistem

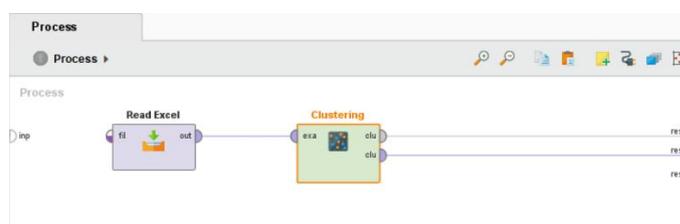
Pada tahap ini akan dijelaskan langkah-langkah penggunaan algoritma K-Means di RapidMiner menggunakan data yang diimpor. Langkah pertama diawali dengan import file excel berisi data yang akan diolah, seperti terlihat pada gambar di bawah ini.



Gambar 2. Tahap Pengolahan 1
Selanjutnya klik opsi Clustering and Segmentation dan pilih metode K-Means untuk melakukan pengelompokan. Data harus terhubung terlebih dahulu dengan operator Clustering agar proses dapat berjalan.



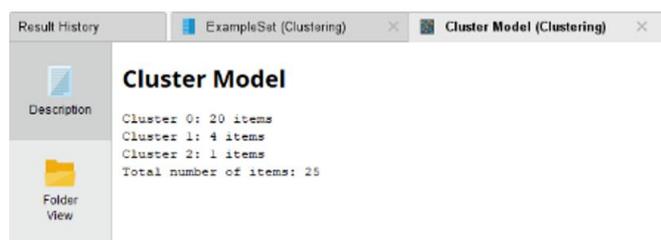
Gambar 3. Tahap Pengolahan 2
Setelah itu sambungkan operator Read Excel dengan operator Clustering seperti terlihat pada gambar di bawah ini.



Gambar 4. Tahap Pengolahan 3

3. Keluaran Sistem RapidMiner

Untuk mendapatkan hasil pengelompokan, langkah selanjutnya adalah klik ikon panah berwarna biru yang terletak di tengah atas atau pada Toolbar. Langkah ini akan menampilkan hasil akhir dan merupakan tahap akhir dalam penggunaan tool RapidMiner, seperti terlihat pada gambar di bawah ini.

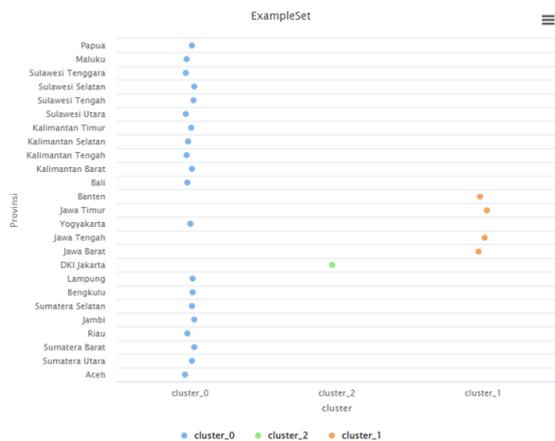


Gambar 5. Nilai Cluster Model RapidMiner

Informasi:

- a) Jumlah Cluster 0 (tinggi) sebanyak 4 item
- b) Jumlah Cluster 2 (sedang) sebanyak 1 item
- c) Jumlah Cluster 1 (rendah) sebanyak 20 item

Dengan demikian, hasil pengelompokan yang dihasilkan oleh RapidMiner dapat dilihat pada gambar berikut.



Gambar 6. Hasil Clustering RapidMiner

Berdasarkan gambar. Terlihat kelompok tinggi berjumlah 1 item, kelompok sedang berjumlah 2 item, sedangkan kelompok rendah berjumlah 0 item.

Berdasarkan penjelasan tahapan penggunaan dan hasil yang telah ditampilkan, dapat dijelaskan bahwa terdapat hubungan antara hasil perhitungan manual menggunakan algoritma K-Means dengan hasil yang diperoleh dengan menggunakan alat seperti Rapid Miner. Hasil perhitungan manual menggunakan algoritma K-Means jika dilakukan dengan menggunakan aplikasi seperti Microsoft Office Excel 2007 menunjukkan kemiripan dengan hasil yang dihasilkan oleh RapidMiner. Hal ini menunjukkan bahwa hasil perhitungan manual dapat diintegrasikan dengan aplikasi RapidMiner untuk analisa lebih lanjut.

