

Analisis Klaster Dampak Kesehatan Akibat Pandemi Covid-19 Menggunakan Metode *K-Means* dan C4.5 *Cluster Analysis of Health Impacts Due to the Covid-19 Pandemic Using the K-Means and C4.5 Methods*

Karina Anindita, Anak Agung Ngurah Bagus Adhitya Wirakusuma, Kristopher May Pamudji,
Dini Wurnaning Budi, I Komang Reka Yudha Putra*
Universitas Bumigora, Mataram, Indonesia
Email: karina.anindita@universitasbumigora.ac.id

Artikel History

Submit: 02 September 2025 *Revisi:* 29 Oktober 2025 *Diterima:* 30 Oktober 2025

Abstrak

Pandemi COVID-19 berdampak luas terhadap sistem kesehatan dan kondisi sosial-ekonomi masyarakat, termasuk di Provinsi Nusa Tenggara Barat. Ketimpangan akses layanan, prevalensi penyakit penyerta, dan disparitas ekonomi menyebabkan tingkat risiko antarwilayah berbeda. Tujuan penelitian ini adalah menganalisis kluster dampak kesehatan akibat pandemi COVID-19 di Nusa Tenggara Barat. Metode penelitian menggunakan metode *K-Means* dan *Decision Tree* C4.5 dengan pendekatan CRISP-DM. Data sekunder diperoleh dari Rumah Sakit Umum Daerah Provinsi Nusa Tenggara Barat, Profil Kesehatan Nusa Tenggara Barat, Badan Pusat Statistik 2022-2023, Riset Kesehatan Dasar 2021, dan WHO. Variabel mencakup kasus COVID-19 per 100.000 penduduk, tingkat vaksinasi, rasio fasilitas kesehatan, prevalensi hipertensi dan diabetes, serta tingkat kemiskinan. Hasil *K-Means* menunjukkan tiga kluster risiko: tinggi (Mataram, Lombok Barat), sedang (Lombok Tengah, Lombok Timur), dan rendah (Sumbawa, Dompu, Bima). Model C4.5 menghasilkan akurasi 80.67% dengan usia dan komorbid sebagai faktor dominan. Kesimpulan temuan ini menegaskan potensi analisis data mining sebagai dasar kebijakan mitigasi dan penguatan ketahanan layanan kesehatan daerah pascapandemi.

Kata Kunci: COVID-19; C4.5; data mining; *K-Means*; Klasterisasi.

Abstract

The COVID-19 pandemic has had a broad impact on the health system and the socio-economic conditions of communities, including in the Province of West Nusa Tenggara. Inequalities in healthcare access, the prevalence of comorbid diseases, and economic disparities have resulted in varying levels of risk across regions. This study aims to analyze clusters of health impacts resulting from the COVID-19 pandemic in West Nusa Tenggara. The research employed the *K-Means* and *C4.5 Decision Tree* methods with a CRISP-DM approach. Secondary data were obtained from the West Nusa Tenggara Provincial General Hospital, the West Nusa Tenggara Health Profile, Statistics Indonesia (BPS) 2022-2023, the 2021 Basic Health Research (Riskesdas), and the World Health Organization (WHO). The variables included COVID-19 cases per 100,000 population, vaccination rates, healthcare facility ratios, prevalence of hypertension and diabetes, and poverty rates. The *K-Means* results identified three risk clusters: high (Mataram, West Lombok), moderate (Central Lombok, East Lombok), and low (Sumbawa, Dompu, Bima). The *C4.5* model achieved an accuracy of 80.67%, with age and comorbidities emerging as dominant factors. These findings highlight the potential of data mining analysis as a foundation for post-pandemic mitigation policies and strengthening regional health system resilience.

Keywords: COVID-19, C4.5; data mining; *K-Means*, Clustering.

Copyright ©2025 by Authors. This is an open access article under the CC-BY-SA license.



*Penulis Korespondensi:

Karina Anindita Universitas Bumigora, Mataram, Indonesia Email: karina.anindita@universitasbumigora.ac.id
Telp/Hp: 085720814858

Cara Sitasi (IEEE Citation Style): K. Anindita, A. A. N. B. A. Wirakusuma, K. M. Pamudji, D. W. Budi, and I. K. R. Y. Putra, "Analisis Klaster Dampak Kesehatan Akibat Pandemi Covid-19 Menggunakan Metode *K-Means* dan C4.5," *Nutriology: Jurnal Pangan, Gizi, Kesehatan*, vol. 6, no. 2, p. 191-198, 2025, <https://doi.org/10.30812/nutriology.v6i2.5323>

