

CLUSTERING KASUS DEMAM BERDARAH DI KABUPATEN BOGOR MENGGUNAKAN K-MEANS

Habillah Darma, Didi Juardi, Asep Jamaludin

Informatika, Universitas Singaperbangsa Karawang

Jl. HS.Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361

hdhmmz@gmail.com

ABSTRAK

Demam Berdarah Dengue (DBD) merupakan penyakit menular yang masih menjadi permasalahan kesehatan di berbagai daerah, termasuk Kabupaten Bogor, yang mengalami peningkatan kasus setiap tahunnya. Penyebaran kasus yang tidak merata di tiap kecamatan menuntut adanya metode analisis yang dapat mengidentifikasi wilayah dengan tingkat risiko tinggi. Penelitian ini menerapkan metode K-Means Clustering untuk mengelompokkan kecamatan berdasarkan jumlah kasus DBD guna memahami pola persebarannya secara lebih sistematis. Data penelitian diperoleh dari Open Data Bogor dalam rentang tahun 2020 hingga 2022, melalui tahapan preprocessing data, penentuan jumlah cluster optimal dengan metode Elbow, serta penerapan algoritma K-Means. Hasil clustering menunjukkan adanya dua kelompok utama yang mengelompokkan kecamatan berdasarkan tingkat kasus DBD, yakni cluster pertama Kecamatan Cibinong dan yang kedua Kecamatan Cileungsi. Pengelompokan ini memberikan wawasan terkait daerah dengan potensi penyebaran tinggi, yang dapat menjadi dasar bagi upaya mitigasi dan intervensi yang lebih tepat sasaran.

Kata kunci : Demam Berdarah Dengue(DBD) , Clustering , K-Means

1. PENDAHULUAN

Demam Berdarah Dengue (DBD) merupakan salah satu penyakit menular yang disebabkan oleh virus dengue dan ditularkan melalui gigitan nyamuk *Aedes aegypti*. Penyakit ini menjadi permasalahan kesehatan yang serius di berbagai daerah tropis dan subtropis, termasuk di Indonesia. Kabupaten Bogor merupakan salah satu wilayah dengan angka kejadian DBD yang cukup tinggi, terutama saat musim hujan. Penyebaran kasus DBD yang tidak merata di setiap wilayah menuntut adanya analisis yang tepat guna mengetahui pola penyebaran dan tingkat risiko di masing-masing daerah[1]

Salah satu permasalahan utama dalam penyebaran DBD adalah pola persebaran kasus yang tidak merata di setiap wilayah. Beberapa kecamatan memiliki jumlah kasus yang jauh lebih tinggi dibandingkan dengan kecamatan lainnya. Hal ini dapat disebabkan oleh berbagai faktor, seperti kepadatan penduduk, kondisi lingkungan yang mendukung perkembangbiakan nyamuk, tingkat kesadaran masyarakat dalam menjaga kebersihan lingkungan, serta efektivitas program pengendalian vektor yang dilakukan oleh pemerintah daerah.

Dampak dari tingginya kasus DBD tidak hanya terbatas pada aspek kesehatan, tetapi juga berdampak secara sosial dan ekonomi. Dari segi kesehatan, peningkatan jumlah penderita DBD dapat membebani fasilitas layanan kesehatan, menyebabkan keterbatasan tempat perawatan di rumah sakit, serta meningkatkan angka kesakitan dan kematian, terutama pada anak-anak dan kelompok rentan. Secara sosial, tingginya kasus DBD dapat menurunkan produktivitas masyarakat karena banyak individu yang harus dirawat atau kehilangan waktu kerja akibat penyakit ini. Dari

sisi ekonomi, biaya perawatan pasien DBD, baik di tingkat individu maupun pemerintah, menjadi beban yang semakin meningkat.

