

Analisis Kluster Persebaran COVID-19 di Sumatera Utara Menggunakan Metode K-Means dengan Visualisasi Berbasis Google Colab

Daniel Rumahorbo¹ Gerhard Hasangapon Parapat² Josua Deo Tampubolon³ Aurella Nasuiton⁴ Fanny Ramdhani⁵ Arnita⁶

Email: danielrumahorbo2004@gmail.com¹ parapatgerhard@gmail.com² josuatampubolon30@gmail.com³ aurelakhairinasution@gmail.com⁴ arnita@unimed.ac.id⁶
Program Studi Ilmu Komputer, Universitas Negeri Medan, Kabupaten Deli Serdang, Provinsi Sumatera Utara, Indonesia^{1,2,3,4,5,6}

Abstrak

Pandemi COVID-19 telah memberikan dampak signifikan terhadap berbagai sektor, termasuk kesehatan, sosial, dan ekonomi. Provinsi Sumatera Utara sebagai salah satu daerah yang cukup padat penduduk menghadapi tantangan besar dalam menangani persebaran kasus COVID-19 yang tidak merata di setiap kabupaten/kota. Penelitian ini bertujuan untuk mengelompokkan wilayah berdasarkan tingkat risiko penyebaran COVID-19 menggunakan metode K-Means Clustering dengan variabel: jumlah penduduk, kepadatan penduduk, jumlah kasus positif, sembuh, meninggal, dan jumlah rumah sakit umum. Data diperoleh dari Badan Pusat Statistik dan Kementerian Kesehatan Republik Indonesia. Hasil analisis menghasilkan yaitu Zona Hijau (C1), Zona Kuning (C2), Zona Oranye (C3), dan Zona Merah (C4). Pengelompokan ini memberikan gambaran strategis terhadap distribusi layanan kesehatan dan perencanaan intervensi kebijakan di tiap wilayah. Visualisasi dilakukan menggunakan Google Colab untuk menampilkan peta klusterisasi kabupaten/kota. Penelitian ini diharapkan dapat menjadi dasar pertimbangan dalam perencanaan kebijakan penanganan pandemi di masa mendatang.

Kata Kunci: COVID-19, K-Means, Sumatera Utara, Data Wilayah, Clustering

Abstract

The COVID-19 pandemic has had a significant impact on various sectors, including health, social, and economic. North Sumatra Province as one of the densely populated areas faces major challenges in dealing with the uneven distribution of COVID-19 cases in each district/city. This study aims to group regions based on the risk level of the spread of COVID-19 using the K-Means Clustering method with the following variables: population, population density, number of positive cases, recoveries, deaths, and number of general hospitals. Data were obtained from the Central Statistics Agency and the Ministry of Health of the Republic of Indonesia. The results of the analysis produced the Green Zone (C1), Yellow Zone (C2), Orange Zone (C3), and Red Zone (C4). This grouping provides a strategic overview of the distribution of health services and policy intervention planning in each region. Visualization was carried out using Google Colab to display a map of district/city clustering. This study is expected to be the basis for consideration in planning policies for handling the pandemic in the future.

Keywords: COVID-19, K-Means, North Sumatra, Regional Data, Clustering



This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

PENDAHULUAN

Pandemi Coronavirus Disease 2019 (COVID-19) yang muncul pertama kali di Wuhan, Tiongkok pada akhir tahun 2019, telah menyebar ke seluruh dunia dengan dampak yang luas di berbagai aspek kehidupan. Di Indonesia, penyebaran virus ini terjadi dengan cepat dan tidak merata di setiap wilayah, termasuk di Provinsi Sumatera Utara yang memiliki 33 kabupaten dan kota. Wilayah ini mengalami fluktuasi jumlah kasus yang signifikan, mulai dari angka infeksi hingga kematian. Persebaran kasus di Sumatera Utara menunjukkan variasi yang cukup besar, sehingga diperlukan pendekatan analitis untuk memahami karakteristik penyebaran virus dan perbedaan antar wilayah tersebut (Nasution et al., 2023). Pentingnya kesehatan sebagai fondasi

kehidupan manusia menuntut perhatian khusus dalam penanganan pandemi. Pandemi ini tidak hanya memberikan tekanan besar pada sektor kesehatan, tetapi juga mengganggu aktivitas sosial, pendidikan, dan ekonomi masyarakat. Berdasarkan data yang dikumpulkan oleh Dinas Kesehatan Provinsi Sumatera Utara, terdapat kebutuhan mendesak untuk melakukan klasifikasi risiko wilayah, agar sumber daya kesehatan dapat dikelola secara lebih tepat sasaran (Nasution et al., 2023).