PENDAHULUAN

Pandemi COVID-19 yang muncul pada akhir 2019 di Wuhan, Tiongkok, telah memberikan dampak multi-dimensi terhadap kesehatan masyarakat, ekonomi, dan sistem sosial global. Menurut data World Health Organization, hingga akhir tahun 2023 terdapat lebih dari 770 juta kasus terkonfirmasi dan lebih dari 7 juta kematian di seluruh dunia [1]. Dampak tersebut tidak hanya pada mortalitas dan morbiditas langsung, tetapi juga mengganggu sistem kesehatan, memperburuk ketimpangan sosial-ekonomi, serta menurunkan akses masyarakat terhadap layanan kesehatan dasar, termasuk imunisasi, deteksi dini penyakit tidak menular, dan gizi masyarakat [2]. Di tingkat nasional, Indonesia mencatat lebih dari 6,7 juta kasus COVID-19 dan sekitar 161 ribu kematian hingga awal 2023 [3]. Selain itu, pandemi memicu peningkatan angka kemiskinan sementara (dari 9,2% menjadi 10,2% pada 2020 menurut BPS), serta mengganggu pelaksanaan layanan kesehatan esensial seperti program imunisasi dasar lengkap dan layanan gizi balita [4].

Secara khusus, Provinsi Nusa Tenggara Barat (NTB) menghadapi tantangan tersendiri. Berdasarkan Profil Kesehatan Provinsi Nusa Tenggara Barat tahun 2023, provinsi ini memiliki rasio tenaga medis per 100.000 penduduk sebesar 37, masih di bawah rata-rata nasional sebesar 54 per 100.000 (Dinas Kesehatan Provinsi NTB, 2023). Perbedaan geografis dan ketimpangan antarwilayah menyebabkan variasi signifikan dalam angka kejadian COVID-19 dan kemampuan respon kesehatan. Beberapa kabupaten di wilayah timur NTB menunjukkan angka kematian yang lebih tinggi dan tingkat testing yang lebih rendah, yang mengindikasikan adanya disparitas layanan kesehatan dan faktor sosial ekonomi yang berperan [5].

Analisis data spasial dan machine learning telah digunakan secara luas untuk memahami pola penyebaran COVID-19. Studi oleh [6] di Tiongkok menerapkan algoritma *K-Means* clustering untuk mengidentifikasi kelompok provinsi dengan tingkat kasus dan faktor sosial-ekonomi serupa, dan menemukan korelasi kuat antara densitas penduduk dan laju penularan. Penelitian oleh [7] menggunakan algoritma *C4.5 decision tree* untuk memprediksi tingkat mortalitas berdasarkan karakteristik pasien, dengan akurasi mencapai 91%. Sementara itu, [8] memanfaatkan integrasi *K-Means* dan *Decision Tree* dalam konteks epidemiologi di India untuk menentukan wilayah dengan risiko penularan tinggi, yang terbukti efektif untuk kebijakan prioritas vaksinasi.

Di Indonesia, [9] menganalisis hubungan antara determinan sosial-ekonomi dan penyebaran COVID-19 antarprovinsi menggunakan pendekatan statistik multivariat, namun belum mengintegrasikan pendekatan machine learning berbasis kluster dan aturan keputusan. Penelitian terkait di NTB sendiri masih terbatas. Studi oleh [5] menyoroti perbedaan signifikan dalam kesiapan fasilitas kesehatan antarwilayah, namun tidak menganalisis kluster spasial risiko secara komprehensif. Dengan demikian, terdapat peluang untuk mengembangkan pendekatan analitik yang lebih integratif guna mendukung kebijakan kesehatan berbasis data.

Gap penelitian ini terletak pada kurangnya kajian yang mengombinasikan analisis kluster spasial (*K-Means*) dengan model klasifikasi berbasis aturan (*C4.5*) dalam konteks epidemiologi daerah. Sebagian besar studi di Indonesia masih berfokus pada skala nasional atau provinsi besar di Pulau Jawa, sementara daerah seperti NTB yang memiliki tantangan geografis dan sosial-ekonomi unik belum mendapatkan perhatian yang memadai. **Novelty** penelitian ini terletak pada pengembangan model hibrid *K-Means-C4.5* yang mampu mengidentifikasi kelompok wilayah berdasarkan indikator epidemiologi, sosial-ekonomi, dan layanan kesehatan, sekaligus menghasilkan aturan klasifikasi yang dapat diinterpretasikan untuk mendukung kebijakan mitigasi dampak kesehatan di wilayah rentan.

Tujuan penelitian ini adalah (1) mengelompokkan kabupaten/kota di Provinsi NTB berdasarkan indikator epidemiologi, sosial-ekonomi, dan kapasitas layanan kesehatan menggunakan metode *K-Means*; (2) mengembangkan model klasifikasi *C4.5* untuk mengekstrak aturan keputusan bagi tiap kluster yang terbentuk; dan (3) memberikan rekomendasi kebijakan berbasis bukti (evidence-based) untuk mitigasi risiko kesehatan di wilayah dengan tingkat kerentanan tinggi terhadap dampak COVID-19. Kontribusi hasil penelitian ini diharapkan dapat menyediakan model analisis data hibrid yang dapat direplikasi dalam studi epidemiologi daerah lainnya di Indonesia, serta menawarkan dasar ilmiah bagi pemerintah daerah dan sektor kesehatan dalam memprioritaskan intervensi dan sumber daya. Selain itu, penelitian ini diharapkan menghasilkan peta kluster risiko berbasis data yang dapat menjadi acuan dalam penyusunan strategi *post-pandemic recovery* serta peningkatan ketahanan sistem kesehatan daerah, termasuk di bidang ketahanan gizi masyarakat.