Metode yang dapat digunakan untuk menganalisis dan mengelompokkan kasus DBD berdasarkan wilayah adalah metode clustering. Clustering merupakan teknik dalam data mining yang bertujuan untuk mengelompokkan data berdasarkan karakteristik yang sama. Dalam penelitian ini, algoritma K-Means digunakan untuk melakukan clustering terhadap data kasus DBD di Kabupaten Bogor. Algoritma ini bekerja dengan membagi data ke dalam sejumlah kluster berdasarkan tingkat kesamaan antar data, sehingga memungkinkan identifikasi wilayah dengan tingkat risiko tinggi, sedang, dan rendah[2].

Dengan menerapkan algoritma K-Means, penelitian ini bertujuan untuk memberikan gambaran pola kasus DBD di Kabupaten Bogor secara lebih sistematis. Hasil analisis ini diharapkan dapat membantu pihak terkait, seperti pemerintah daerah dan dinas kesehatan, dalam menentukan langkah-langkah strategis untuk pencegahan dan pengendalian DBD secara lebih efektif. Selain itu, hasil clustering juga dapat menjadi dasar dalam perencanaan kebijakan kesehatan masyarakat guna menekan angka kasus DBD di masa mendatang[3].

Melalui penelitian ini, diharapkan bahwa penggunaan metode clustering dengan algoritma K-Means dapat memberikan wawasan baru dalam analisis epidemiologi penyakit menular. Dengan memahami pola penyebaran DBD secara lebih akurat, upaya mitigasi dan penanggulangan penyakit ini dapat dilakukan dengan lebih optimal, sehingga dapat

mengurangi dampak kesehatan yang ditimbulkan di Kabupaten Bogor.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian sebelumnya digunakan sebagai acuan untuk penelitian ini. Penelitian yang dilakukan oleh Miftahul Arif Aldi dan Zaehol Fatah sebelumnya menunjukkan bahwa algoritma K-Means Clustering telah banyak diterapkan dalam pengelompokan data, termasuk dalam sektor pariwisata untuk menganalisis pola kunjungan wisatawan asing di Indonesia. Studi-studi terdahulu membuktikan bahwa metode ini efektif dalam mengidentifikasi tren kunjungan, membantu pengambilan keputusan strategis, serta memberikan wawasan terkait segmentasi wisatawan.

Studi tambahan yang dilakukan oleh Ilham Rinaldi dan Harmayani, menggunakan algoritma K-Means Clustering untuk menganalisis pola cuaca berdasarkan data suhu dan kelembapan dari BMKG. Hasil penelitian menunjukkan bahwa algoritma ini efektif dalam mengelompokkan data cuaca menjadi dua kategori utama, yaitu cuaca hujan dan cuaca dingin lembab, yang dapat digunakan untuk perencanaan dan mitigasi bencana.

Selain itu, Akhmad Thoriq Basalamah dan Resad Setyadi menganalisis tingkat penyelesaian pendidikan di 34 provinsi Indonesia dengan metode K-Means Clustering menggunakan perangkat lunak RapidMiner. Data diperoleh dari Badan Pusat Statistik (BPS) dan dikategorikan ke dalam tiga kluster: tinggi (K0), sedang (K1), dan rendah (K2). Hasil menunjukkan bahwa semakin tinggi jenjang pendidikan, semakin menurun tingkat penyelesaiannya, terutama pada tingkat SMA. Provinsi di Indonesia bagian timur, seperti Papua, Gorontalo, dan Nusa Tenggara Timur, memiliki tingkat penyelesaian pendidikan yang lebih rendah dibandingkan wilayah lain, sehingga memerlukan perhatian lebih dalam kebijakan pendidikan.

2.2. Demam Berdarah Dengue (DBD)

A Demam Berdarah Dengue (DBD) adalah penyakit yang disebabkan oleh virus dengue dan ditularkan melalui gigitan nyamuk *Aedes aegypti* dan *Aedes albopictus*. Penyakit ini menjadi masalah kesehatan global, terutama di daerah tropis dan subtropis, termasuk Indonesia. Gejala umum DBD meliputi demam tinggi, nyeri sendi, pendarahan ringan, hingga kondisi yang lebih parah seperti syok dengue. Pencegahan penyakit ini dilakukan melalui pengendalian vektor nyamuk dan upaya promotif dalam Masyarakat[4].