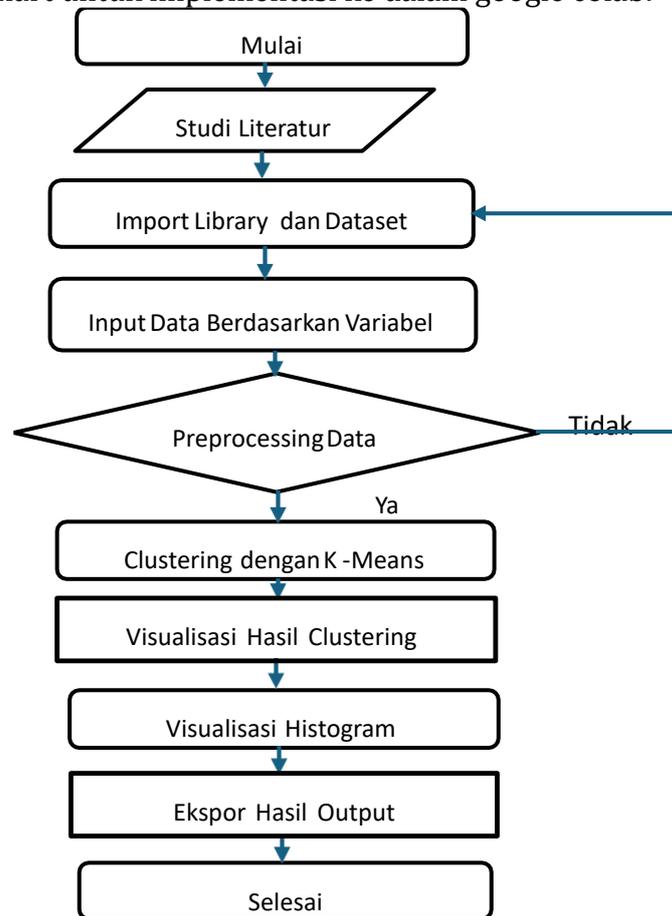
Salah satu pendekatan yang dapat digunakan untuk memahami pola penyebaran COVID-19 secara geografis adalah metode clustering, khususnya K-Means Clustering. Metode ini merupakan teknik partisi data yang membagi objek ke dalam sejumlah cluster berdasarkan kemiripan karakteristik. Dalam konteks ini, objek yang dimaksud adalah kabupaten/kota di Sumatera Utara, dan karakteristik yang digunakan mencakup jumlah penduduk, kepadatan penduduk, jumlah kasus positif, pasien sembuh, dan jumlah kematian (Damanik et al., 2021). Algoritma K-Means untuk mengelompokkan daerah berdasarkan tingkat penyebaran COVID-19 ke dalam tiga kategori: tinggi, sedang, dan rendah (Damanik et al., 2021). Pendekatan ini efektif dalam memberikan gambaran kondisi epidemiologis tiap daerah, sehingga mendukung pengambilan kebijakan berbasis data (Nasution et al., 2023). Mereka juga menunjukkan bahwa variabel jumlah penduduk memiliki pengaruh paling besar dalam pembentukan klaster. Melalui metode K-Means, masing-masing kabupaten/kota dapat dikelompokkan berdasarkan tingkat risiko COVID-19, sehingga pemerintah daerah memiliki informasi yang lebih akurat dalam menentukan prioritas intervensi. Contohnya, daerah yang masuk dalam klaster dengan tingkat kematian tinggi dapat menjadi sasaran utama untuk peningkatan layanan kesehatan, penguatan edukasi protokol kesehatan, serta distribusi vaksin dan logistik medis lainnya (Damanik et al., 2021). Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengelompokkan wilayah kabupaten/kota di Provinsi Sumatera Utara ke dalam zona risiko penyebaran COVID-19 berdasarkan data tahun 2021 menggunakan algoritma K-Means Clustering. Diharapkan hasil pengelompokan ini dapat menjadi dasar ilmiah dalam perumusan strategi mitigasi pandemi di masa mendatang.