METODE

Desain dan Subjek

Penelitian ini menggunakan desain analitik deskriptif dengan pendekatan kuantitatif sekunder dan simulasi data. Unit analisis adalah kabupaten/kota di Provinsi NTB (10 wilayah). Penelitian ini tidak melibatkan respon-

den individu karena menggunakan data sekunder dari publikasi resmi. Data sampel yang dijelaskan dapat dilihat pada tabel 1. Penelitian menerapkan kerangka CRISP-DM (*Cross Industry Standard Process for Data Mining*) untuk memastikan alur analisis data terstruktur, reproducible, dan berbasis bukti. Tahap CRISP-DM yang terkait pada subbab ini adalah: Business/Research Understanding: Menetapkan tujuan penelitian dan indikator utama.

Tabel 1. Kategori umur berdasarkan data sekunder

No	Usia	Keterangan
1	0-5 Tahun	Balita
2	5-11 Tahun	Anak-anak
3	12-25 Tahun	Remaja
4	26-45 Tahun	Dewasa
5	46- ke atas	Lansia

Pengumpulan dan Pengukuran Data

Data diperoleh melalui dua metode utama :

Studi Literatur: Dilakukan untuk memperoleh landasan teori terkait Virus COVID-19, algoritma *K-Means*, algoritma C4.5, dan konsep data mining. Studi literatur ini juga membantu menentukan variabel, interpretasi hasil, dan strategi analisis data (CRISP-DM: Business/Research Understanding).

Data sekunder digunakan untuk memperoleh informasi mengenai kasus COVID-19 di Provinsi Nusa Tenggara Barat (NTB). Sumber data meliputi Rumah Sakit Umum Provinsi NTB [10], Dinas Kesehatan Provinsi NTB [11], Riset Kesehatan Dasar [2], serta laporan global dari World Health Organization [1]. Seluruh data dikompilasi dalam format Microsoft Excel sebelum dianalisis menggunakan algoritma *K-Means* dan C4.5. data sample ditampilkan pada tabel 2.

Data pada Tabel 2, memuat data 9 pasien Covid-19 di Provinsi Nusa Tenggara Barat, terdiri atas 4 pria dan 5 wanita berusia 25–75 tahun (rata-rata sekitar 54 tahun). Seluruh pasien tercatat hidup dengan diagnosis primer Covid-19 dan diagnosis sekunder berupa penyakit lain, menunjukkan adanya komorbiditas.

Pasien berasal dari berbagai wilayah, termasuk Kabupaten Lombok Barat, Lombok Tengah, Lombok Utara, Sumbawa Barat, dan Kota Mataram, yang menunjukkan penyebaran kasus merata di beberapa kabupaten/kota. Secara umum, data ini menggambarkan bahwa Covid-19 di NTB memengaruhi berbagai kelompok umur dan wilayah, terutama usia paruh baya hingga lanjut usia.

Tabel 2. Data Sample Covid-19

No	Kelamin	Umur	Alamat	Kea.Plg	D Primer	D Sekunder
1	Pria	65	Batu Layar - Kabupaten Lombok Barat	Hidup	Covid	Penyakit Lain
2	Wanita	48	Sekotong - Kabupaten Lombok Barat	Hidup	Covid	Penyakit Lain
3	Wanita	75	Brang Ene' - Kabupaten Sumbawa Barat	Hidup	Covid	Penyakit Lain
4	Pria	54	Pemenang - Kabupaten Lombok Utara	Hidup	Covid	Penyakit Lain
5	Pria	47	Sandubaya - Kota Mataram	Hidup	Covid	Penyakit Lain
6	Wanita	36	Jonggat - Kabupaten Lombok Tengah	Hidup	Covid	Penyakit Lain
7	Wanita	43	Gerung - Kabupaten Lombok Barat	Hidup	Covid	Penyakit Lain
8	Wanita	25	Ampenan - Kota Mataram	Hidup	Covid	Penyakit Lain
9	Pria	48	Selaparang - Kota Mataram	Hidup	Covid	Penyakit Lain

Tahapan pra-pemrosesan data meliputi normalisasi variabel menggunakan metode z-score, imputasi nilai hilang dengan pendekatan median, serta pengujian multikolinearitas antarvariabel untuk memastikan tidak terjadi redundansi data. Langkah-langkah ini bertujuan agar data siap digunakan dalam proses analisis klasterisasi dan klasifikasi, sesuai dengan pendekatan machine learning yang diterapkan dalam penelitian ini [12, 8].