2.3. Clustering

Clustering adalah metode dalam analisis data yang bertujuan untuk mengelompokkan sekumpulan data berdasarkan kemiripan karakteristik tertentu. Dalam konteks penelitian ini, clustering digunakan untuk mengelompokkan wilayah berdasarkan jumlah

kasus DBD sehingga dapat diidentifikasi daerah dengan tingkat risiko tinggi, sedang, dan rendah. Clustering memungkinkan pemetaan pola distribusi kasus penyakit sehingga dapat membantu dalam pengambilan keputusan yang lebih efektif.[5]

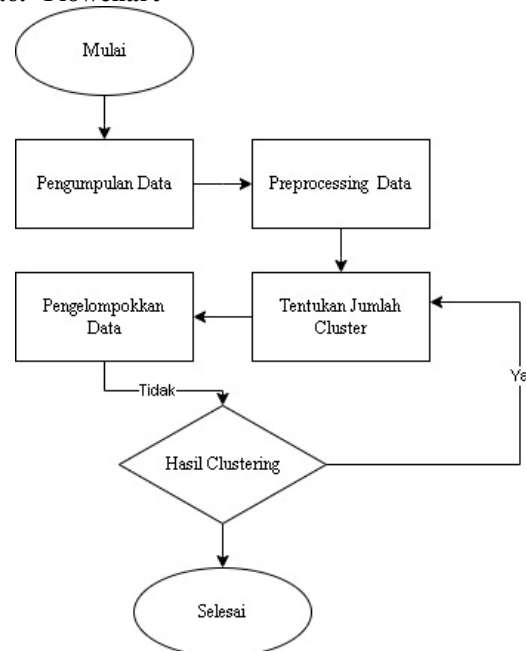
2.4. Data Mining

Data mining adalah proses menemukan pola atau informasi berharga dari kumpulan data yang besar. Metode ini sering digunakan dalam berbagai bidang, termasuk kesehatan, untuk menganalisis dan mengekstrak informasi dari data yang kompleks. Salah satu teknik dalam data mining adalah clustering, yang digunakan untuk mengelompokkan data berdasarkan karakteristik tertentu. Penerapan data mining dalam epidemiologi dapat membantu mengidentifikasi pola penyebaran penyakit dan menentukan langkah mitigasi yang lebih efektif[6].

2.5. K-Means

K-Means adalah algoritma clustering yang bekerja dengan cara menentukan sejumlah pusat kluster secara acak, lalu mengelompokkan data berdasarkan jarak terdekat ke pusat kluster tersebut. Proses iteratif dilakukan untuk menyesuaikan pusat kluster hingga mencapai konvergensi. Algoritma ini banyak digunakan dalam berbagai bidang, termasuk epidemiologi, untuk memahami pola penyebaran penyakit.[7]

2.6. Flowchart



Gambar 1 . Flowchart K-Means

Flowchart di atas menggambarkan proses clustering, dimulai dari pengumpulan data, lalu dilakukan preprocessing untuk membersihkan dan mempersiapkan data. Selanjutnya, jumlah cluster ditentukan sebelum data dikelompokkan sesuai

dengan metode clustering yang digunakan. Jika hasil clustering sudah sesuai, proses selesai, tetapi jika belum, data akan dikelompokkan kembali hingga hasilnya optimal. Tahapan algoritma K-means yang digunakan untuk menganalisis data kasus demam berdarah digambarkan dalam flowchart di atas.

3. METODOLOGI PENELITIAN

3.1. Pengumpulan Data

Pengumpulan Data dalam penelitian ini dilakukan dengan mengumpulkan informasi terkait kasus DBD di Kabupaten Bogor dari sumber yang valid dan terpercaya. Dataset tersebut di ambil dari OPEN DATA BOGOR.