METODE PENELITIAN

Penelitian ini menerapkan metode *K-Means Clustering* untuk melakukan pengelompokan wilayah di Provinsi Sumatera Utara berdasarkan persebaran kasus COVID-19. Tahapan dimulai dengan pengumpulan data sekunder dari sumber resmi seperti Badan Pusat Statistik (BPS) dan Kementerian Kesehatan Republik Indonesia. Data yang dikumpulkan mencakup variabel-variabel seperti jumlah penduduk, jumlah kasus positif, jumlah pasien sembuh, dan jumlah kematian akibat COVID-19 pada masing-masing kabupaten/kota. Selanjutnya dilakukan pra-pemrosesan data (*preprocessing*) yang meliputi pembersihan data dari nilai kosong atau tidak konsisten, seleksi atribut relevan, serta normalisasi agar semua variabel berada dalam skala yang setara. Prosedur ini mengikuti prinsip-prinsip dalam proses *Knowledge Discovery in Database (KDD)* seperti *data cleansing*, *data transformation*, dan *data integration*, sebagaimana dijelaskan dalam penelitian sebelumnya yang juga mengelompokkan wilayah di Sumatera Utara berdasarkan tingkat sebaran COVID-19 menggunakan K-Means (Damanik et al., 2021).

Untuk menentukan jumlah klaster (k) yang optimal, digunakan metode Elbow. Dalam metode ini, analisis dilakukan terhadap nilai *Within-Cluster Sum of Squares (WCSS)* pada berbagai kemungkinan jumlah klaster. Titik tekuk pada grafik Elbow menunjukkan jumlah klaster optimal. Berdasarkan hasil visualisasi ini, ditetapkan jumlah klaster sebanyak empat yang masing-masing menggambarkan tingkat risiko zona: zona hijau (sangat rendah), zona kuning (rendah), zona oranye (sedang), dan zona merah (tinggi). Implementasi algoritma dan visualisasi dilakukan menggunakan Google Colaboratory dengan bantuan pustaka

pemrograman Python untuk keperluan analisis data dan pemetaan kluster. Hasil dari klusterisasi ini kemudian dianalisis dan disajikan dalam bentuk visual seperti *histogram* guna mempermudah interpretasi dan pengambilan keputusan berbasis data. Pendekatan ini diharapkan mampu menghasilkan pemetaan risiko wilayah yang akurat dan dapat menjadi dasar pertimbangan dalam pengambilan kebijakan penanganan pandemi di tingkat lokal. Berikut adalah flowchart untuk implementasi ke dalam google colab.



Gambar 1. Flowchart

HASIL PENELITIAN DAN PEMBAHASAN

Penggunaan masker sebagai langkah preventif dalam menghadapi pandemi COVID-19 telah terbukti menjadi salah satu strategi paling efektif dalam mengurangi risiko penularan virus. Masker berfungsi ganda: melindungi individu yang memakainya dari paparan langsung partikel virus dan pada saat yang sama mencegah individu yang telah terinfeksi menyebarkan virus kepada orang lain (Masriani 2022). Efektivitas penggunaan masker ini sangat bergantung pada konsistensi dan kesadaran masyarakat dalam menerapkannya. Oleh karena itu, kampanye edukasi yang berkelanjutan mengenai pentingnya protokol kesehatan seperti pemakaian masker, menjaga jarak, dan mencuci tangan perlu terus digencarkan agar dampak penyebaran virus dapat diminimalisir. Dalam konteks pengolahan data epidemiologis seperti COVID-19, data mining memainkan peran krusial. Data mining atau penambangan data merupakan proses eksplorasi dan analisis terhadap kumpulan data besar guna menemukan pola-pola tersembunyi yang bermanfaat untuk pengambilan keputusan. Dalam penelitian ini, algoritma K-Means Clustering digunakan sebagai teknik utama dalam proses data mining untuk mengelompokkan kabupaten/kota di Sumatera Utara berdasarkan tingkat risiko penyebaran COVID-19. Penggunaan algoritma ini memungkinkan analisis data multivariat seperti jumlah kasus positif,

angka kesembuhan, kematian, hingga jumlah fasilitas kesehatan secara efisien, sehingga memberikan gambaran menyeluruh tentang kondisi tiap wilayah.

