

## Analisis Pengelompokan kabupaten/kota di Provinsi Sumatera Utara berdasarkan Penyakit Menular Menggunakan Algoritma K-Means

Suriaty Padang<sup>1</sup>, Setiani Hulu<sup>2</sup>, Sardo Parningotan Sipayung<sup>3</sup>

<sup>1,2,3</sup>Universitas Katolik Santo Thomas  
Jln. Setia Budi No. 479 F Sumatera utara, Indonesia

e-mail: <sup>1</sup>[suriatyp12@gmail.com](mailto:suriatyp12@gmail.com), <sup>2</sup>[setianyhulu@gmail.com](mailto:setianyhulu@gmail.com), <sup>3</sup>[pinsarsiphom@gmail.com](mailto:pinsarsiphom@gmail.com).

Artikel Info : Diterima : 15-01-2026 | Direvisi : 19-01-2026 | Disetujui : 19-01-2026

**Abstrak** - Penyakit menular masih menjadi salah satu permasalahan utama kesehatan masyarakat di Provinsi Sumatera Utara. Penyebaran penyakit yang tidak merata antar kabupaten dan kota menyebabkan upaya penanggulangan penyakit menular belum berjalan secara optimal dan sering kali tidak tepat sasaran. Beberapa penyakit menular yang masih memerlukan perhatian khusus meliputi *Tuberculosis*, kusta, malaria, dan demam berdarah. Variasi tingkat kejadian penyakit tersebut menunjukkan perlunya suatu metode analisis yang mampu menggambarkan pola penyebaran penyakit secara terstruktur sebagai dasar dalam penentuan prioritas kebijakan kesehatan berbasis wilayah.

Penelitian ini menerapkan pendekatan *data mining* dengan metode *clustering* menggunakan algoritma K-Means untuk mengelompokkan kabupaten/kota di Provinsi Sumatera Utara berdasarkan karakteristik penyakit menular. Data yang digunakan mencakup indikator penemuan kasus *Tuberculosis*, tingkat keberhasilan pengobatan *Tuberculosis*, jumlah kasus kusta, tingkat kesakitan malaria, dan tingkat kesakitan demam berdarah. Wilayah kajian meliputi Kabupaten Tapanuli Tengah, Toba Samosir, Labuhanbatu, Simalungun, Dairi, Karo, Deli Serdang, Langkat, Nias Selatan, Pakpak Bharat, Serdang Bedagai, Batu Bara, Padang Lawas, dan Labuhanbatu Utara. Tahapan penelitian meliputi praproses data, penentuan jumlah *cluster*, perhitungan jarak menggunakan Euclidean Distance, serta proses iterasi hingga diperoleh pengelompokan yang stabil.

Hasil penelitian menunjukkan bahwa kabupaten/kota di Provinsi Sumatera Utara dapat dikelompokkan ke dalam tiga cluster, yaitu wilayah dengan tingkat penyakit menular tinggi, sedang, dan rendah. Pengelompokan ini diharapkan dapat menjadi dasar pendukung pengambilan keputusan dalam penentuan prioritas penanganan penyakit menular oleh pemerintah daerah.

**Kata kunci:** Penyakit menular, Data mining, clustering, k-means, Sumatera Utara

**Abstracts** - *Infectious diseases remain a major public health problem in North Sumatra Province. The uneven distribution of infectious diseases across districts and cities has resulted in suboptimal disease control efforts that are often not well targeted. Several infectious diseases that still require special attention include tuberculosis, leprosy, malaria, and dengue fever. Variations in disease incidence among regions indicate the need for an analytical method capable of describing disease distribution patterns in a structured manner to support regional-based health policy prioritization.*

*This study applies a data mining approach using clustering methods with the K-Means algorithm to group districts and cities in North Sumatra Province based on infectious disease characteristics. The data used include indicators of tuberculosis case detection, tuberculosis treatment success rates, the number of leprosy cases, malaria morbidity rates, and dengue fever morbidity rates. The study area covers Tapanuli Tengah, Toba Samosir, Labuhanbatu, Simalungun, Dairi, Karo, Deli Serdang, Langkat, Nias Selatan, Pakpak Bharat, Serdang Bedagai, Batu Bara, Padang Lawas, and Labuhanbatu Utara. The research stages consist of data preprocessing, determining the number of clusters, distance calculation using Euclidean Distance, and iterative processes until stable clustering results are obtained.*

*The results show that districts and cities in North Sumatra Province can be grouped into three clusters, namely regions with high, medium, and low levels of infectious diseases. This clustering is expected to support decision-making in determining priority areas for infectious disease control by local governments.*

