



Pemanfaatan Algoritma K-Means untuk Klastering Spasial Beban Kasus Pneumonia pada Kelompok Balita di Wilayah dengan Kepadatan Populasi Tinggi

Rosi Windi Chrisamudra, Safrizal Abdurrahman*

Fakultas Teknik dan Desain, Sistem Informasi, Universitas Pembangunan Jaya, Tangerang Selatan, Indonesia

Email: ¹rosi.windichrisamudra@student.upj.ac.id, ^{2,*}safrizal.abdurrahman@upj.ac.id

Email Penulis Korespondensi: safrizal.abdurrahman@upj.ac.id

Abstrak—Pneumonia merupakan salah satu penyebab utama kematian balita di Indonesia, termasuk di Kota Bandung yang mencatat ribuan kasus setiap tahun. Penelitian ini bertujuan untuk menganalisis dan mengelompokkan kasus pneumonia pada balita menggunakan algoritma K-Means berdasarkan data sekunder dari Open Data Bandung tahun 2023. Tahapan penelitian meliputi pra-pemrosesan data dengan variabel utama Beban Kasus Absolut, penentuan jumlah kluster optimal dengan Metode Silhouette, penerapan K-Means dengan metrik *Euclidean Distance*, serta evaluasi hasil menggunakan Indeks Davies-Bouldin (DBI). Hasil penelitian menunjukkan tiga kluster risiko yaitu rendah (12 kecamatan), sedang (12 kecamatan), dan tinggi (6 kecamatan) dengan Kecamatan Arcamanik, Bandung Kidul, Bandung Kulon, Buahbatu, Cibeunying Kidul, dan Kiaracondong teridentifikasi sebagai wilayah dengan risiko tinggi. Evaluasi kualitas kluster menghasilkan nilai DBI sebesar 0,487, yang menunjukkan pemisahan kluster cukup baik. Kesimpulan dari penelitian ini adalah teknik *clustering* dapat digunakan sebagai alat bantu analisis spasial untuk mendukung kebijakan kesehatan berbasis data, yang diharapkan menjadi acuan bagi pemerintah daerah dalam mengoptimalkan alokasi sumber daya dan merancang intervensi pencegahan pneumonia yang lebih efektif.

Kata Kunci: Pneumonia; Kematian Balita; Analisis; K-Means; Klastering

Abstract—Pneumonia is one of the leading causes of death among toddlers in Indonesia, including in the city of Bandung, which records thousands of cases each year. This study aims to analyze and group cases of pneumonia in toddlers using the K-Means algorithm based on secondary data from Open Data Bandung in 2023. The research stages included data pre-processing with the main variable of Absolute Case Load, determining the optimal number of clusters using the Silhouette Method, applying K-Means with the Euclidean Distance metric, and evaluating the results using the Davies-Bouldin Index (DBI). The results show three risk clusters low (12 subdistricts), medium (12 subdistricts), and high (6 subdistricts) with Arcamanik, Bandung Kidul, Bandung Kulon, Buahbatu, Cibeunying Kidul, and Kiaracondong subdistricts identified as high-risk areas. The cluster quality evaluation produced a DBI value of 0.487, indicating fairly good cluster separation. The conclusion of this study is that clustering techniques can be used as a spatial analysis tool to support data-based health policies, which are expected to serve as a reference for local governments in optimizing resource allocation and designing more effective pneumonia prevention interventions.