3.2. Preprocessing Data

Preprocessing Data dalam penelitian ini bertujuan untuk memastikan data yang digunakan dalam clustering memiliki kualitas yang baik dan bebas dari kesalahan. Proses ini mencakup beberapa tahapan penting, seperti pembersihan data, di mana data yang tidak valid, duplikat, atau memiliki nilai kosong akan dihapus atau diperbaiki. Selanjutnya, dilakukan normalisasi data agar semua variabel berada dalam skala yang sama, sehingga tidak ada variabel yang mendominasi dalam proses clustering. Selain itu, data juga akan ditransformasi sesuai dengan kebutuhan analisis, misalnya dengan mengubah format data atau mengonversi data kategorikal menjadi numerik. Tahap preprocessing ini sangat penting karena akan meningkatkan akurasi hasil clustering dengan algoritma K-Means[8].

3.3. Tentukan Jumlah Cluster

Penentuan jumlah kluster optimal dalam algoritma K-Means dilakukan menggunakan Sum of Squared Errors (SSE), yang mengukur total jarak kuadrat antara setiap titik data dengan pusat klasternya. Nilai SSE yang lebih kecil menunjukkan bahwa data lebih terkelompok dengan baik. Titik optimal jumlah kluster ditentukan dari grafik Elbow, yaitu pada titik di mana penurunan SSE mulai melambat, menyerupai bentuk siku. Dengan metode ini, jumlah kluster yang dipilih dapat lebih akurat, sehingga hasil clustering lebih bermakna dalam memahami pola penyebaran kasus DBD di Kabupaten Bogor[9].

3.4. Pengelompokan Data

Pengelompokan data dalam penelitian ini dilakukan menggunakan algoritma K-Means, yang bertujuan untuk membagi wilayah di Kabupaten Bogor ke dalam beberapa kluster berdasarkan jumlah kasus DBD. Setelah jumlah kluster ditentukan, algoritma K-Means diterapkan dengan memilih pusat kluster awal secara acak, lalu setiap data dihitung jaraknya ke pusat kluster terdekat. Proses ini berulang hingga pusat kluster mengalami konvergensi, yaitu ketika tidak ada perubahan signifikan dalam pengelompokan data. Hasil clustering ini kemudian

dianalisis untuk mengidentifikasi daerah dengan tingkat risiko tinggi, sedang, dan rendah, sehingga dapat menjadi dasar bagi pemerintah dalam mengambil langkah pencegahan penyebaran DBD di Kecamatan atau Kabupaten Bogor [10].

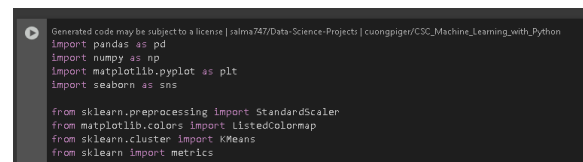
3.5. Hasil Clustering

Hasil clustering dalam penelitian ini berfokus pada dua variabel utama, yaitu kecamatan dan jumlah kasus DBD di Kabupaten Bogor. Setelah menerapkan algoritma K-Means, setiap kecamatan dikelompokkan ke dalam kluster berdasarkan tingkat kasus DBD yang terjadi. Kecamatan dengan jumlah kasus tinggi akan masuk dalam kluster berisiko tinggi, sedangkan yang lebih rendah akan dikelompokkan dalam kluster risiko sedang atau rendah[11].

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Setelah memperoleh dataset kasus DBD dari Open Data Bogor untuk periode 2020 hingga 2022, langkah selanjutnya adalah melakukan pemrosesan data menggunakan teknik machine learning. Pemrosesan ini dilakukan di Google Colab, yang telah terintegrasi dengan dataset tersebut, sehingga memungkinkan analisis data secara efisien dan optimal. Dalam tahap ini, data akan melalui serangkaian proses, termasuk pembersihan, normalisasi, serta penerapan algoritma clustering K-Means untuk mengelompokkan wilayah berdasarkan jumlah kasus DBD[12].



```
Generated code may be subject to a license | salma747/Data-Science-Projects | cuongpiger/CSC_Machine_Learning_with_Python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.preprocessing import StandardScaler
from matplotlib.colors import ListedColormap
from sklearn.cluster import KMeans
from sklearn import metrics
```