**Keywords :** *infectious diseases, data mining, clustering, K-Means, North Sumatra*



## PENDAHULUAN

Penyakit menular masih menjadi salah satu permasalahan kesehatan masyarakat yang memerlukan perhatian serius, khususnya di wilayah dengan jumlah penduduk besar dan kondisi lingkungan yang beragam seperti Provinsi Sumatera Utara. Tingginya kepadatan penduduk, mobilitas masyarakat, serta perbedaan kondisi geografis dan kualitas pelayanan kesehatan menyebabkan penyebaran penyakit menular tidak merata antar kabupaten dan kota. Kondisi tersebut berdampak pada belum optimalnya upaya penanganan penyakit menular karena kebijakan kesehatan yang diterapkan cenderung bersifat umum dan belum sepenuhnya mempertimbangkan karakteristik wilayah secara spesifik.

Beberapa penyakit menular yang masih menjadi perhatian utama di Provinsi Sumatera Utara meliputi *Tuberculosis* (TB), kusta, malaria, dan demam berdarah dengue (DBD). Penyakit-penyakit tersebut memiliki karakteristik penyebaran yang berbeda, baik dari sisi mekanisme penularan, faktor resiko, maupun dampaknya terhadap kesehatan masyarakat dan sangat dipengaruhi oleh faktor lingkungan, kepadatan penduduk, sanitasi, iklim, serta kualitas pelayanan kesehatan. Data yang dipublikasikan oleh Badan Pusat Statistik menunjukkan bahwa tingkat kejadian penyakit menular di Provinsi Sumatera Utara bervariasi antar wilayah, baik di daerah perkotaan maupun pedesaan (*Kasus Penyakit Menurut Kabupaten\_Kota Dan Jenis Penyakit Di Provinsi Sumatera Utara, 2024, n.d.*). Variasi ini menunjukkan bahwa pendekatan penanganan penyakit menular tidak dapat disamaratakan, sehingga diperlukan analisis berbasis data yang komprehensif untuk memahami pola penyebaran penyakit berdasarkan wilayah.

Pendekatan berbasis data diperlukan agar pengolahan data kesehatan dapat menghasilkan informasi yang akurat, objektif, dan relevan bagi pengambilan keputusan. Pengolahan data secara sistematis dan terstruktur diharapkan mampu memberikan gambaran pola penyebaran penyakit menular yang lebih jelas, mengidentifikasi wilayah dengan tingkat kerentanan yang berbeda, serta mendukung pemerintah daerah dalam menyusun kebijakan kesehatan yang efektif dan tepat sasaran (Fitri et al., 2025). Selain itu, pendekatan ini juga penting untuk meningkatkan efisiensi alokasi sumber daya kesehatan dan mendukung perencanaan program pencegahan penyakit menular. Oleh karena itu, dibutuhkan metode analisis yang mampu menggali informasi tersembunyi dari data kesehatan dalam jumlah besar dan bersifat multidimensi.

Salah satu pendekatan yang dapat digunakan untuk tujuan tersebut adalah *data mining*. *Data mining* merupakan proses untuk mengekstraksi informasi, pola, atau pengetahuan yang bernilai dari kumpulan data berukuran besar menggunakan teknik dan algoritma tertentu (Ordila et al., 2020). Melalui proses data mining, data kesehatan yang sebelumnya hanya bersifat deskriptif dapat diolah menjadi informasi yang lebih bermakna. Dalam konteks kesehatan masyarakat, *data mining* telah banyak dimanfaatkan untuk menganalisis tren penyakit, memetakan wilayah berisiko, melakukan pengelompokan wilayah berdasarkan karakteristik kesehatan serta mendukung sistem pengambilan keputusan berbasis data. Salah satu teknik *data mining* yang sering digunakan adalah *clustering*, yaitu proses pengelompokan data ke dalam beberapa kelompok berdasarkan tingkat kemiripan karakteristik tertentu (Annas & Wahab, 2023).

