JITET (Jurnal Informatika dan Teknik Elektro Terapan)
Vol. 13 No. 2, pISSN: 2303-0577 eISSN: 2830-7062 http://dx.doi.org/10.23960/jitet.v13i2.6276

PEMANFAATAN K-MEANS UNTUK PEMETAAN EPIDEMIOLOGI KASUS KUSTA DI KABUPATEN CIREBON

Egi Susanto Dimin[1*], Nining Rahaningsih[2], Raditya Danar Dana[3], Cep Lukman Rohmat[4]

[1,2,3,4]STMIK IKMI Cirebon; Jl. Perjuangan No. 10 B Majasem Kec. Kesambi Kota Cirebon; telp.(0231)490480

Received: 24 Februari 2025
Accepted: 27 Maret 2025
Published: 14 April 2025

Keywords:
Kusta;
Klasterisasi;
Algoritma K-Means;
Kabupaten Cirebon.

Corespondent Email:
egisusanto382@gmail.com

Abstrak. Kusta merupakan masalah kesehatan masyarakat di Kabupaten Cirebon, dengan 240 kasus aktif pada 2022–2023. Penyebaran penyakit ini menunjukkan perlunya strategi pengendalian berbasis data. Penelitian ini menerapkan algoritma K-Means untuk mengelompokkan wilayah berdasarkan tingkat kerawanan kasus kusta karena kemampuannya mengidentifikasi pola distribusi epidemiologi. Metode yang digunakan mencakup pengumpulan data dari portal data terbuka Kabupaten Cirebon, preprocessing untuk memastikan kualitas data, penerapan K-Means, serta evaluasi hasil klasterisasi menggunakan Davies-Bouldin Index. Dataset terdiri dari jumlah kasus, lokasi geografis, jenis kusta, dan faktor demografi. Hasil analisis menunjukkan bahwa wilayah Kabupaten Cirebon terbagi dalam tiga klaster kerawanan: tinggi, sedang, dan rendah. Wilayah dengan tingkat kerawanan tinggi memiliki kepadatan kasus lebih besar serta keterbatasan akses layanan kesehatan. Temuan ini dapat menjadi dasar bagi pemerintah dalam menyusun kebijakan intervensi yang lebih terarah, seperti distribusi tenaga medis dan edukasi masyarakat. Studi ini menekankan pentingnya pemanfaatan teknologi berbasis data untuk meningkatkan efektivitas program pengendalian kusta di masa depan.

Abstract. *Leprosy is a public health issue in Cirebon Regency, with 240 active cases in 2022–2023. The spread of this disease indicates the need for data-driven control strategies. This research applies the K-Means algorithm to cluster regions based on the level of susceptibility to leprosy cases due to its ability to identify epidemiological distribution patterns. The methods used include data collection from the Cirebon Regency open data portal, preprocessing to ensure data quality, application of K-Means, and evaluation of the clustering results using the Davies-Bouldin Index. The dataset consists of the number of cases, geographic location, type of leprosy, and demographic factors. The analysis results show that the Cirebon Regency area is divided into three vulnerability clusters: high, medium, and low. Regions with a high level of vulnerability have a higher case density and limited access to healthcare services. These findings can serve as a basis for the government in formulating more targeted intervention policies, such as the distribution of medical personnel and community education. This study emphasizes the importance of utilizing data-based technology to enhance the effectiveness of leprosy control programs in the future.*

1. PENDAHULUAN

Kusta adalah penyakit menular yang dapat menyebabkan disabilitas permanen jika tidak ditangani dengan baik dan tepat. Kabupaten Cirebon merupakan salah satu wilayah di Indonesia dengan jumlah kasus aktif yang cukup tinggi, mencapai 240 kasus pada periode 2022–2023. Penyebaran kasus yang tidak merata menunjukkan perlunya strategi