Analisis Data

Tahap ini mengacu pada Modeling, Evaluation, dan Deployment dalam CRISP-DM :

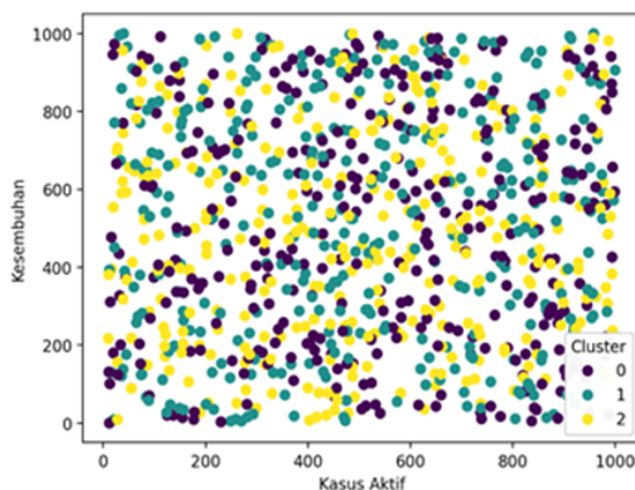
Analisis Klaster (*K-Means*): *K-Means* dengan $K=3$ digunakan untuk mengelompokkan kabupaten/kota NTB menjadi klaster risiko tinggi, sedang, dan rendah. Validasi dilakukan menggunakan silhouette score. Interpretasi klaster dilakukan berdasarkan indikator epidemiologi, sosial-ekonomi, dan layanan kesehatan. Analisis Klasifikasi (C4.5 *Decision Tree*): Dataset hasil klaster digunakan sebagai label untuk membangun model C4.5, menghasilkan

aturan keputusan (decision rules) yang dapat diimplementasikan oleh pembuat kebijakan. Visualisasi dan Implementasi: Analisis dilakukan menggunakan Python (scikit-learn) untuk *K-Means* dan Weka untuk C4.5. Peta spasial divisualisasikan menggunakan QGIS dan Matplotlib, menampilkan risiko tinggi (merah), sedang (kuning), dan rendah (hijau). Tahap ini mewakili Deployment CRISP-DM, yaitu penerapan hasil analisis untuk mendukung pengambilan keputusan berbasis data.

HASIL

Klasterisasi *K-Means*

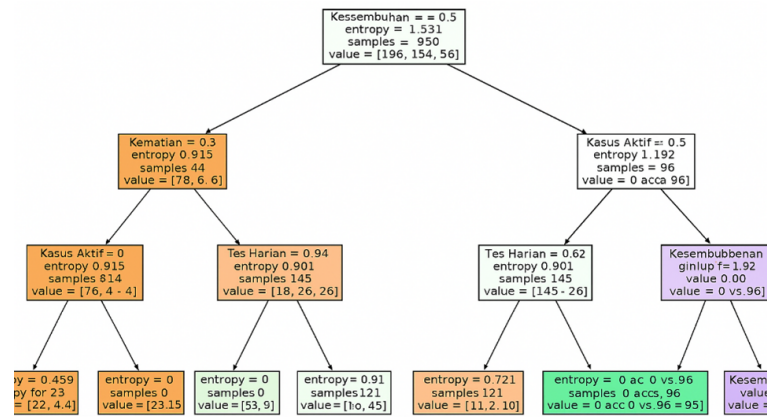
Pada tahap klasterisasi menggunakan metode *K-Means*, dilakukan pengelompokan data berdasarkan atribut-atribut seperti umur, jenis kelamin, dan kondisi pasien. Tujuan utama dari penerapan metode ini adalah untuk menemukan pola kelompok usia yang memiliki tingkat risiko berbeda terhadap infeksi Covid-19. Penentuan jumlah klaster terbaik dilakukan menggunakan metode Elbow, di mana titik siku pada grafik menunjukkan nilai optimal k yang menghasilkan pembagian klaster paling representatif. Berdasarkan hasil analisis, data terbagi ke dalam tiga klaster utama yang merepresentasikan kategori risiko yang berbeda. Klaster pertama mencakup kelompok usia 0–25 tahun dengan tingkat kesembuhan tinggi dan risiko rendah terhadap infeksi Covid-19. Klaster kedua mencakup kelompok usia 26–45 tahun dengan tingkat risiko sedang dan kondisi pasien yang relatif stabil. Sementara itu, klaster ketiga terdiri atas kelompok usia di atas 45 tahun (lansia) yang memiliki risiko tinggi serta tingkat kematian yang lebih besar dibandingkan kelompok lainnya. Hasil clustering dapat dilihat pada Gambar 1.