Algoritma *K-Means* merupakan salah satu metode *clustering* yang paling populer dan banyak digunakan karena memiliki konsep yang sederhana, efisien, dan serta mudah diimplementasikan pada berbagai jenis data, khususnya data numerik. Algoritma ini bekerja dengan mengelompokkan data ke dalam sejumlah *cluster* berdasarkan jarak terdekat terhadap pusat *cluster* (*centroid*) menggunakan ukuran jarak tertentu, seperti *Euclidean Distance* (Ahmed et al., 2020). Selain itu, *K-Means* juga memiliki keunggulan dalam hal kecepatan komputasi sehingga cocok untuk digunakan untuk pengolahan data besar. Beberapa penelitian sebelumnya menunjukkan bahwa algoritma *K-Means* efektif digunakan dalam bidang kesehatan, antara lain untuk pengelompokan pasien rumah sakit berdasarkan tingkat keparahan penyakit (Keparahan et al., 2025), pengelompokan data rekam medis pasien berdasarkan jenis penyakit (Ordila et al., 2020), serta pemetaan wilayah resiko penyakit menular seperti demam berdarah (Mubarak & Kholijah, 2023).

Meskipun demikian, penelitian-penelitian sebelumnya umumnya masih berfokus pada satu jenis penyakit atau wilayah tertentu dan belum secara khusus mengelompokkan kabupaten dan kota di Provinsi Sumatera Utara berdasarkan beberapa indikator penyakit menular secara bersamaan. Selain itu, sebagian peneliti masih terbatas pada analisis *clustering* tanpa mengaitkannya secara langsung dengan upaya penentuan prioritas kebijakan wilayah dan pengambilan keputusan strategis. Padahal, pengguna beberapa indikator penyakit menular secara simultan dapat memberikan gambaran kondisi kesehatan wilayah yang lebih komprehensif. Hal ini menunjukkan adanya celah penelitian yang perlu dikaji lebih lanjut, khususnya dalam pemanfaatan algoritma *K-Means* untuk mendukung penentuan prioritas penanganan penyakit menular berbasis wilayah secara lebih komprehensif dan terintegrasi.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengelompokkan kabupaten dan kota di Provinsi Sumatera Utara berdasarkan karakteristik penyakit menular menggunakan algoritma *K-Means*. Nilai kebaruan penelitian ini terletak pada penggunaan beberapa indikator penyakit menular secara simultan untuk menghasilkan pengelompokan wilayah yang lebih komprehensif, objektif, dan berbasis data. Selain itu, penelitian ini juga diharapkan mampu memberikan gambaran tingkat kerentanan wilayah terhadap penyakit

menular serta membantu pemerintah daerah dalam menentukan daerah dalam menentukan prioritas. Hasil penelitian diharapkan dapat memberikan kontribusi dalam penyediaan informasi berbasis data yang mendukung perumusan kebijakan kesehatan masyarakat yang lebih terarah dan efektif.

## **METODE PENELITIAN**

Metodologi penelitian ini dirancang secara sistematis untuk mencapai tujuan penelitian, yaitu mengelompokkan kabupaten/kota di Provinsi Sumatera Utara berdasarkan tingkat penyakit menular. Tahapan utama penelitian ini meliputi akuisisi data, pra-pemrosesan data, normalisasi, proses clustering menggunakan algoritma K-Means, serta evaluasi hasil clustering menggunakan Davies–Bouldin Index (DBI).

### **1. Kerangka Penelitian**

Penelitian ini menggunakan metode *unsupervised learning* yaitu *K-Means clustering* untuk mengelompokkan kabupaten/kota di Provinsi Sumatera Utara berdasarkan pola penyakit menular. Clustering dipilih karena mampu menemukan kelompok dengan karakteristik serupa tanpa memerlukan label target (Annas & Wahab, 2023). K-Means secara luas digunakan pada data numerik kesehatan untuk mengidentifikasi pola persebaran penyakit yang dapat mendukung perencanaan intervensi kesehatan (Maharani Putri et al., 2025). Data penelitian berupa indikator penyakit menular seperti angka penemuan kasus Tuberkulosis (TBC), angka keberhasilan pengobatan TBC, penemuan kasus baru kusta per 100.000 penduduk, angka kesakitan malaria per 1.000 penduduk, dan angka kesakitan DBD per 100.000 penduduk. Data ini diambil dari portal resmi statistik Provinsi Sumatera Utara (<https://sumut.bps.go.id/id>) Setelah data dikumpulkan, dilakukan *pra-pemrosesan* termasuk penanganan missing value dan *normalisasi Min–Max* untuk menyetarakan skala antar atribut agar tidak memengaruhi perhitungan jarak dalam K-Means (Muhammad Raqib Syahkur et al., 2024). Proses clustering dilakukan dengan  $K = 3$  untuk memperoleh kluster wilayah dengan tingkat penyakit rendah, sedang, dan tinggi. Evaluasi kualitas hasil clustering dilakukan menggunakan *Davies–Bouldin Index (DBI)*, di mana nilai yang lebih rendah menunjukkan kluster yang lebih baik (*Cluster Evaluation Using DBI*, 2023).