berbasis data untuk memetakan wilayah rawan. Salah satu metode yang efektif untuk analisis ini adalah algoritma K-Means, yang mampu mengelompokkan data ke dalam klaster berdasarkan karakteristik tertentu. Penelitian yang dilakukan oleh [1] menunjukkan bahwa algoritma K-Means dapat membantu mengelompokkan distribusi sosial ekonomi masyarakat, sehingga pola distribusi lebih mudah dipahami. [2] juga menyoroti penggunaan algoritma ini untuk memetakan wilayah prioritas vaksinasi COVID-19, memberikan wawasan bagi pemerintah dalam menentukan prioritas intervensi. Dalam studi [3] algoritma K-Means dibandingkan dengan K-Medoids untuk analisis kemiskinan di Indonesia, menghasilkan wawasan yang relevan tentang wilayah dengan tingkat kemiskinan tinggi. Pada sektor pendidikan, penelitian [4] mengidentifikasi disparitas fasilitas pendidikan antar wilayah menggunakan algoritma K-Means. Sementara itu, [5] mengaplikasikan algoritma ini dalam manajemen stok obat di fasilitas kesehatan, memberikan solusi yang lebih efisien dalam pengelolaan sumber daya. Oleh karena itu, penelitian ini bertujuan menerapkan algoritma K-Means untuk mengidentifikasi wilayah di Kabupaten Cirebon berdasarkan tingkat kerawanan kasus kusta. Dengan pendekatan ini, strategi intervensi yang lebih terarah dan efisien dapat dirancang.

2. TINJAUAN PUSTAKA

Algoritma K-Means merupakan salah satu metode *unsupervised learning* yang banyak digunakan dalam proses pengelompokan data berdasarkan karakteristik yang serupa. Metode ini bekerja dengan menetapkan sejumlah pusat klaster (*centroid*), lalu mengelompokkan data berdasarkan kedekatan nilai terhadap titik pusat tersebut [6].

Dalam beberapa penelitian sebelumnya, metode klasterisasi telah digunakan untuk mengelompokkan wilayah berdasarkan tingkat risiko penyebaran penyakit. Misalnya, studi yang dilakukan oleh [7] menunjukkan bahwa metode ini dapat diterapkan dalam pemetaan cakupan vaksinasi COVID-19 di suatu daerah. Dengan menerapkan pendekatan serupa, metode K-Means dapat dimanfaatkan untuk mengidentifikasi wilayah rawan penyebaran kusta berdasarkan jumlah kasus, kepadatan penduduk, serta faktor lingkungan yang berpengaruh terhadap transmisi penyakit ini.

Dalam penelitian yang dilakukan oleh [8] algoritma K-Means diterapkan untuk menentukan penerima Bantuan Langsung Tunai (BLT) berdasarkan kriteria tertentu, sehingga proses seleksi menjadi lebih sistematis dan tepat sasaran.

Sementara itu, penelitian [9] menggunakan DBI untuk mengevaluasi pengelompokan usaha kecil dan menengah (UKM) berdasarkan kinerja bisnis mereka, yang membantu dalam menentukan strategi pengembangan usaha.

Selain DBI, Silhouette Score juga sering digunakan untuk menilai keakuratan klasterisasi. Studi yang dilakukan oleh [10] mengenai klasterisasi data kemiskinan di Provinsi Jawa Barat menunjukkan bahwa Silhouette Score dapat memberikan gambaran yang jelas tentang sejauh mana data berada dalam klaster yang tepat dan efektif.

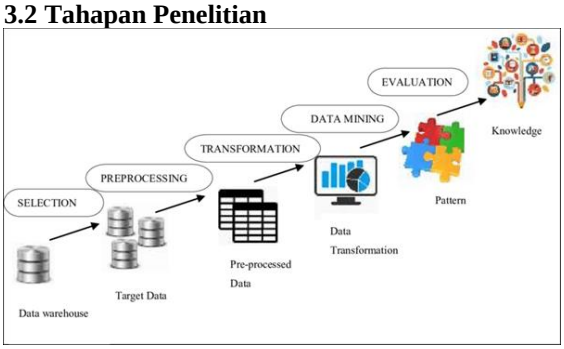
3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode clustering algoritma K-Means untuk mengidentifikasi dan mengelompokkan wilayah di Kabupaten Cirebon berdasarkan tingkat kerawanan kasus kusta. Metode ini bertujuan untuk memberikan pengelompokan yang jelas terkait penyebaran kasus kusta sehingga dapat membantu pemerintah dalam menyusun strategi penanganan yang lebih tepat sasaran. Algoritma K-Means dipilih karena memiliki kemampuan yang baik dalam mengelompokkan data ke dalam beberapa cluster yang berbasis pada kesamaan karakteristik, khususnya pada data epidemiologi.[11]

3.1 Metode pengumpulan

Penelitian ini menggunakan data sekunder yang diperoleh dari portal resmi Kabupaten Cirebon (https://opendata.cirebonkab.go.id), mencakup jumlah kasus kusta, distribusi wilayah, dan karakteristik terkait. Data diidentifikasi, diunduh, lalu melalui tahap preprocessing untuk memastikan kualitasnya, seperti pembersihan anomali, penyesuaian format, dan normalisasi, sebelum dianalisis menggunakan algoritma K-Means.