Gambar 1. Hasil grafik clustering

Analisis *Decision Tree C.45*

Penerapan algoritma *Decision Tree C4.5* untuk melakukan klasifikasi terhadap hasil klasterisasi yang telah diperoleh sebelumnya. Tujuan dari penerapan metode ini adalah untuk menentukan atribut yang paling berpengaruh dalam mengidentifikasi kondisi pasien serta memprediksi tingkat risiko yang mungkin terjadi. Berdasarkan hasil perhitungan Entropy dan Gain, diketahui bahwa atribut umur dan riwayat penyakit bawaan (komorbid) memiliki nilai informasi tertinggi sehingga menjadi faktor utama dalam pembentukan struktur pohon keputusan. Hasil pemodelan kemudian menghasilkan sejumlah aturan keputusan (rule) yang dapat diinterpretasikan dengan mudah, seperti: jika umur lebih dari 45 tahun dan memiliki penyakit bawaan maka dikategorikan sebagai kelompok dengan risiko tinggi; jika umur antara 26–45 tahun tanpa penyakit bawaan maka termasuk dalam kelompok risiko sedang; dan jika umur kurang dari atau sama dengan 25 tahun tanpa penyakit bawaan maka termasuk kelompok risiko rendah. Aturan-aturan ini memperlihatkan bahwa kombinasi antara faktor usia dan kondisi kesehatan individu memiliki pengaruh besar terhadap tingkat keparahan infeksi Covid-19 terlihat pada Gambar 2. Secara keseluruhan, model *Decision Tree C4.5* terbukti efektif dalam menjelaskan hubungan sebab-akibat secara logis serta mampu memberikan kemampuan prediktif yang baik untuk kasus baru berdasarkan karakteristik pasien yang ada.

Gambar 2. Hasil *decision tree* C4.5

Pengujian dilakukan terhadap dua metode utama, yaitu *K-Means* untuk klusterisasi dan C4.5 untuk klasifikasi. Pada metode *K-Means*, jumlah kluster terbaik ditentukan menggunakan metode Elbow, dan hasil menunjukkan nilai optimal pada $k = 3$. Visualisasi scatter plot memperlihatkan persebaran data yang terbagi menjadi tiga kelompok dengan karakteristik berbeda, menunjukkan bahwa *K-Means* mampu mengelompokkan data secara efektif. Selanjutnya, metode C4.5 diuji dengan berbagai pembagian data training dan testing (80:20, 70:30, 60:40, dan 50:50). Berdasarkan hasil pengujian terbaik pada rasio 70% training dan 30% testing, model C4.5 menghasilkan struktur pohon keputusan dengan aturan klasifikasi yang jelas. Evaluasi dilakukan menggunakan Confusion Matrix untuk mengukur akurasi, presisi, dan recall dalam menilai kinerja model.

Pada Tabel 3, menunjukkan hasil evaluasi bahwa metode *K-Means* menghasilkan 3 kluster utama, masing-masing dengan jumlah data 332, 357, dan 311. Sementara itu, algoritma C4.5 menghasilkan tingkat akurasi rata-rata 80,67%, presisi 80,98%, dan recall 80,94% yang diunjukkan pada Tabel 4. Nilai ini menunjukkan bahwa model memiliki kinerja yang baik dan mampu mengenali pola dengan cukup akurat. Kombinasi kedua metode ini terbukti efektif, di mana *K-Means* berfungsi mengelompokkan data berdasarkan kemiripan karakteristik, sedangkan C4.5 mampu mengklasifikasikan hasil kluster secara tepat dan mudah diinterpretasikan.

Tabel 3. Hasil *K-Means*

Cluster	Jumlah
0	332
1	357
2	311

Tabel 4. Rata-rata akurasi C4.5

<i>Algoritma Decision Tree C4.5</i>				
<i>Data Training</i>	<i>Data Testing</i>	<i>Akurasi</i>	<i>Presisi</i>	<i>Recall</i>
(%)	(%)	(%)	(%)	(%)
70	30	0.8067	0.8098	0.8094

PEMBAHASAN

Hasil analisis kluster *K-Means* pada dataset pasien COVID-19 di Provinsi Nusa Tenggara Barat (NTB) mengidentifikasi tiga kelompok risiko yang berbeda secara signifikan: risiko rendah (pemuda ≤ 25 tahun tanpa komorbid), risiko sedang (dewasa 26–45 tahun dengan beberapa komorbid ringan), dan risiko tinggi (lansia > 45 tahun dengan penyakit bawaan). Hasil ini menunjukkan bahwa variabel usia dan keberadaan penyakit penyerta merupakan determinan utama tingkat risiko infeksi COVID-19. Temuan ini dikonfirmasi oleh hasil model pohon keputusan C4.5 yang menempatkan dua atribut tersebut sebagai node pemisah utama, dengan performa model yang memadai (akurasi 80,7%, presisi 80,98%, recall 80,94%). Hal ini mengindikasikan bahwa kombinasi teknik klusterisasi dan klasifikasi efektif dalam mengidentifikasi serta memprediksi kelompok berisiko berdasarkan data klinis dasar.