### **2. Pengumpulan Data**

Data yang digunakan merupakan data penyakit menular di Provinsi Sumatera Utara yang meliputi Tuberkulosis (TB), kusta, malaria, dan demam berdarah dengue (DBD). Data diperoleh dari situs resmi Badan Pusat Statistik (BPS) Provinsi Sumatera Utara (<https://sumut.bps.go.id>) dalam bentuk data agregat tingkat kabupaten/kota. Dari total 35 kabupaten/kota, penelitian ini menggunakan 14 kabupaten/kota yang memiliki data lengkap pada kelima indikator tersebut. Pemilihan ini dilakukan untuk menjaga konsistensi dan kualitas data yang dianalisis. Data yang diperoleh disusun dalam bentuk tabel numerik dan digunakan sebagai data masukan (input) untuk proses pra-pemrosesan dan analisis clustering menggunakan algoritma K-Means (Mubarak & Kholijah, 2023)

### **3. Pra-Pemrosesan Data**

Sebelum dilakukan clustering, data diperiksa untuk memastikan tidak ada nilai kosong atau inkonsistensi. Seluruh atribut kemudian dinormalisasi menggunakan metode *Min–Max Normalization* sehingga nilai berada pada rentang 0–1. Normalisasi diperlukan agar semua variabel memiliki kontribusi yang seimbang dalam penghitungan jarak pada algoritma K-Means (Annas & Wahab, 2023; Sasmita & Muliono, 2025).

Rumus Min–Max:

$$V_n = \frac{V_n - Min}{Max - Min}$$

### **4. Algoritma K-Means**

Algoritma K-Means digunakan untuk mengelompokkan kabupaten/kota di Provinsi Sumatera Utara berdasarkan kemiripan indikator penyakit menular yang telah dinormalisasi. Proses pengelompokan dilakukan dengan menghitung jarak Euclidean antara setiap data dan pusat kluster (*centroid*), kemudian memperbarui centroid secara iteratif sampai terbentuk kluster yang stabil dan optimal (Ahmed et al., 2020). Prosesnya meliputi:

Menentukan jumlah kluster ( $K = 3$ )

1. Menentukan posisi centroid awal

2. Menghitung jarak setiap data ke centroid

$$d(x, c) = \sqrt{\sum_{i=1}^m (x_i - c_i)^2}$$

3. Menetapkan data ke kluster terdekat
4. Menghitung ulang centroid berdasarkan anggota kluster

$$c_k = \frac{1}{n_k} \sum_{i=1}^{n_k} x_i$$

5. Iterasi hingga centroid stabil

Algoritma ini diulangi hingga tidak terjadi perubahan signifikan pada posisi centroid dan keanggotaan kluster.

### 5. Evaluasi kluster

Kualitas hasil pengelompokan menggunakan algoritma K-Means dievaluasi dengan Davies–Bouldin Index (DBI). DBI merupakan metrik internal yang mengukur kekompakan kluster (intra-cluster) dan keterpisahan antar kluster (inter-cluster). Semakin kecil nilai DBI, semakin baik kualitas kluster yang dihasilkan karena menunjukkan kluster yang semakin kompak dan semakin terpisah satu sama lain. Perhitungan DBI dilakukan melalui dua tahap utama. Pertama, dihitung nilai dispersi tiap kluster ( $S_i$ ) sebagai rata-rata jarak data terhadap centroid kluster. Kedua, dihitung jarak antar centroid kluster ( $M_{ij}$ ). Nilai DBI dirumuskan sebagai: (Sholeh & Aeni, 2023)

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left( \frac{S_i + S_j}{M_{ij}} \right)$$

## HASIL DAN PEMBAHASAN

### 1. Deskripsi Data Penelitian

Data penelitian mencakup angka penemuan kasus dan tingkat kesakitan dari 14 kabupaten/kota yang bersumber dari BPS Provinsi Sumatera Utara. Terdapat variasi signifikan, misalnya Deli Serdang dengan angka TB tertinggi, sementara Nias Selatan menonjol pada kasus malaria dan DBD.