3.2 Tahapan Penelitian



Gambar 1. Tahapan Metode Penelitian

Dalam penelitian ini, analisis data dilakukan dengan pendekatan *Knowledge Discovery in Databases* (KDD) yang terdiri dari beberapa tahapan: pemilihan data, praproses data, transformasi data, *data mining*, hingga interpretasi dan evaluasi hasil. Tahapan KDD ini bertujuan

untuk mendapatkan pengetahuan baru dari data epidemiologi kasus kusta di Kabupaten Cirebon melalui metode pengelompokan (clustering) berbasis algoritma K-Means.

4. HASIL DAN PEMBAHASAN

Pada tahap ini, akan memaparkan hasil analisis clustering menggunakan algoritma K-Means terhadap data kasus kusta di Kabupaten Cirebon. Dalam proses pengolahan ini menggunakan *tools* RapidMiner Studio versi AI Studio 2024.1.0.

4.1 Data

Dataset yang digunakan adalah data epidemiologi kasus kusta di Kabupaten Cirebon. Yang diambil dari platform data terbuka resmi yang dikelola oleh Pemerintah Kabupaten Cirebon, (https://opendata.cirebonkab.go.id). Dengan jumlah data sebanyak 240 *record.*

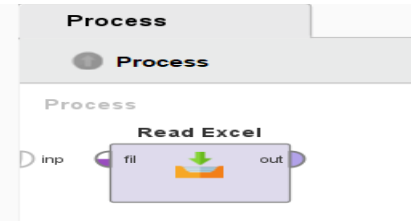
Tabel 1. Dataset

id	kode_pro	nama_pro	bps_kode	bps_nama	bps_kode	bps_nama	kemendag	kemendag	nama_pus	usia	jenis_kust	jumlah_p	satuan	tahun
1	32	JAWA BAR	3209	KABUPATE	3209010	WALED	320901	WALED	WALED	ANAK	PAUSI BAS	0	ORANG	2023
2	32	JAWA BAR	3209	KABUPATE	3209010	WALED	320901	WALED	CIBOGO	ANAK	PAUSI BAS	0	ORANG	2023
3	32	JAWA BAR	3209	KABUPATE	3209011	PASALEMA	320932	PASALEMA	PASALEMA	ANAK	PAUSI BAS	0	ORANG	2023
4	32	JAWA BAR	3209	KABUPATE	3209020	CILEDUG	320902	CILEDUG	CILEDUG	ANAK	PAUSI BAS	0	ORANG	2023
5	32	JAWA BAR	3209	KABUPATE	3209021	PABUARAN	320933	PABUARAN	PABUARAN	ANAK	PAUSI BAS	0	ORANG	2023
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
240	32	JAWA BAR	3209	KABUPATE	3209231	KALIWEDI	320929	KALIWEDI	KALIWEDI	DEWASA	MULTI BAC	3	ORANG	2023

4.2 Data Selection

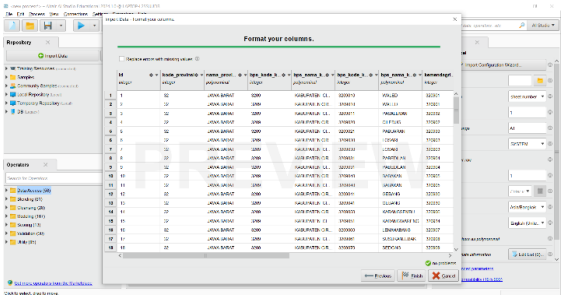
Tahapan pemilihan data melibatkan identifikasi dan seleksi data yang relevan untuk memenuhi tujuan penelitian, yaitu pengelompokan wilayah di Kabupaten Cirebon berdasarkan tingkat kerawanan kasus kusta. Data diperoleh dari portal *open data* Kabupaten Cirebon yang mencakup atribut-atribut penting.