Selain metode analisis, peran database sebagai infrastruktur penyimpanan data menjadi sangat penting. database, atau dikenal juga sebagai data warehouse, adalah gudang informasi terstruktur yang memungkinkan data disimpan, dikelola, dan diakses secara sistematis (Hardiani 2022). Dalam penelitian ini, data yang diperoleh dari BPS dan Kementerian Kesehatan diintegrasikan dan disimpan dengan pendekatan manajemen database yang baik agar memudahkan dalam tahap preprocessing dan analisis lebih lanjut. Pengelolaan database yang akurat menjadi kunci dalam menjamin validitas hasil klasterisasi dan kesesuaian data dengan kondisi lapangan. Pendekatan data mining berbasis database mampu meningkatkan efisiensi dalam pengambilan keputusan, karena proses penyimpanan, integrasi, dan transformasi data dilakukan secara terpadu (Anggiat 2022). Hal ini terlihat dalam implementasi penelitian yang dilakukan melalui Google Colaboratory, di mana seluruh proses mulai dari impor data, preprocessing, perhitungan K-Means, hingga visualisasi hasil dalam bentuk klaster dan histogram dilakukan secara otomatis dan terstruktur. Dengan demikian, integrasi antara metode K-Means, database, dan data mining secara keseluruhan memberikan hasil yang tidak hanya akurat secara analitik, tetapi juga praktis dalam mendukung kebijakan penanganan pandemi di wilayah Sumatera Utara.

Penentuan empat cluster awal pada analisis K-Means ini dilakukan secara manual berdasarkan pengamatan terhadap sebaran data kasus COVID-19 dari 33 kabupaten/kota di Sumatera Utara. Setiap cluster atau zona (Zona Hijau, Kuning, Oranye, dan Merah) merepresentasikan tingkat keparahan penyebaran COVID-19 di suatu wilayah, yang dilihat dari jumlah total kasus, jumlah sembuh, dan jumlah meninggal. Nilai pusat awal (centroid) dari masing-masing zona diperoleh dengan menghitung rata-rata dari beberapa kabupaten/kota yang dianggap mewakili karakteristik zona tersebut. Misalnya, Zona Hijau yang memiliki jumlah kasus sangat rendah dihitung dari rata-rata data wilayah seperti Nias, Gunung Sitoli, dan Pakpak Barat. Zona Kuning dihitung dari kabupaten dengan jumlah kasus menengah seperti Asahan, Karo, dan Tebing Tinggi. Zona Oranye berasal dari wilayah dengan penyebaran cukup tinggi seperti Deli Serdang, Serdang Bedagai, dan Tapanuli Utara. Sementara itu, Zona Merah dihitung dari kota-kota besar dengan angka kasus tertinggi seperti Medan, Binjai, dan Pematang Siantar. Dengan pendekatan ini, diperoleh nilai centroid awal untuk masing-masing zona, yakni: Zona Hijau (Kasus: 1.230, Sembuh: 1.209, Meninggal: 3), Zona Kuning (33.983, 32.802, 1.181), Zona Oranye (9.606, 9.036, 570), dan Zona Merah (73.694, 72.517, 1.177). Nilai-nilai inilah yang kemudian digunakan sebagai pusat awal pada proses pengelompokan data menggunakan algoritma K-Means.