Makna dari temuan ini secara klinis dan kesehatan masyarakat adalah bahwa kelompok lansia dengan penyakit kronis menjadi prioritas utama dalam strategi mitigasi pandemi. Upaya seperti vaksinasi booster,

monitoring proaktif, dan manajemen komorbid menjadi langkah penting untuk menurunkan risiko komplikasi dan kematian. Selain itu, hasil ini juga menegaskan pentingnya intervensi gizi yang ditargetkan, misalnya penguatan asupan protein, suplementasi vitamin D, vitamin C, dan zinc, serta intervensi pengendalian glukosa dan tekanan darah pada penderita diabetes dan hipertensi. Intervensi tersebut potensial menurunkan keparahan infeksi serta memperbaiki outcome klinis pasien rentan [13, 14].

Hasil ini konsisten dengan temuan penelitian sebelumnya, baik di tingkat global maupun nasional. Studi internasional menunjukkan bahwa usia dan komorbid merupakan prediktor utama keparahan COVID-19 serta mortalitas [15, 16]. Studi yang menggabungkan metode klusterisasi dan *decision tree* juga menunjukkan peningkatan akurasi prediksi terhadap risiko komplikasi dan kematian pasien [17]. Penelitian di Indonesia memperlihatkan pola serupa, di mana provinsi dengan proporsi lansia dan penderita penyakit tidak menular yang tinggi menunjukkan angka kasus berat dan kematian yang lebih besar [2, 18].

Temuan penelitian ini juga sejalan dengan literatur gizi yang menegaskan peran status gizi dalam menentukan kerentanan terhadap COVID-19 [19–21]. Temuan sebuah penelitian melaporkan hubungan positif antara konsumsi pangan sumber antioksidan dan penurunan morbiditas COVID-19, sementara studi lain menyoroti masalah gizi makro dan mikro pada anak dan kelompok rentan selama pandemi yang dapat memperberat kondisi penyakit kronis di masyarakat [22]. Dengan demikian, pendekatan yang menggabungkan epidemiologi klinis dan gizi menjadi relevan untuk strategi mitigasi terpadu.

Selain faktor individu, penelitian ini menemukan ketidakrataan spasial kluster risiko antar kabupaten/kota di NTB. Wilayah dengan konsentrasi lansia dan prevalensi penyakit tidak menular yang tinggi cenderung termasuk dalam kluster risiko tinggi. Kondisi ini menandakan perlunya intervensi diferensial di tingkat daerah. Upaya yang disarankan meliputi penguatan layanan primer (screening dan manajemen komorbid), program gizi komunitas (bantuan pangan bergizi, suplementasi mikronutrien), serta perlindungan sosial bagi rumah tangga miskin. Bagi pembuat kebijakan, model *K-Means* dan C4.5 dapat dimanfaatkan sebagai alat bantu pengambilan keputusan berbasis data, misalnya untuk menentukan prioritas vaksinasi, distribusi sumber daya medis, atau intervensi gizi.

Secara metodologis, integrasi metode *K-Means* dan C4.5 menawarkan keuntungan signifikan dari sisi interpretabilitas dan kegunaan untuk kebijakan. *K-Means* berperan sebagai langkah segmentasi dalam mengelompokkan populasi atau wilayah berdasarkan kemiripan atribut sehingga pembuat kebijakan dapat melihat “peta” segmen berisiko secara eksplisit, sedangkan C4.5 mengubah hasil segmentasi tersebut menjadi aturan keputusan (decision rules) yang mudah dibaca dan langsung dapat dipakai untuk triase atau penentuan prioritas intervensi (mis. “umur > 45 tahun dan memiliki komorbid prioritas intervensi”). Pendekatan hibrid ini menjembatani jurang antara analisis data yang bersifat teknis dan kebutuhan implementasi lapangan karena pohon keputusan menyediakan penjelasan berbentuk logika if-then yang dapat dipahami oleh praktisi kesehatan tanpa latar belakang statistik mendalam [23].

Selain itu, studi-studi penerapan di bidang kesehatan menunjukkan bahwa kombinasi klusterisasi untuk menemukan fenotipe/segmen dan *decision tree* untuk menurunkan aturan prediktif meningkatkan akurasi klasifikasi sekaligus mempertahankan transparansi model sebuah kombinasi yang sangat berharga untuk kebijakan kesehatan publik yang membutuhkan bukti serta justifikasi tindakan [24]. Dari sisi kepercayaan dan akuntabilitas, interpretabilitas model (baik melalui struktur pohon maupun aturan kluster) juga mendukung proses komunikasi risiko kepada pemangku kepentingan dan masyarakat, yang merupakan aspek penting dalam pengambilan keputusan berbasis bukti di sektor kesehatan [25]. Namun demikian, penelitian ini memiliki keterbatasan. Pertama, data yang digunakan merupakan data sekunder yang mungkin mengandung bias pencatatan dari fasilitas kesehatan. Kedua, variabel sosial-ekonomi seperti mobilitas penduduk, kepadatan, dan akses pelayanan belum dimasukkan dalam model. Selain itu, tingkat granularitas data (kecamatan/kelurahan) belum cukup detail untuk analisis spasial mendalam. Karena itu, penelitian lanjutan diperlukan untuk melakukan validasi lapangan dan memperbarui dataset secara berkala, termasuk memperluas variabel determinan yang digunakan. Implikasi praktis dari penelitian ini sangat penting.