Selanjutnya, *import* data untuk memasukan data yang akan digunakan dan diuji dalam bentuk format .xls atau .xlsx. Berikut adalah cara untuk melakukan *import file Microsoft Excel.*



Gambar 2. Operator Read Excel

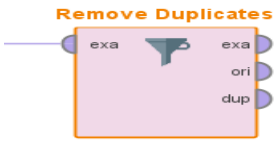
Kemudian, setelah dilakukannya proses “Read Excel” menggunakan RapidMiner, hasil dataset yang sudah diimpor ditampilkan dalam format columns atau bentuk tabel yang dapat dilihat pada Gambar 3.



Gambar 3. Hasil Import Data

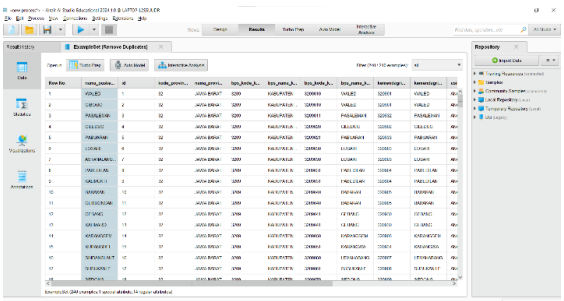
4.3 Preprocessing Data

Preprocessing data adalah tahap persiapan dan pembersihan data sebelum data tersebut digunakan dalam proses analisis atau pembelajaran mesin. Langkah ini bertujuan untuk memastikan data dalam kondisi optimal, bersih, dan siap diolah, sehingga analisis atau model yang dihasilkan lebih akurat dan efektif. Selanjutnya adalah menghilangkan duplikat pada data menggunakan operator *Remove Duplicates.*



Gambar 4. Operator Remove Duplicates

Hasil dari penggunaan operator *Remove Duplicates* tampak pada gambar dibawah ini.



Gambar 5. Hasil Dari Remove Duplicates

Operator *Remove Duplicates* digunakan untuk menghapus baris duplikat dalam dataset. Ini berguna ketika ingin memastikan bahwa data tidak memiliki nilai yang berulang, yang dapat menyebabkan analisis yang tidak akurat. Dari hasil gambar tersebut tidak ditemukan adanya data yang duplikat.

Selanjutnya adalah memilih atribut yang dibutuhkan dengan menggunakan operator *Select Attributes.*

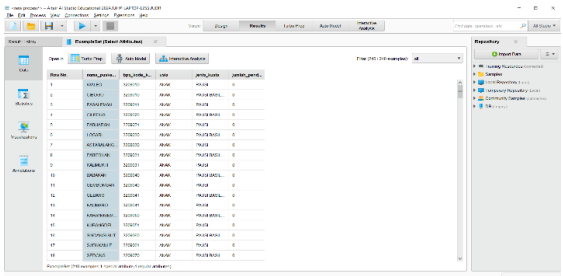
Gambar 6. Operator Select Attributes

Selanjutnya, memilih beberapa atribut yang akan digunakan untuk analisis berikutnya.

Tabel 2. Pemilihan Atribut Di Operator *Select Attributes*

PARAMETER	ISI
Type	Include Attributes
Attribute filter type	A Subset
Select Attributes	Nama puskesmas, bps kode kecamatan, usia, jenis kusta, jumlah penderita kusta.

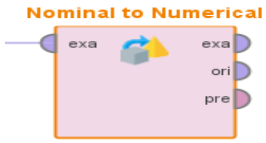
Hasil dari penggunaan operator *Select Attributes* dapat dilihat seperti gambar dibawah ini.



Gambar 7. Hasil Dari Operator Select Attributes

4.4 Data Transformations

Selanjutnya adalah mengubah tipe data dari *polynominal* ke *numeric* dengan menggunakan operator *Nominal to Numerical.*



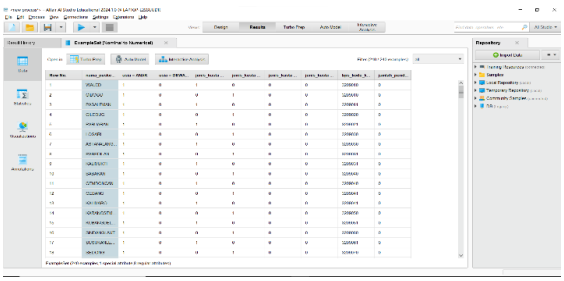
Gambar 8. Operator Nominal To Numerical

Melakukan pemilihan atribut yang akan diubah seperti pada tabel di bawah.