**Tabel 1. Sampel Jumlah Penderita Penyakit Menular di Sumatra utara.**

Kabupaten/Kota	Jumlah Kasus Penyakit - Angka Penemuan TBC	Jumlah Kasus Penyakit - Angka Keberhasilan Pengobatan TBC	Jumlah Kasus Penyakit - Penemuan Kasus Baru Kusta per 100.000 Penduduk	Jumlah Kasus Penyakit - Angka Kesakitan Malaria per 1.000 Penduduk	Jumlah Kasus Penyakit - Angka Kesakitan DBD per 100.000 Penduduk
Tapanuli Tengah	745	969	0.97	0.01	14.7
Toba Samosir	644	588	-	0.02	105.1
Labuhan Batu	1385	890	2.69	0.44	3.3
Simalungun	1950	1836	0.59	-	66.1
Dairi	1117	996	0.67	0.03	243.5
Karo	1222	701	-	0.01	53.3
Deli Serdang	5972	4640	0.24	0.02	46
Langkat	2566	2021	0.46	-	46
Nias Selatan	461	312	0.6	2.77	220
Pakpak Bharat	174	131	-	-	183.1
Serdang Bedagai	1287	1044	0.16	1.8	33.7
Batu Bara	923	733	2.52	1.53	57.6
Padang Lawas	626	580	1.96	-	38.9
Labuhan Batu Utara	720	781	1.86	1.13	113.1

Sebelum dilakukan proses clustering, data terdahulu melalui tahap praproses untuk menangani adanya data kosong pada indikator kusta dan malaria agar tidak mempengaruhi perhitungan jarak. Selanjutnya, dilakukan normalisasi data untuk menyamakan skala antara atribut dengan rentan nilai 0 hingga 1. Proses normalisasi ini bertujuan mencegah dominasi atribut tertentu dalam perhitungan jarak Euclidean pada algoritma K-Means, sehingga setiap indikator penyakit menular memiliki kontribusi yang seimbang dalam proses pengelompokan.

**2. Hasil Praproses dan Normalisasi Data**

Sebelum proses clustering, data melalui tahap praproses yang meliputi penanganan data kosong dan normalisasi. Normalisasi dilakukan untuk menyamakan skala antar atribut yang memiliki satuan dan rentang nilai berbeda, sehingga mencegah dominasi atribut tertentu dalam perhitungan jarak menggunakan *Euclidean Distance* pada algoritma K-Means. Tabel berikut menunjukkan hasil normalisasi data.

**Tabel 2. Hasil Normalisasi**

Kabupaten/Kota	X1	X2	X3	X4	X5
Tapanuli Tengah	0.0985	0.1859	0.3202	0.0000	0.0475
Toba Samosir	0.0811	0.1014	0.3953	0.0036	0.4238
Labuhan Batu	0.2089	0.1683	1.0000	0.1558	0.0000
Simalungun	0.3063	0.3781	0.1700	0.2790	0.2614
Dairi	0.1626	0.1918	0.2016	0.0072	1.0000
Karo	0.1808	0.1264	0.3953	0.0000	0.2082
Deli Serdang	1.0000	1.0000	0.0316	0.0036	0.1778
Langkat	0.4126	0.4192	0.1186	0.2790	0.1778
Nias Selatan	0.0495	0.0401	0.1739	1.0000	0.9022
Pakpak Bharat	0.0000	0.0000	0.3953	0.2790	0.7485
Serdang Bedagai	0.1920	0.2025	0.0000	0.6486	0.1266
Batu Bara	0.1292	0.1335	0.9328	0.5507	0.2261
Padang Lawas	0.0780	0.0996	0.7115	0.2790	0.1482
Labuhan Batu Utara	0.0942	0.1442	0.6719	0.4058	0.4571

Hasil normalisasi data pada Tabel di atas diperoleh melalui tahap praproses menggunakan metode *Min-Max Normalization* yaitu dengan mentransformasikan setiap nilai data ke dalam rentang 0 hingga 1 berdasarkan nilai minimum dan maksimum pada masing-masing variabel. Normalisasi ini bertujuan untuk menyamakan skala data agar tidak terjadi dominasi atribut tertentu dalam proses perhitungan jarak pada algoritma K-Means.

Selanjutnya, proses clustering dilakukan secara interaktif dengan memperbarui nilai centroid pada setiap iterasi. Pada Iterasi kedua, dilakukan untuk memperbarui centroid berdasarkan hasil iterasi pertama. Proses ini menghasilkan pengelompokan yang stabil, sehingga iterasi dihentikan dan hasil clustering dinyatakan optimal.