Gambar 2 . Import Library

- Disini kita memakai Library pandas , pandas tersebut untuk wadah menganalisis data.
- Disini kita memakai library Numpy menyediakan fungsi yang siap pakai untuk memudahkan kita melakukan perhitungan saintifik seperti matriks, aljabar, statistik, dan sebagainya.
- Matplotlib.pyplot adalah submoduledari library matplotlib yang digunakan untuk visualisasi grafik dan plot dalam python.
- Seaborn yakni Library yang dibangun di atas Matplotlib, khusus digunakan untuk visualisasi data dengan tampilan yang lebih menarik dan informatif.



kps_kode_kecamatan	kps_nama_kecamatan	kementerian_kode_kecamatan	kementerian_nama_kecamatan	jumlah_kasus_berdarah	status	tahun
3301140	SABAHAN MADANG	330105	SABAHAN MADANG	16 Orang	2020	
3301140	SABAHAN MADANG	330105	SABAHAN MADANG	34 Orang	2021	
3301140	SABAHAN MADANG	330105	SABAHAN MADANG	45 Orang	2022	
3301220	BOJONG GEDE	330113	BOJONG GEDE	44 Orang	2020	
3301220	BOJONG GEDE	330113	BOJONG GEDE	104 Orang	2021	

Gambar 3 . Dataset Dbd

Setelah dataset berhasil diimpor, kita dapat menampilkan data yang telah kita ambil untuk melihat struktur serta isinya sebelum melanjutkan ke tahap analisis lebih lanjut.

4.2. Preprocessing Data

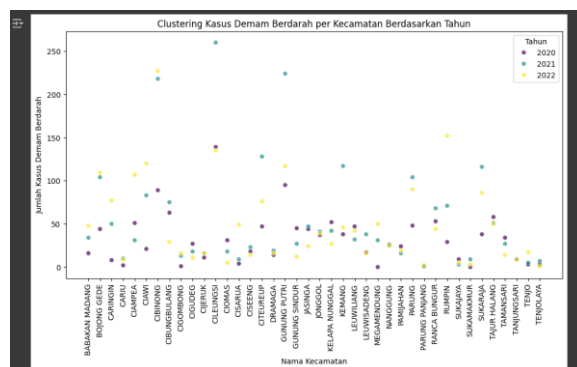
id	kode_provinsi	nama_provinsi	bps_kode_kabupaten_kota	
1	False	False	False	
2	False	False	False	
3	False	False	False	
4	False	False	False	
5	False	False	False	
...	...	...	...	
116	False	False	False	
117	False	False	False	
118	False	False	False	
119	False	False	False	
120	False	False	False	

id	bps_nama_kabupaten_kota	bps_kode_kecamatan	bps_nama_kecamatan	
1	False	False	False	
2	False	False	False	
3	False	False	False	
4	False	False	False	
5	False	False	False	
...	...	...	...	
116	False	False	False	
117	False	False	False	
118	False	False	False	
119	False	False	False	
120	False	False	False	

id	kemendagri_kode_kecamatan	kemendagri_nama_kecamatan	
1	False	False	
2	False	False	
3	False	False	
4	False	False	
5	False	False	
...	...	...	
116	False	False	
117	False	False	
118	False	False	
119	False	False	

Gambar 4 . Gambar Preprocessing

Untuk memastikan bahwa data yang diolah valid, kita perlu melakukan preprocessing data guna memeriksa dataset. Jika tidak terdapat nilai null, maka kita bisa langsung melanjutkan ke tahap berikutnya.



Gambar 5 . Scatter Plot Tahun

Scatter Plot digunakan untuk menggambarkan distribusi jumlah kasus per tahun di setiap kecamatan. Setiap tahun ditandai dengan warna yang berbeda, sehingga memudahkan dalam mengamati perubahan jumlah kasus dari waktu ke waktu.