Tabel 3. Pemilihan Atribut Di Operator Nominal To Numerical

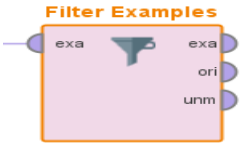
PARAMETER	ISI
Attribute filter type	A Subset
Select Attributes	Jenis kusta dan usia

Hasil dari penggunaan operator *Nominal to Numerical* seperti pada gambar dibawah.



Gambar 9. Hasil Dari Operator Nominal To Numerical

Selanjutnya adalah menyaring atau membersihkan data bernilai 0 pada atribut jumlah_penderita_kusta dengan menggunakan operator *Filter Examples.*



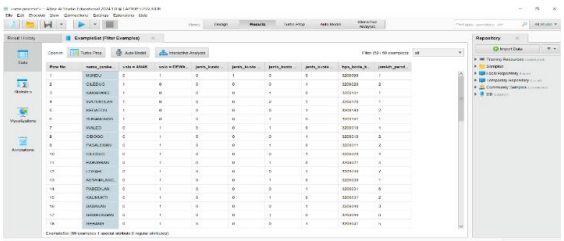
Gambar 10. Operator Filter Examples

Selanjutnya yaitu memfilter atribut jumlah_penderita_kusta seperti pada tabel di bawah.

Tabel 4. Pemilihan Atribut Di Operator Filter Examples

PARAMETER	ISI
Filters	Add Filters
Condition Class	Custom Filters

Hasil dari penggunaan operator *Filter Examples* seperti pada gambar dibawah ini.

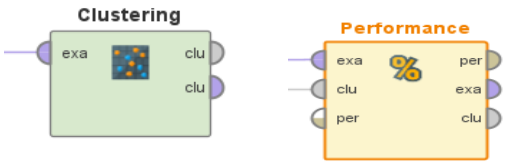


Gambar 11. Hasil Dari Operator Filter Examples

Pada gambar tersebut terlihat bahwa baris yang berisi nilai 0 pada atribut jumlah_penderita_kusta terlah dibersihkan dan pada atribut tersebut hanya terdapat angka saja.

4.5 Data Mining

Algoritma K-Means adalah salah satu metode *clustering* yang digunakan untuk mengelompokkan data ke dalam sejumlah kelompok atau *cluster* berdasarkan kesamaan antar data.



Gambar 12. Operator Clustering Dan Performance

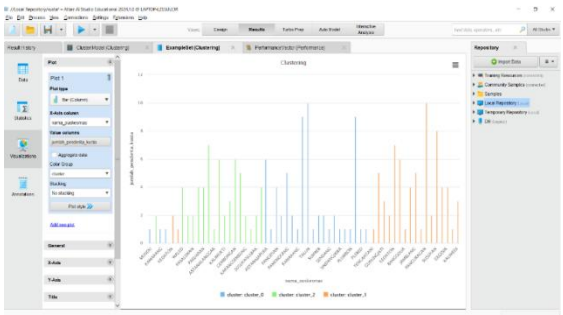
Selanjutnya mencari nilai k yang optimal menggunakan nilai DBI dengan cara melakukan percobaan hingga sebanyak 7 cluster.

Tabel 5. Nilai DBI

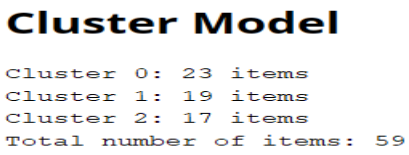
K	DBI
2	0.519
3	0.449
4	0.527
5	0.507
6	0.487
7	0.506

Nilai Davies-Bouldin Index (DBI) diperoleh dari hasil perhitungan otomatis yang dilakukan oleh algoritma K-Means di RapidMiner**.** Kemudian disalin hasil dari percobaan 2 cluster sampai 7 cluster pada operator *Performance*. Dari hasil berikut Nilai DBI terbaik (terendah) tercapai pada K = 3 (DBI = 0.449), menunjukkan bahwa tiga cluster menjadi jumlah klaster optimal berdasarkan metrik ini. Nilai DBI yang lebih rendah mengindikasikan bahwa cluster lebih kompak secara internal dan lebih terpisah dari cluster lainnya.