**Tabel 3. Hasil Normalisasi**

Kabupaten/Kota	Jumlah Kasus Penyakit - Angka Penemuan TBC	Jumlah Kasus Penyakit - Angka eberhasilan Pengobatan TBC	Jumlah Kasus Penyakit - Angka Penemuan Kasus Baru Kusta per 100.000 Penduduk	Jumlah Kasus Penyakit - Angka Kesakitan Malaria per 1.000 Penduduk	Jumlah Kasus Penyakit - Angka Kesakitan DBD per 100.000 Penduduk
Tapanuli Tengah	0.0985	0.1859	0.3202	0.0000	0.0475
Toba Samosir	0.0811	0.1014	0.3953	0.0036	0.4238
Labuhan Batu	0.2089	0.1683	1.0000	0.1558	0.0000
Simalungun	0.3063	0.3781	0.1700	0.2790	0.2614
Dairi	0.1626	0.1918	0.2016	0.0072	1.0000
Karo	0.1808	0.1264	0.3953	0.0000	0.2082

Deli Serdang	1.0000	1.0000	0.0316	0.0036	0.1778
Langkat	0.4126	0.4192	0.1186	0.2790	0.1778
Nias Selatan	0.0495	0.0401	0.1739	1.0000	0.9022
Pakpak Bharat	0.0000	0.0000	0.3953	0.2790	0.7485
Serdang Bedagai	0.1920	0.2025	0.0000	0.6486	0.1266
Batu Bara	0.1292	0.1335	0.9328	0.5507	0.2261
Padang Lawas	0.0780	0.0996	0.7115	0.2790	0.1482
Labuhan Batu Utara	0.0942	0.1442	0.6719	0.4058	0.4571

### 3. Hasil Clustering Menggunakan Algoritma K-Means

Proses pengelompokan wilayah dilakukan menggunakan algoritma K-Means dengan jumlah cluster sebanyak tiga cluster ( $K = 3$ ). Penentuan jumlah cluster ini bertujuan untuk mengelompokkan kabupaten/kota ke dalam kategori tingkat penyakit menular tinggi, sedang, dan rendah. Pengelompokan dilakukan secara interaktif dengan menghitung jarak setiap data terhadap pusat *cluster (centroid)* menggunakan *Euclidean Distance* sehingga diperoleh kondisi stabil, dimana tidak berubah ke anggotaan cluster. Berdasarkan hasil iterasi algoritma K-Means hingga mencapai kondisi stabil, seluruh kabupaten/kota di Provinsi Sumatera Utara berhasil dikelompokkan ke dalam tiga cluster yang memiliki karakteristik berbeda berdasarkan indikator penyakit menular yang digunakan.

**Tabel 4. Centroid**

Tabel Centroid Akhir Hasil Pengelompokan K-Means ( $K = 3$ )						
	Cluster	X1 (TBC)	X2 (Keberhasilan TBC)	X3 (Kusta)	X4 (Malaria)	X5 (DBD)
nilai tertinggi	C1	1.0000	1.0000	0.0316	0.0036	0.1778
nilai menengah	C2	0.2401	0.2600	0.1156	0.5516	0.3670
nilai rendah	C3	0.1148	0.1279	0.5582	0.1868	0.3622

### 4. Hasil Pengelompokan Wilayah

Berdasarkan hasil pengelompokan wilayah menggunakan algoritma K-Means, kabupaten dan kota di Provinsi Sumatera Utara terbagi ke dalam tiga cluster dengan karakteristik penyakit menular yang berbeda. Cluster 1 hanya terdiri dari Kabupaten Deli Serdang, yang memiliki nilai tertinggi pada indikator penemuan kasus dan keberhasilan pengobatan Tuberkulosis (X1 dan X2), sehingga menunjukkan dominasi beban kasus TB yang sangat tinggi dibandingkan wilayah lain dan menempatkannya sebagai daerah prioritas utama dalam penanganan penyakit menular. Cluster 2 mencakup Kabupaten Simalungun, Langkat, Nias Selatan, dan Serdang Bedagai, yang secara umum memiliki tingkat penyakit menular sedang dengan variasi indikator tertentu, seperti tingginya angka malaria dan DBD di Nias Selatan serta malaria di Serdang Bedagai, yang mengindikasikan adanya pengaruh faktor lingkungan dan geografis. Sementara itu, Cluster 3 mencakup sebagian besar wilayah lainnya, yaitu Tapanuli Tengah, Toba Samosir, Labuhan Batu, Dairi, Karo, Pakpak Bharat, Batu Bara, Padang Lawas, dan Labuhan Batu Utara, yang secara umum menunjukkan tingkat penyakit menular relatif lebih rendah, meskipun beberapa wilayah masih memiliki nilai tinggi pada indikator tertentu seperti kusta dan DBD. Hasil pengelompokan ini menegaskan adanya perbedaan tingkat risiko penyakit menular antar wilayah dan dapat dijadikan dasar dalam penentuan prioritas kebijakan kesehatan berbasis wilayah di Provinsi Sumatera Utara.