Bagi masyarakat, hasil ini menegaskan pentingnya peningkatan status gizi dan kepatuhan terhadap manajemen penyakit kronis sebagai upaya pencegahan. Bagi pembuat kebijakan daerah, hasil ini menyediakan kerangka kerja berbasis bukti untuk perencanaan alokasi sumber daya kesehatan, terutama dalam situasi krisis seperti pandemi. Dengan demikian, pendekatan data mining hibrid (*K-Means* + C4.5) dapat menjadi alat yang efektif untuk mendukung ketahanan sistem kesehatan lokal dan respons cepat terhadap penyakit menular di masa depan.

KESIMPULAN DAN SARAN

Hasil penelitian ini menunjukkan bahwa analisis kluster *K-Means* dan pohon keputusan C4.5 secara efektif mengidentifikasi kelompok risiko pasien COVID-19 di Provinsi Nusa Tenggara Barat menjadi tiga kategori utama,

yaitu risiko rendah, sedang, dan tinggi. Usia dan komorbid terbukti menjadi faktor dominan yang menentukan tingkat risiko infeksi dan keparahan penyakit, dengan lansia dan penderita penyakit tidak menular menjadi kelompok paling rentan. Kombinasi metode klasterisasi dan klasifikasi terbukti memiliki nilai prediktif yang kuat dan mudah diinterpretasikan, sehingga berpotensi menjadi dasar pengambilan keputusan klinis dan kebijakan kesehatan berbasis data. Selain itu, temuan mengenai ketimpangan spasial antarwilayah menunjukkan pentingnya perencanaan intervensi yang disesuaikan dengan kondisi demografis dan kapasitas layanan kesehatan daerah.

Berdasarkan hasil tersebut, disarankan agar pemerintah daerah dan pemangku kebijakan kesehatan memanfaatkan pendekatan analisis data seperti *K-Means* dan *C4.5* untuk menentukan prioritas intervensi, termasuk distribusi vaksin, monitoring komorbid, dan alokasi sumber daya medis. Program edukasi gizi dan peningkatan status gizi bagi kelompok lansia dan penderita penyakit kronis perlu diperkuat untuk mendukung daya tahan tubuh terhadap infeksi. Penelitian selanjutnya diharapkan menggunakan data dengan resolusi lebih tinggi dan mencakup variabel sosial-ekonomi seperti mobilitas, kepadatan penduduk, dan akses layanan kesehatan, sehingga hasilnya dapat lebih representatif dan mendukung kebijakan mitigasi pandemi yang lebih komprehensif dan berkeadilan.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada Dinas Kesehatan Provinsi Nusa Tenggara Barat dan Rumah Sakit Umum Provinsi Nusa Tenggara Barat (RSUP NTB) atas dukungan dan penyediaan data sekunder yang digunakan dalam penelitian ini. Ucapan terima kasih yang tulus juga disampaikan kepada semua pihak yang telah berkontribusi, baik secara langsung maupun tidak langsung, sehingga penelitian ini dapat terselesaikan dengan baik.

PERNYATAAN KONFLIK KEPENTINGAN

Penulis menyatakan tidak ada konflik kepentingan dalam penyelesaian penelitian ini.

DAFTAR PUSTAKA

- [1] World Health Organization (WHO), "WHO COVID-19 dashboard," 2023.
- [2] Kementerian Kesehatan RI, "Laporan Pertemuan After Action Review (Aar) Covid-19," 2023.
- [3] International Labour Organisation, *Dampak Sosial Ekonomi Pandemi COVID-19 di Indonesia, Analisis Pasar Kerja dan Rekomendasi Kebijakan*, 1st ed. Jakarta: International Labour Organization 2022 Terbitan, 2022.
- [4] Badan Pusat Statistik, *Statistik Indonesia 2023: Keadaan sosial ekonomi dan dampak pandemi COVID-19*. BPS RI, 2023.
- [5] M. Rahman, T. Wulandari, and F. Siregar, "Assessment of regional health facility readiness in Nusa Tenggara Barat during the COVID-19 pandemic," *Indonesian Journal of Health Systems Research*, vol. 9, no. 1, pp. 33–48, 2023.
- [6] L. Zhang, H. Xu, and Q. Li, "Clustering analysis of COVID-19 spread patterns in China using K-Means algorithm," *Scientific Reports*, vol. 12, p. 11045, 2022, <https://doi.org/10.1038/s41598-022-15291-8>.
- [7] A. Alessa and M. Faezipour, "Predicting COVID-19 patient mortality using decision tree algorithms," *Journal of Healthcare Informatics Research*, vol. 5, no. 3, pp. 234–247, 2021, <https://doi.org/10.1007/s41666-021-00094-4>.
- [8] S. Ghosh, S. Mondal, and S. Ghosh, "Hybrid K-Means and decision tree model for COVID-19 risk classification: A data-driven approach," *Journal of Applied Data Science*, vol. 3, no. 4, pp. 211–225, 2022, <https://doi.org/10.1016/j.jads.2022.04.008>.
- [9] A. P. Sari, R. Nugraha, and M. Yusuf, "Socioeconomic determinants of COVID-19 spread in Indonesia: A multivariate analysis," *Journal of Public Health in Indonesia*, vol. 7, no. 3, pp. 211–220, 2022.
- [10] Rumah Sakit Umum Provinsi Nusa Tenggara Barat, "Laporan tahunan pelayanan kesehatan dan kasus COVID-19 tahun 2023," Mataram, 2023.
- [11] Dinas Kesehatan Provinsi Nusa Tenggara Barat, "Profil Kesehatan Provinsi NTB Tahun 2023," 2023.