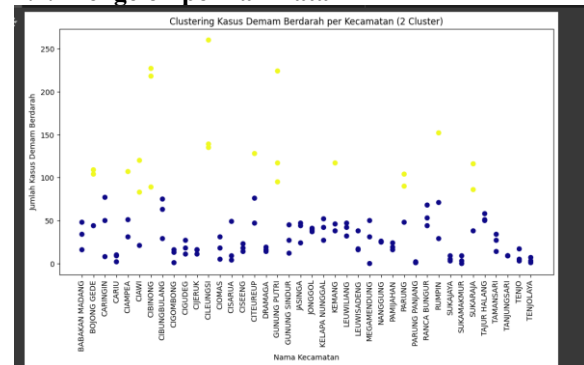
4.3. Tentukan Jumlah Cluster

```
[58] kmeans = KMeans(n_clusters=2,random_state=10)
      kmeans.fit(x)
```

Gambar 6 . Kmeans Clustering

Scatter plot yang ditampilkan pada gambar 5 , masih mencakup variabel tahun dengan variasi warna yang menunjukkan perbedaan tahun (2020, 2021, dan 2022). Namun, untuk proses clustering yang lebih sederhana dan fokus, saya akan mengabaikan variabel tahun dan hanya mempertimbangkan dua fitur utama, yaitu Nama Kecamatan dan Jumlah Kasus Demam Berdarah. Dengan demikian, hasil clustering akan lebih spesifik dalam mengelompokkan kecamatan berdasarkan tingkat kasus yang terjadi, tanpa adanya pengaruh variasi tahun. Pendekatan ini bertujuan untuk memperoleh pola atau kecenderungan kasus demam berdarah di setiap kecamatan secara lebih jelas dan terstruktur[13].

4.4. Pengelompokkan Data

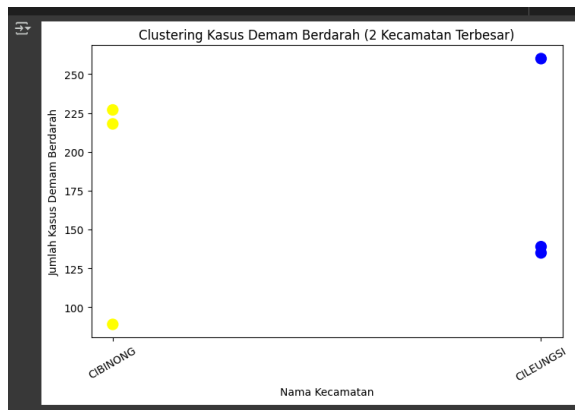


Gambar 7 . Scatter Plot 2 Clustering

Setelah dilakukan proses clustering menggunakan algoritma K-Means dengan jumlah cluster sebanyak 2, diperoleh hasil seperti yang ditampilkan pada scatter plot di atas. Tidak seperti scatter plot sebelumnya yang masih mempertimbangkan variabel tahun, pada visualisasi ini clustering hanya difokuskan pada dua variabel utama, yaitu Nama Kecamatan dan Jumlah Kasus Demam Berdarah. Tujuannya adalah untuk mengelompokkan kecamatan berdasarkan intensitas kasus demam berdarah yang tercatat, tanpa adanya pengaruh perbedaan waktu atau tahun[14].

4.5. Hasil Clustering

Dari hasil clustering tersebut, terlihat adanya dua kelompok data yang cukup jelas. Cluster pertama (berwarna biru tua) merepresentasikan kecamatan-kecamatan dengan jumlah kasus yang cenderung rendah hingga sedang. Sementara itu, cluster kedua (berwarna kuning) menggambarkan kecamatan dengan jumlah kasus yang tergolong tinggi. Visualisasi ini membantu untuk melihat pola distribusi kasus secara lebih jelas dan memberikan gambaran umum kecamatan mana saja yang tergolong rawan terhadap penyebaran demam berdarah[15].



Gambar 8 . Scatter Plot Final 2 Clustering

Setelah dilakukan proses clustering menggunakan algoritma K-Means dengan jumlah cluster sebanyak 2, diperoleh hasil seperti yang ditampilkan pada scatter plot di atas. Berbeda dengan scatter plot sebelumnya yang mencakup seluruh kecamatan, visualisasi ini lebih terfokus pada dua kecamatan dengan jumlah kasus tertinggi, yaitu Cibinong dan Cileungsi. Clustering dilakukan berdasarkan dua variabel utama, yakni Nama Kecamatan dan Jumlah Kasus Demam Berdarah, tanpa mempertimbangkan faktor waktu atau tahun.