**Tabel 5. Hasil akhir pengelompokan kabupaten/kota menggunakan K-Means**

Kabupaten/Kota	X1 (TBC)	X2 (Keberhasilan TBC)	X3 (Kusta)	X4 (Malaria)	X5 (DBD)	HASIL CLUSTER
Deli Serdang	1	1	0.03162055 3	0.00362318 8	0.17776852 6	Cluster1

Simalungun	0.306 3	0.3781	0.1700	0.2790	0.2614	cluster2
Langkat	0.412 6	0.4192	0.1186	0.2790	0.1778	cluster2
Nias Selatan	0.049 5	0.0401	0.1739	1.0000	0.9022	cluster2
Serdang Bedagai	0.192 0	0.2025	0.0000	0.6486	0.1266	cluster2
Tapanuli Tengah	0.098 5	0.1859	0.3202	0.0000	0.0475	cluster3
Toba Samosir	0.081 1	0.1014	0.3953	0.0036	0.4238	cluster3
Labuhan Batu	0.208 9	0.1683	1.0000	0.1558	0.0000	cluster3
Dairi	0.162 6	0.1918	0.2016	0.0072	1.0000	cluster3
Karo	0.180 8	0.1264	0.3953	0.0000	0.2082	cluster3
Pakpak Bharat	0.000 0	0.0000	0.3953	0.2790	0.7485	cluster3
Batu Bara	0.129 2	0.1335	0.9328	0.5507	0.2261	cluster3
Padang Lawas	0.078 0	0.0996	0.7115	0.2790	0.1482	cluster3
Labuhan Batu Utara	0.094 2	0.1442	0.6719	0.4058	0.4571	cluster3

Evaluasi kualitas hasil clustering pada penelitian ini dilakukan menggunakan metode Davies–Bouldin Index (DBI), yang bertujuan untuk mengukur tingkat kekompakan data di dalam masing-masing cluster sekaligus menilai tingkat pemisahan antar cluster yang terbentuk. Berdasarkan hasil perhitungan, diperoleh nilai DBI sebesar 0,3051, yang menunjukkan bahwa hasil pengelompokan menggunakan algoritma K-Means memiliki kualitas yang baik. Nilai DBI yang relatif kecil mengindikasikan bahwa setiap cluster memiliki tingkat homogenitas yang tinggi serta batas pemisahan antar cluster yang cukup jelas. Pengelompokan yang dihasilkan membentuk tiga cluster yang merepresentasikan kategori tingkat penyakit menular tinggi, sedang, dan rendah, sehingga mencerminkan adanya perbedaan pola penyebaran penyakit menular antar kabupaten dan kota di Provinsi Sumatera Utara. Wilayah yang termasuk dalam cluster tingkat tinggi umumnya ditandai dengan nilai indikator Tuberkulosis dan beberapa penyakit menular lainnya yang relatif dominan, yang kemungkinan dipengaruhi oleh faktor kepadatan penduduk, mobilitas masyarakat yang tinggi, serta kondisi lingkungan dan sanitasi. Sementara itu, cluster tingkat sedang menunjukkan karakteristik nilai indikator penyakit menular yang lebih seimbang tanpa adanya dominasi satu jenis penyakit tertentu, namun tetap memerlukan perhatian karena berpotensi mengalami peningkatan kasus apabila upaya pencegahan tidak dilakukan secara berkelanjutan. Adapun cluster tingkat rendah menggambarkan wilayah dengan tingkat penyakit menular yang relatif lebih terkendali, meskipun tetap memerlukan pemantauan untuk menjaga stabilitas kondisi kesehatan masyarakat. Secara implementatif, hasil clustering ini dapat dimanfaatkan oleh pemerintah daerah dan instansi kesehatan sebagai dasar dalam penentuan prioritas penanganan penyakit menular berbasis wilayah, khususnya dalam perencanaan kebijakan, alokasi sumber daya kesehatan, penyusunan program pencegahan, serta peningkatan layanan kesehatan pada wilayah berisiko tinggi, sekaligus mempertahankan capaian kesehatan pada wilayah dengan tingkat penyakit menular rendah.

## KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penerapan tahapan data preprocessing dan algoritma K-Means mampu mengelompokkan wilayah kabupaten/kota di Provinsi Sumatera Utara ke dalam tiga cluster berdasarkan tingkat penyakit menular, yaitu cluster tingkat tinggi, sedang, dan rendah. Proses clustering menghasilkan kualitas pengelompokan yang baik, yang ditunjukkan oleh nilai Davies–Bouldin Index (DBI) sebesar 0,3051, menandakan bahwa cluster yang terbentuk memiliki tingkat kekompakan yang tinggi dan pemisahan antar cluster yang jelas. Hasil pengelompokan ini memperlihatkan adanya perbedaan karakteristik

penyakit menular antar wilayah yang dapat dijadikan dasar dalam perencanaan kebijakan kesehatan, khususnya dalam penentuan prioritas penanganan penyakit, alokasi sumber daya kesehatan, serta penyusunan program pencegahan yang lebih tepat sasaran. Dengan demikian, metode K-Means terbukti efektif dan layak digunakan sebagai pendekatan analitis dalam pemetaan dan pengambilan keputusan terkait pengendalian penyakit menular di Provinsi Sumatera Utara.

#### DAFTAR PUSTAKA

- Ahmed, M., Seraj, R., & Islam, S. M. S. (2020). The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics (Switzerland)*, 9(8), 1–12. <https://doi.org/10.3390/electronics9081295>
- Annas, M., & Wahab, S. N. (2023). Data Mining Methods: K-Means Clustering Algorithms. *International Journal of Cyber and IT Service Management (IJCITSM)*, 3(1), 40–47. <https://iiast.iaic-publisher.org/ijcitsm/index.php/IJCITSM/article/view/122>
- Fitri, A. A., Ula, M., & Agusniar, C. (2025). A Comparative Study of K-Means and K-Medoids for Clustering Dengue Fever Risk Areas in Medan. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 6(4), 193–201. <https://doi.org/10.30865/json.v6i4.8702>
- Kasus Penyakit Menurut Kabupaten\_Kota dan Jenis Penyakit di Provinsi Sumatera Utara, 2024.* (n.d.).
- Keparahan, T., Alasi, P.-G., Al, E., Alasi, G., & Putri, R. A. (2025). Penerapan K-Means Clustering untuk Pengelompokan Pasien Rumah Sakit berdasarkan Penerapan K-Means Clustering untuk Pengelompokan Pasien Rumah Sakit berdasarkan Tingkat Keparahan Penyakit. *Jurnal Multimedia Dan Teknologi Informasi*, 07(03), 12–475. <https://journal.cattleyadf.org/index.php/jatilima/index>
- Maharani Putri, F., Astuti, R., Prihartono, W., & Ryan Hamonangan. (2025). Application of the K-Means Method for Disease Clustering in Medical Records at Puskesmas Jatiwangi. *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, 4(2), 1263–1268. <https://doi.org/10.59934/jaiea.v4i2.888>
- Mubarak, H., & Kholijah, G. (2023). Analisis Klaster Dalam Pengelompokan Kabupaten/Kota Di Provinsi Jambi Berdasarkan Penyakit Menular Menggunakan Metode K-Means. *Jurnal Statistika Dan Komputasi*, 2(1), 20–30. <https://doi.org/10.32665/statkom.v2i1.1597>
- Muhammad Raqib Syahkur, Hartama, D., & Solikhun, S. (2024). Evaluasi Jumlah Cluster pada Algoritma K-Means++ Menggunakan Silhouette dan Elbow dengan Validasi Nilai DBI dalam Mengelompokkan Gizi Balita. *JST (Jurnal Sains Dan Teknologi)*, 13(3), 487–496. <https://doi.org/10.23887/jstundiksha.v13i3.86419>
- Ordila, R., Wahyuni, R., Irawan, Y., & Yulia Sari, M. (2020). PENERAPAN DATA MINING UNTUK PENGELOMPOKAN DATA REKAM MEDIS PASIEN BERDASARKAN JENIS PENYAKIT DENGAN ALGORITMA CLUSTERING (Studi Kasus : Poli Klinik PT.Inecda). *Jurnal Ilmu Komputer*, 9(2), 148–153. <https://doi.org/10.33060/jik/2020/vol9.iss2.181>
- Sholeh, M., & Aeni, K. (2023). Perbandingan Evaluasi Metode Davies Bouldin, Elbow dan Silhouette pada Model Clustering dengan Menggunakan Algoritma K-Means. *STRING (Satuan Tulisan Riset Dan Inovasi Teknologi)*, 8(1), 56. <https://doi.org/10.30998/string.v8i1.16388>