Clustering data dan analisis data. Data kasus COVID-19 tahun 2021 yang dikumpulkan dari berbagai kabupaten dan kota di Provinsi Sumatera Utara, dilakukan analisis pengelompokan menggunakan metode *K-Means Clustering*. Pada iterasi pertama, data berhasil dikelompokkan ke dalam empat cluster atau zona, yaitu Zona Hijau (C1), Zona Kuning (C2), Zona Oranye (C3), dan Zona Merah (C4). Berdasarkan hasil pengelompokan, Zona Hijau memiliki total 1.230 kasus konfirmasi, 1.209 sembuh, dan 3 kematian; Zona Kuning mencatat 33.983 kasus konfirmasi, 32.802 sembuh, dan 1.181 kematian; Zona Oranye memiliki 9.606 kasus konfirmasi, 9.036 sembuh, dan 570 kematian; sedangkan Zona Merah mencatat angka tertinggi dengan 73.694 kasus konfirmasi, 72.517 sembuh, dan 1.177 kematian. Hasil analisis menunjukkan bahwa mayoritas kabupaten/kota berada pada Zona Kuning, sedangkan hanya sebagian kecil daerah masuk dalam Zona Oranye dan Zona Merah. Zona Hijau menunjukkan tingkat penyebaran yang sangat rendah, sedangkan Zona Merah mencerminkan tingkat penyebaran dan dampak COVID-19 yang paling tinggi. Proses pengelompokan dilakukan

berdasarkan perhitungan jarak terpendek dari masing-masing data ke pusat cluster, dan hasil pengelompokan tersebut ditampilkan dalam tabel perhitungan jarak pusat cluster pada iterasi pertama. Tabel iterasi pertama pada proses klasterisasi K-Means disusun berdasarkan perhitungan jarak Euclidean antara setiap data kabupaten/kota dengan pusat cluster (centroid) awal yang telah ditentukan sebelumnya. Setiap centroid mewakili satu zona, yaitu: C1 (Asahan), C2 (Medan), C3 (Deli Serdang), dan C4 (Binjai), dengan nilai masing-masing yang menggambarkan karakteristik jumlah kasus positif, sembuh, dan meninggal akibat COVID-19. Dalam proses ini, setiap kabupaten/kota dihitung jaraknya ke keempat pusat cluster tersebut. Jarak yang paling kecil menunjukkan bahwa kabupaten/kota tersebut lebih dekat (dalam konteks data) ke cluster tersebut dibanding cluster lainnya. Berdasarkan hasil perhitungan jarak, masing-masing daerah kemudian dimasukkan ke dalam cluster dengan jarak terpendek. Dari hasil tersebut, diperoleh pengelompokan awal sebagai berikut: 28 kabupaten/kota tergabung dalam Cluster C1, yaitu cluster dengan karakteristik jumlah kasus yang rendah hingga sedang; 3 kabupaten/kota masuk ke Cluster C2 yang merepresentasikan wilayah dengan kasus tinggi seperti Medan; masing-masing 1 kabupaten/kota masuk ke Cluster C3 dan C4, yaitu Deli Serdang dan Binjai, yang juga memiliki karakteristik penyebaran kasus cukup tinggi namun dengan pola yang berbeda dibanding Medan. Jumlah total kasus dalam tiap cluster juga dihitung untuk menunjukkan tingkat penyebaran dalam masing-masing kelompok. Proses ini menjadi tahap awal dalam iterasi algoritma K-Means sebelum pusat cluster diperbarui pada iterasi selanjutnya.

Pada tahap iterasi kedua dalam metode klasterisasi K-Means diperoleh berdasarkan hasil pembaruan pusat cluster (centroid) dari iterasi sebelumnya. Setelah data 33 kabupaten/kota di Sumatera Utara dikelompokkan ke dalam empat cluster berdasarkan jarak Euclidean terpendek terhadap centroid awal (Asahan, Medan, Deli Serdang, dan Binjai), langkah berikutnya adalah menghitung ulang posisi pusat cluster baru. Perhitungan ini dilakukan dengan mengambil nilai rata-rata dari jumlah kasus konfirmasi positif, jumlah sembuh, dan jumlah meninggal dari seluruh anggota dalam masing-masing cluster hasil iterasi pertama. Hasilnya dirangkum dalam Tabel 5, yang menunjukkan bahwa cluster 1 didominasi oleh wilayah dengan kasus menengah ke rendah (rata-rata positif 1.842, sembuh 1.752, meninggal 42); cluster 2 adalah wilayah dengan jumlah kasus sangat tinggi, didominasi oleh Medan (positif 26.021, sembuh 25.655, meninggal 375); cluster 3 memiliki karakteristik sedang-tinggi yang diwakili oleh Deli Serdang; sementara cluster 4 berisi Binjai dengan karakteristik cukup tinggi namun dengan profil kematian lebih rendah dari cluster lainnya. Setelah pusat cluster baru diperoleh, dilakukan perhitungan jarak Euclidean ulang antara tiap kabupaten/kota terhadap keempat centroid baru ini. Perhitungan jarak tersebut disajikan dalam Tabel 6 dan didasarkan pada rumus jarak Euclidean tiga dimensi, yakni:

$$d = \sqrt{(x_1 - c_1)^2 + (x_2 - c_2)^2 + (x_3 - c_3)^2}$$

Di mana x_1 , x_2 , x_3 adalah nilai kasus positif, sembuh, dan meninggal dari masing-masing kabupaten/kota, sedangkan c_1 , c_2 , c_3 adalah nilai rata-rata dari pusat cluster yang baru. Setiap kabupaten/kota kemudian kembali dikelompokkan ke dalam cluster terdekat berdasarkan hasil perhitungan jarak baru tersebut. Proses ini menunjukkan dinamika perpindahan cluster apabila ada perubahan signifikan dari komposisi centroid, serta bertujuan untuk meminimalkan variasi dalam masing-masing cluster. Dengan demikian, iterasi kedua ini memberikan hasil klasterisasi yang lebih representatif berdasarkan pola sebaran COVID-19 di setiap daerah. Langkah ini mengacu pada prinsip utama algoritma K-Means seperti dijelaskan oleh peneliti Indonesia, yang menyatakan bahwa proses iteratif pada K-Means bertujuan untuk meminimalkan nilai within-cluster sum of squares (WCSS) agar tercapai hasil klasterisasi yang optimal dan konvergen (Adiyarta et al. 2020)

Berdasarkan hasil pengelompokan menggunakan metode K-Means, wilayah kabupaten/kota di Provinsi Sumatera Utara dapat dikelompokkan ke dalam empat zona berdasarkan tingkat persebaran dan dampak COVID-19, yaitu zona hijau (sangat rendah), zona kuning (rendah), zona oranye (sedang), dan zona merah (tinggi). Zona hijau (kelompok C4) terdiri dari wilayah dengan kasus COVID-19 yang sangat rendah hingga nihil, kepadatan penduduk yang rendah hingga sedang, serta fasilitas kesehatan yang terbatas, seperti Nias, Nias Barat, dan Simalungun. Zona kuning dan oranye (mayoritas kelompok C3) mencakup daerah dengan tingkat kasus sedang hingga cukup tinggi, serta memiliki jumlah rumah sakit yang lebih memadai, seperti Asahan, Dairi, dan Deli Serdang. Sementara itu, zona merah (juga termasuk dalam kelompok C4) mencakup kota-kota besar seperti Medan, Binjai, dan Tebing Tinggi yang memiliki kepadatan penduduk sangat tinggi dan jumlah kasus positif serta kematian yang signifikan. Pengelompokan ini memberikan gambaran tingkat risiko tiap daerah dan dapat menjadi dasar dalam perumusan kebijakan penanganan COVID-19 yang lebih tepat sasaran.