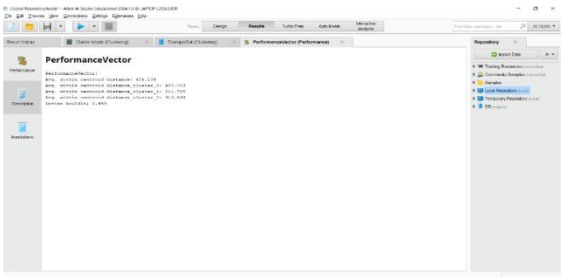
Pada hasil pengujian data terdapat beberapa output yang dihasilkan oleh software RapidMiner diantaranya yaitu:



Gambar 13. Tampilan Column Bar



Gambar 14. Tampilan Text View



Gambar 15. Tampilan Performance Vektor

4.6 Evaluasi

Proses pengolahan data yang dilakukan dengan metode K-Means ini membuat model *Clustering* dimana jumlah cluster yang dimodel adalah 2 – 7. Pada tiap cluster yang dibuat, di evaluasi dengan operator *Cluster Distance Performance* untuk mengetahui nilai DBI tiap cluster. Nilai DBI yang terkecil menunjukkan jumlah cluster yang optimal. Berikut hasil DBI tiap cluster :

Tabel 6. Nilai DBI Tiap Cluster

K	DBI
2	0.519
3	0.449
4	0.527
5	0.507
6	0.487
7	0.506

Pada pengujian cluster di RapidMiner menunjukan terdapat 3 cluster yang ditentukan, dan memiliki nilai rata-rata jarak centroidnya adalah 436.234, jarak rata-rata pada cluster 0 memiliki nilai 460.053, sedangkan jarak rata-rata pada cluster 1 memiliki nilai 511.729, dan jarak rata-rata pada cluster 2 memiliki nilai centroidnya 319.633 dengan nilai DBI nya adalah 0.449.

Berikut adalah karakteristik masing-masing klaster yang disajikan dalam bentuk tabel:

Tabel 7. Karakteristik Tiap Cluster

Karakteristik	Cluster 0	Cluster 1	Cluster 2
Usia Anak	*8.7%*	*10.5%*	*5.9%*
Usia Dewasa	*91.3%*	*89.5%*	*94.1%*
Jenis Kusta Kering	*0%*	*0%*	*0%*
Pausi Basiler/Kering	*43.5%*	*47.4%*	*41.2%*
Multi Basiler/Basah	*52.2%*	*52.6%*	*58.8%*
Rata-rata kasus/Kecamatan	2.826 kasus	3.684 kasus	3.353 kasus

5. KESIMPULAN

a) Penelitian ini berhasil menerapkan algoritma K-Means untuk mengelompokkan wilayah di Kabupaten Cirebon berdasarkan tingkat kerawanan kasus kusta. Pengelompokan menghasilkan tiga klaster utama: Cluster 0 (Rendah): Yang terdiri dari 23 anggota, wilayah dengan kasus kusta paling sedikit, menunjukkan kebutuhan intervensi yang relatif minimal. Cluster 1 (Tinggi): Yang terdiri dari 19 anggota, wilayah dengan jumlah kasus tertinggi, menjadi prioritas utama untuk tindakan kesehatan. Cluster 2 (Sedang): Yang terdiri dari 17 anggota, wilayah dengan tingkat kerawanan menengah, memerlukan pengawasan dan intervensi moderat.

b) Hasil dari pengelompokan ini menyediakan informasi yang relevan untuk memprioritaskan alokasi sumber daya kesehatan, terutama pada wilayah dengan tingkat kerawanan tinggi. Dengan adanya pembagian yang jelas, intervensi kesehatan dapat difokuskan pada area yang paling membutuhkan, sehingga meningkatkan efektivitas program penanganan kusta di Kabupaten Cirebon.

c) Pengukuran performa menggunakan Davies-Bouldin Index (DBI) menghasilkan nilai 0.449, yang menunjukkan hasil klasterisasi memiliki validitas yang baik. Hal ini membuktikan bahwa algoritma K-Means mampu secara akurat memetakan wilayah rawan kusta berdasarkan data epidemiologi.