Dari hasil clustering tersebut, terlihat adanya dua kelompok data yang cukup jelas. Cluster pertama (berwarna biru tua) merepresentasikan kecamatan dengan jumlah kasus dari yang terendah sampai yang tertinggi di angka 260, sementara cluster kedua (berwarna kuning) sama seperti cluster pertama merepresentasikan kecamatan dengan jumlah kasus dari yang terendah sampai yang tertinggi di angka 227.

4.6. Pembahasan

Hasil pembahasan menunjukkan bahwa setelah dilakukan clustering menggunakan algoritma K-Means dengan jumlah cluster sebanyak 2, data kasus Demam Berdarah berhasil dikelompokkan berdasarkan jumlah kasus yang terjadi di dua kecamatan dengan angka kasus tertinggi, yaitu Cibinong dan Cileungsi.

Dari hasil clustering ini, diperoleh dua kelompok utama, yaitu:

- Cluster pertama (berwarna biru tua) → Merepresentasikan kecamatan dengan jumlah kasus dari yang terendah hingga yang tertinggi mencapai 260 kasus.
- Cluster kedua (berwarna kuning) → Merepresentasikan kecamatan dengan jumlah kasus dari yang terendah hingga yang tertinggi mencapai 227 kasus.

Perbedaan jumlah kasus pada kedua cluster ini menunjukkan bahwa terdapat variasi tingkat penyebaran Demam Berdarah di dua kecamatan tersebut. Dengan adanya hasil clustering ini, terjawab sudah dari penelitian kami bahwa pengelompokan berdasarkan jumlah kasus dapat memberikan

gambaran yang lebih jelas mengenai tingkat persebaran penyakit di setiap wilayah.

Hasil ini membuktikan bahwa metode K-Means Clustering mampu mengelompokkan daerah dengan karakteristik yang berbeda, sehingga dapat membantu pihak berwenang dalam menetapkan prioritas intervensi kesehatan.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil clustering menggunakan algoritma K-Means, dapat disimpulkan bahwa terdapat dua kelompok kecamatan dengan tingkat kasus Demam Berdarah yang berbeda. Cluster pertama (biru tua) merepresentasikan kecamatan dengan jumlah kasus mulai dari yang terendah hingga tertinggi di angka 260 kasus, sedangkan cluster kedua (kuning) memiliki jumlah kasus dari yang terendah hingga tertinggi di angka 227 kasus. Hasil ini menunjukkan bahwa terdapat perbedaan signifikan dalam tingkat persebaran kasus di wilayah tersebut, sehingga metode clustering ini dapat membantu dalam memetakan daerah dengan risiko tinggi serta menjadi dasar bagi perencanaan strategi pencegahan yang lebih tepat.

Oleh karena itu, disarankan agar pihak terkait seperti pemerintah daerah, instansi kesehatan, dan masyarakat dapat lebih aktif dalam melakukan upaya pencegahan dan pengendalian Demam Berdarah. Wilayah yang termasuk dalam cluster dengan jumlah kasus tinggi sebaiknya menjadi prioritas dalam intervensi kesehatan, seperti peningkatan kesadaran masyarakat melalui edukasi, pemberantasan sarang nyamuk, serta optimalisasi program fogging dan sanitasi lingkungan. Selain itu, penelitian lebih lanjut perlu dilakukan untuk memperbaiki kekurangan dalam studi ini, seperti mempertimbangkan faktor lain yang dapat memengaruhi penyebaran kasus, misalnya kepadatan penduduk, kondisi lingkungan, dan faktor cuaca. Dengan demikian, analisis yang lebih komprehensif dapat dilakukan untuk meningkatkan efektivitas strategi penanggulangan penyakit ini di masa mendatang.