KESIMPULAN

Penelitian ini berhasil menganalisis persebaran kasus COVID-19 di Provinsi Sumatera Utara menggunakan metode K-Means Clustering, yang mengelompokkan wilayah ke dalam empat zona risiko (Hijau, Kuning, Oranye, dan Merah) berdasarkan variabel jumlah kasus, kepadatan penduduk, dan fasilitas kesehatan. Hasilnya menunjukkan bahwa wilayah urban seperti Medan dan Binjai masuk dalam Zona Merah dengan kasus tertinggi, sedangkan daerah rural seperti Nias termasuk Zona Hijau dengan kasus minimal. Temuan ini menegaskan bahwa faktor demografi, geografi, dan kapasitas kesehatan berpengaruh signifikan terhadap penyebaran COVID-19. Untuk pengembangan penelitian selanjutnya, disarankan melakukan normalisasi data (Z-score) dan pemilihan centroid yang lebih optimal (misalnya dengan Elbow Method) guna meningkatkan akurasi clustering. Penambahan variabel seperti tingkat vaksinasi, mobilitas penduduk, dan kebijakan lokal juga dapat memperdalam analisis. Dari sisi kebijakan, pemerintah daerah sebaiknya memprioritaskan alokasi sumber daya kesehatan di Zona Merah dan Oranye, sekaligus meningkatkan pengawasan di Zona Hijau untuk mengantisipasi potensi lonjakan kasus. Implementasi sistem pemantauan real-time berbasis data juga dapat menjadi solusi jangka panjang untuk penanganan pandemi yang lebih efektif. Dengan demikian, pendekatan berbasis data-driven policy ini diharapkan dapat mendukung strategi penanganan COVID-19 yang lebih tepat sasaran di Sumatera Utara.

DAFTAR PUSTAKA

- A. M. Rahman and S. N. Hidayat, "Penerapan Metode K-Means untuk Klasterisasi Wilayah Berdasarkan Jumlah Kasus COVID-19 di Indonesia," *Jurnal Sains dan Teknologi Informasi*, vol. 6, no. 1, pp. 45–52, 2022.
- Anggiat, *Penerapan Data Mining dalam Pengolahan Database untuk Pengambilan Keputusan*. Medan: Penerbit Teknologi Data, 2022.
- D. L. Fithri, F. Nugraha, and N. Latifah, "Visualisasi Data Mining Cluster Covid-19 Menggunakan K-means dan Aplikasi Google Colaboratory," *Jurnal Unitek*, vol. 15, no. 2, pp. 159–165, 2022.
- I Adiyarta, E. Mulyani, and R. Prasetyo, "Clustering Pengelompokan Penyebaran COVID-19 di Indonesia dengan Algoritma K-Means," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 3, pp. 1234–1240, 2020.
- J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. San Francisco: Morgan Kaufmann, 2012.
- K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010.

- M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, "Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster," *International Journal of Advanced Science and Technology*, vol. 38, pp. 111–121, 2020.
- M. Fauzan and H. N. Wahyudi, "Pemanfaatan Data Clustering dalam Penanganan Wabah COVID-19 di Indonesia Menggunakan K-Means dan DBSCAN," *Jurnal Media Informatika Budidarma*, vol. 5, no. 1, pp. 77–84, 2021.
- M. Masriani, *Efektivitas Penggunaan Masker dalam Mencegah Penyebaran COVID-19*. Banda Aceh: Pustaka Kesehatan, 2022.
- N. Putri and D. Saputra, "Visualisasi Data Epidemiologi COVID-19 Menggunakan Google Colab dan Metode Elbow-KMeans," *Prosiding Seminar Nasional Teknologi Informasi dan Komputer (SENTIKA)*, pp. 234–240, 2022.
- R. Hardiani, *Manajemen Basis Data dan Implementasinya dalam Dunia Digital*. Jakarta: Cipta Informatika, 2022.
- S. R. Nasution, R. F. Sari, and R. Widyasari, "Analisis Klaster dengan Metode K-Means pada Informasi dan Komputer Bln-Thn, Vol.XX, No.X, hal. XX-XX Penyebaran Kasus Covid-19 Berdasarkan Kabupaten/Kota di Sumatera Utara," 2023.
- T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York: Springer, 2009.
- W. S. Lestari and T. Hidayat, "Analisis Penerapan Data Mining Menggunakan K-Means Clustering untuk Segmentasi Wilayah Penyebaran COVID-19," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 7, no. 2, pp. 137–144, 2022.
- Y. F. S. Y. Damanik, Sumarno, I. Gunawan, D. Hartama, and I. O. Kirana, "Penerapan Data Mining untuk Pengelompokan Penyebaran Covid-19 di Sumatera Utara Menggunakan Algoritma K- Means," 2021.