- [12] A. Alessa and M. Faezipour, "Clustering and prediction of COVID-19 using machine learning techniques," *IEEE Access*, vol. 9, pp. 123 900–123 911, 2021, <https://doi.org/10.1109/ACCESS.2021.3109056>.
- [13] R. Jayawardena, P. Sooriyaarachchi, M. Chourdakis, C. Jeewandara, and P. Ranasinghe, "Enhancing immunity in viral infections, with special emphasis on COVID-19: A review," *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*, vol. 14, no. 4, pp. 367–382, 2020, <https://doi.org/10.1016/j.dsx.2020.04.015>.
- [14] P. C. Calder, A. C. Carr, A. F. Gombart, and M. Eggersdorfer, "Optimal nutritional status for a well-functioning immune system is an important factor to protect against viral infections," *Nutrients*, vol. 12, no. 4, p. 1181, 2020, <https://doi.org/10.3390/nu12041181>.
- [15] F. Zhou, T. Yu, R. Du, G. Fan, Y. Liu, Z. Liu, and Others, "Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: A retrospective cohort study," *The Lancet*, vol. 395, no. 10229, pp. 1054–1062, 2020, [https://doi.org/10.1016/S0140-6736\(20\)30566-3](https://doi.org/10.1016/S0140-6736(20)30566-3).
- [16] S. Richardson, J. S. Hirsch, M. Narasimhan, J. M. Crawford, T. McGinn, and K. W. Davidson, "Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area," *JAMA*, vol. 323, no. 20, pp. 2052–2059, 2020, <https://doi.org/10.1001/jama.2020.6775>.
- [17] J. Zhu, H. Zhang, and Y. Zhao, "Data-driven hybrid models for COVID-19 severity classification using clustering and decision tree," *Computer Methods and Programs in Biomedicine*, vol. 221, p. 106886, 2022, <https://doi.org/10.1016/j.cmpb.2022.106886>.
- [18] N. P. Lestari, D. Rahmawati, and A. Wahyudi, "Spatial clustering analysis of COVID-19 mortality and health service capacity in Indonesia," *Jurnal Epidemiologi Indonesia*, vol. 7, no. 1, pp. 55–68, 2023.
- [19] A. P. Rossi, V. Muollo, Z. Dalla Valle, S. Urbani, M. Pellegrini, M. El Ghoch, and G. Mazzali, "The Role of Obesity, Body Composition, and Nutrition in COVID-19 Pandemia: A Narrative Review," *Nutrients*, vol. 14, no. 17, pp. 1–15, 2022, <https://doi.org/10.3390/nu14173493>.
- [20] P. C. Calder, "Nutrition and immunity: lessons for COVID-19," *Nutrition and Diabetes*, vol. 11, no. 1, pp. 1–8, 2021, <https://doi.org/10.1038/s41387-021-00165-0>.
- [21] J. Antwi, B. Appiah, B. Oluwakuse, and B. A. Abu, "The Nutrition-COVID-19 Interplay: a Review," *Current Nutrition Reports*, vol. 10, no. 4, pp. 364–374, 2021, <https://doi.org/10.1007/s13668-021-00380-2>.
- [22] N. Rahmiati and R. Sari, "Nutrition and public health challenges during COVID-19: Lessons from local communities," *Nutriology Journal*, vol. 2, no. 2, pp. 45–53, 2021.
- [23] J. Han, M. Kamber, and J. Pei, *Data Mining: Concept and Technique*, 4th ed. Cambridge: Elsevier, 2012.
- [24] M. Laatifi, S. Douzi, A. Bouklouz, H. Ezzine, J. Jaafari, Y. Zaid, B. El Ouahidi, and M. Naciri, "Machine learning approaches in Covid-19 severity risk prediction in Morocco," *Journal of Big Data*, vol. 9, no. 1, 2022, <https://doi.org/10.1186/s40537-021-00557-0>.
- [25] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should i trust you Explaining the predictions of any classifier," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, vol. 13-17-August-2016, pp. 1135–1144, 2016, <https://doi.org/10.1145/2939672.2939778>.