DAFTAR PUSTAKA

- [1] T. Tendean and W. Purba, "Analisis Cluster Provinsi Indonesia Berdasarkan Produksi Bahan Pangan Menggunakan Algoritma K-Means," *Jurnal Sains dan Teknologi*, vol. 1, no. 2, pp. 5–11, 2020.
- [2] A. Yahya and R. Kurniawan, "Implementasi Algoritma K-Means untuk Pengelompokan Data Penjualan Berdasarkan Pola Penjualan: Implementation of K-Means Algorithm for Clustering Sales Data Based on Sales Patterns," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 1, pp. 350–358, 2025.
- [3] S. Salsabila and L. A. Susanti, "OPTIMASI PENGELOMPOKAN PROVINSI BERDASARKAN INDIKATOR SDGs MENGGUNAKAN K-MEANS DAN

- ANALISIS FEATURE IMPORTANCE,” *Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika*, vol. 6, no. 1, pp. 51–59, 2025.
- [4] M. A. Ramadhan, E. Poerwandono, Y. Akbar, and A. Z. Hidayat, “Analisis Clustering Penyakit Menular pada Manusia di Jakarta Timur Menggunakan Algoritma K-Means,” *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, vol. 9, no. 1, pp. 52–58, 2025.
- [5] G. M. M. Sujak, H. N. Rofiq, and F. I. Tawakal, “Implementasi K-Means Clustering untuk Optimalisasi Anggaran Penyakit Tidak Menular: Implementation of K-Means Clustering for Optimizing Non-Communicable Disease Budgets,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 1, pp. 67–74, 2025.
- [6] I. Rinaldi and H. Harmayani, “PENGUNAAN ALGORITMA CLUSTERING K-MEANS UNTUK MENGELOMPOKKAN POLA CUACA,” *JOURNAL OF SCIENCE AND SOCIAL RESEARCH*, vol. 8, no. 1, pp. 343–348, 2025.
- [7] M. A. Aldi and Z. Fatah, “IMPLEMENTASI K-MEANS CLUSTERING DALAM PENGELOMPOKAN DATA KUNJUNGAN WISATAWAN ASING DI INDONESIA,” *Jurnal Ilmiah Multidisiplin Ilmu*, vol. 2, no. 1, pp. 13–19, 2025.
- [8] S. H. Abdullah and Z. Fatah, “Analisis Produksi Cabai Rawit Indonesia Menggunakan Algoritma K-Means Clustering,” *JURNAL ILMIAH SAINS TEKNOLOGI DAN INFORMASI*, vol. 3, no. 1, pp. 66–74, 2025.
- [9] A. Septianingsih, “Analisis k-means clustering pada pemetaan provinsi indonesia berdasarkan indikator rumah layak huni,” *Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika*, vol. 3, no. 1, pp. 224–241, 2022.
- [10] 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: Comparison of K-Means and K-Medoids on Poor Data Clustering in Indonesia,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 2, no. 2, pp. 35–41, 2022.
- [11] W. Rosida and Y. A. Wijaya, “Klasterisasi Penyakit HIV/AIDS di Jawa Barat Menggunakan Algoritma K-Means Clustering,” *Blend Sains Jurnal Teknik*, vol. 1, no. 4, pp. 306–315, 2023.
- [12] B. Nurhakim, I. Septiani, K. Anam, and D. Pratama, “Penerapan Algoritma K-Means Clustering Dalam Menganalisis Resiko Penyakit Stroke,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 318–322, 2024.
- [13] A. Basalamah and R. Setyadi, “Penerapan algoritma K-Means clustering pada tingkat penyelesaian pendidikan di Provinsi Indonesia,” *Jurnal Informatika dan Teknologi Komputer (J-ICOM)*, vol. 4, no. 2, pp. 114–121, 2023.
- [14] K. A. Kholil, N. Rahaningsih, and R. D. Dana, “PENERAPAN DATA MINING UNTUK CLUSTERING PENYAKIT DIARE MENGGUNAKAN ALGORITMA K-MEANS,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 3, pp. 3124–3131, 2024.
- [15] P. S. Rosiana, A. A. Mohsa, and Y. Umaidah, “Implementasi K-Means Dalam Pengelompokan Penyebaran Penyakit Dbd Di Jawa Barat,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 11, no. 3, 2023.