Row No.	id	Provinsi	cluster	No	Jumlah Kas...	Jumlah Sem...	Jumlah Men...
1	1	Aceh	cluster_0	1	5000	4800	200
2	2	Sumatera Lt...	cluster_0	2	8000	7500	500
3	3	Sumatera Ba...	cluster_0	3	6000	5800	200
4	4	Riau	cluster_0	4	7000	6800	200
5	5	Jambi	cluster_0	5	3000	2900	100
6	6	Sumatera Sel...	cluster_0	6	10000	9500	500
7	7	Bengkulu	cluster_0	7	2000	1900	100
8	8	Lampung	cluster_0	8	7500	7000	500
9	9	DKI Jakarta	cluster_2	9	120000	115000	5000
10	10	Jawa Barat	cluster_1	10	80000	76000	4000
11	11	Jawa Tengah	cluster_1	11	60000	55000	5000
12	12	Yogyakarta	cluster_0	12	15000	14000	1000
13	13	Jawa Timur	cluster_1	13	70000	65000	5000
14	14	Banten	cluster_1	14	40000	38000	2000
15	15	Bali	cluster_0	15	20000	19000	1000
16	16	Kalimantan B...	cluster_0	16	5000	4800	200

Figure 7. RapidMiner Calculation Data Display

SIMPULAN DAN SARAN (PENUTUP)

Hasil analisis menunjukkan bahwa algoritma K-Means merupakan alat yang efektif untuk mengelompokkan wilayah berdasarkan tingkat prevalensi COVID-19. Mengelompokkan data ke dalam tiga kategori utama prevalensi: tinggi, sedang, dan rendah. Pengelompokan ini memberikan wawasan yang lebih mendalam mengenai perbedaan karakteristik sebaran antar wilayah dan mendukung pengembangan kebijakan yang lebih spesifik dan tepat sasaran.

Penggunaan RapidMiner sebagai alat analisis telah terbukti memudahkan pengolahan data yang besar dan kompleks. Hasil clustering yang dihasilkan oleh RapidMiner konsisten dengan perhitungan manual menggunakan algoritma K-Means, yang menunjukkan keandalan alat ini untuk analisis yang lebih efisien. Hasil ini juga membuka peluang penerapan yang lebih luas, seperti mengembangkan strategi pembendungan yang lebih efektif dalam mengalokasikan sumber daya medis dan memprioritaskan area berisiko tinggi. Selain itu, pendekatan berbasis data mining ini memiliki potensi besar untuk mendukung sistem peringatan dini terhadap lonjakan infeksi COVID-19 di masa depan. Pendekatan yang dapat menganalisis data dalam skala besar dan kompleks ini memberikan kontribusi yang signifikan dalam memahami pola sebaran kasus dan membantu mengembangkan kebijakan yang lebih efisien dan berbasis bukti.

DAFTAR PUSTAKA

- Sitohang, R., & Dewi, A. S. (2016). *Data mining: Teori dan aplikasi*. Informatika.
- Suyanto, S. (2018). *Machine learning dan data mining: Mengolah data menjadi informasi berharga*. Andi.
- Girsang, S. R., & Simarmata, J. (2017). *Algoritma dan pemrograman data mining*. Elex Media Komputindo.
- Kusrini, A., & Luthfi, E. (2009). *Algoritma data mining*. Andi.

- Pradana, M. A., & Sihombing, R. (2020). *Data mining dan aplikasinya: Teori dan implementasi*. Informatika.
- Wibisono, H. (2021). *Pengolahan data dengan data mining menggunakan algoritma K-Means*. Bumi Aksara.
- Siregar, A. M. (n.d.). *Data mining: Pengolahan data menjadi informasi dengan Rapidminer*. Grup CV Kekata. Retrieved from <http://www.rapidminer.com>
- Tan, P.-N., Steinbach, M., & Kumar, V. (2006). *Introduction to data mining*. Pearson Addison-Wesley.
- Pradana, M. A., & Sihombing, R. (2020). *Data mining dan aplikasinya: Teori dan implementasi*. Informatika.
- Sari, M. R. M. (2024). *Metode clustering analisis data penyakit berbahaya*. Serasi ssMedia Teknologi. Retrieved from <http://www.serasi.com>