Dalam konteks literatur terdahulu, temuan ini sejalan dengan penelitian sebelumnya yang menyoroti efektivitas algoritma K-Means dalam mengelompokkan wilayah berdasarkan faktor risiko kesehatan. [12]

UCAPAN TERIMA KASIH

Penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan, bimbingan, serta kontribusi dalam proses penelitian ini.

DAFTAR PUSTAKA

[1] D. Syaputri, P. H. Noprita, and S. Romelah, "Implementasi Algoritma K-Means untuk Pengelompokan Distribusi Sosial Ekonomi Masyarakat Berdasarkan Demografi Kependudukan," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 1–6, 2021, doi: 10.57152/malcom.v1i1.5.

[2] R. Kurniawan, M. S. Hasibuan, and R. Hasibuan, "Klasterisasi Wilayah Prioritas Vaksin Menggunakan Algoritma K-Means Clustering," *Media Online*, vol. 4, no. 3, pp. 1585–1592, 2023, doi: 10.30865/klik.v4i3.1334.

[3] N. T. Luchia, H. Handayani, F. S. Hamdi, D. Erlangga, and S. F. Octavia, "Perbandingan K-Means dan K-Medoids Pada Pengelompokan Data Miskin di Indonesia," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 2, pp. 35–41, 2022, doi: 10.57152/malcom.v2i2.422.

[4] A. Amrullah, I. Purnamasari, B. N. Sari, G. Garno, and A. Voutama, "Analisis Cluster Faktor Penunjang Pendidikan Menggunakan Algoritma K-Means (Studi Kasus: Kabupaten Karawang)," *J. Inform. dan Rekayasa Elektron.*, vol. 5, no. 2, pp. 244–252, 2022, doi: 10.36595/jire.v5i2.701.

[5] T. Asy Aria, M. Julkarnain, and F. Hamdani, "Penerapan Algoritma K-Means Clustering Untuk Data Obat," *Media Online*, vol. 4, no. 1, pp. 649–657, 2023, doi: 10.30865/klik.v4i1.1117.

[6] F. P. Dewi, P. S. Aryni, and Y. Umaidah, "Implementasi Algoritma K-Means Clustering Seleksi Siswa Berprestasi Berdasarkan Keaktifan dalam Proses Pembelajaran," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 7, no. 2, pp. 111–121, 2022, doi: 10.14421/jiska.2022.7.2.111-121.

[7] D. S. Saputri, G. M. Putra, and M. F. Larasati, "Implementation of the K-Means Clustering Algorithm for the Covid-19 Vaccinated Village in the Ujung Padang Sub-District," *J. Tek. Inform.*, vol. 3, no. 2, pp. 261–267, 2022,

[Online]. Available: https://doi.org/10.20884/1.jutif.2022.3.2.165

[8] Y. Filki, "Algoritma K-Means Clustering dalam Memprediksi Penerima Bantuan Langsung Tunai (BLT) Dana Desa," *J. Inform. Ekon. Bisnis*, vol. 4, no. 4, pp. 166–171, 2022, doi: 10.37034/infeb.v4i4.166.

[9] D. Marcelina, A. Kurnia, and T. Terttiaavini, "Analisis Klaster Kinerja Usaha Kecil dan Menengah Menggunakan Algoritma K-Means Clustering," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 293–301, 2023, doi: 10.57152/malcom.v3i2.952.

[10] N. N. Fransiska R, D. S. Anggraeni, and U. Enri, "Pengelompokkan Data Kemiskinan Provinsi Jawa Barat Menggunakan Algoritma K-Means dengan Silhouette Coefficient," *Tematik*, vol. 9, no. 1, pp. 29–35, 2022, doi: 10.38204/tematik.v9i1.901.

[11] R. N. Fahmi, M. Jajuli, and N. Sulistiyowati, "Analisis Pemetaan Tingkat Kriminalitas di Kabupaten Karawang menggunakan Algoritma K-Means," *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 4, no. 1, pp. 67–79, 2021, doi: 10.31539/intecoms.v4i1.2413.

[12] P. S. Rosiana, A. A. Mohsa, and Y. Umaidah, "Implementasi K-Means Dalam Pengelompokan Penyebaran Penyakit Dbd Di Jawa Barat," *J. Inform. dan Tek. Elektro Terap.*, vol. 11, no. 3, pp. 782–788, 2023, doi: 10.23960/jitet.v11i3.3344.