Keywords: Pneumonia; Child Mortality; Analysis; K-Means; Clustering

1. PENDAHULUAN

Pneumonia, sering dijuluki sebagai "pembunuh balita yang terlupakan" (*the forgotten killer of children*), tetap menjadi tantangan kesehatan masyarakat yang signifikan di Indonesia. Penyakit ini merupakan infeksi pernapasan akut yang menyerang paru-paru dan tercatat sebagai penyebab utama morbiditas dan mortalitas balita di dunia, menyumbang persentase signifikan dari seluruh kematian anak di bawah lima tahun (Husna M et al., 2022). Di tingkat nasional, Indonesia mencatat ribuan kasus setiap tahun, dan Kota Bandung merupakan salah satu kota penyumbang kasus tertinggi di Jawa Barat. Fenomena peningkatan prevalensi kasus di wilayah perkotaan, seperti yang teramati di Kota Semarang pada tahun 2020, menunjukkan bahwa masalah ini bersifat sistemik dan membutuhkan penanganan yang berkelanjutan. Dalam konteks Kota Bandung, pertumbuhan penduduk yang pesat di berbagai kecamatan menambah tantangan dalam memastikan distribusi sarana dan pelayanan kesehatan yang merata. Meskipun Pusat Kesehatan Masyarakat (Puskesmas) berfungsi sebagai garda terdepan dalam upaya promotif dan preventif, efektivitas intervensi sangat bergantung pada kualitas pelayanan dan kemampuan analisis data yang mendalam (Ramadhan et al., 2021).

Masalah utama yang dihadapi dalam penanggulangan pneumonia adalah distribusi kasus yang tidak merata (*heterogeneous distribution*), yang jika tidak dianalisis secara spasial dapat menyebabkan alokasi sumber daya dan tenaga kesehatan yang tidak optimal (Rakuasa et al., 2021). Laporan rutin dari fasilitas kesehatan, yang dicatat dalam Sistem Informasi Kesehatan (SIK) atau hasil Surveilans Kesehatan yang di *publish* pada Open Data Bandung, menyediakan data mentah yang harus diolah lebih lanjut untuk mengklasifikasikan wilayah berdasarkan kesamaan tingkat beban kasus sebagai dasar penentuan prioritas. Tantangan implementasi SIK di fasilitas pelayanan kesehatan sering kali terkendala oleh kurangnya kuantitas dan kualitas sumber daya manusia, sehingga menyulitkan pengambilan keputusan yang berbasis data (Chotimah S, 2022). Mengingat kejadian pneumonia dipengaruhi oleh faktor risiko tidak langsung seperti kondisi lingkungan fisik rumah (kepadatan, ventilasi), status gizi anak, dan tingkat sosial ekonomi, identifikasi wilayah yang memiliki karakteristik kasus serupa menjadi langkah krusial (Abimayu & Rahmawati, 2023).

Kajian penelitian sejenis telah banyak dilakukan dengan menggunakan teknik *clustering* dan *machine learning* untuk analisis penyakit dan sumber daya kesehatan (Abdillah, Susilo, & Ihksan, 2023; Akbar et al., 2025). Penelitian sebelumnya telah menerapkan algoritma K-Means untuk pengelompokan penyebaran pneumonia pada balita di Kota Bandung (Vitalaya & Prasetio, 2020), klasterisasi gejala ISPA (Ariyanto, 2022), serta pemetaan kasus penyakit



menggunakan analisis sebaran spasial (Rakuasa et al., 2021; Rosiana et al., 2023). Penelitian juga mengulas pemanfaatan teknologi *data mining* dan *machine learning* untuk diagnosis dan klasifikasi data kesehatan (Abdillah, Susilo, & Ihksan, 2023; Akbar et al., 2025). Meskipun studi terdahulu telah membuktikan kemampuan metodologis K-Means, sebagian besar studi terbatas pada data insiden lama atau tidak secara eksplisit memfokuskan klastering pada kebutuhan kebijakan lokal yang mendesak.

Pendekatan *clustering* dipilih karena lebih relevan untuk identifikasi daerah rawan tanpa adanya label kelas yang telah ditentukan sebelumnya, berbeda dengan metode klasifikasi. Perbedaan mendasar (*research gap*) dari penelitian ini dengan studi sebelumnya terletak pada: pertama, analisis kluster risiko terbaru per kecamatan di Kota Bandung berdasarkan metrik beban kasus absolut tahun 2023, merefleksikan kondisi aktual pasca-pandemi dan dinamika pertumbuhan kota terbaru (Vitalaya & Prasetio, 2020). Kedua, penelitian ini secara eksplisit mengintegrasikan metode Silhouette untuk penentuan kluster optimal dan Indeks Davies-Bouldin (DBI) sebagai metrik evaluasi yang ketat, guna memastikan kualitas pemisahan kluster yang valid dan bermakna (Ghazal et al., 2021).

Urgensi penelitian ini terletak pada pentingnya penyediaan peta kerawanan yang diperbarui dan terverifikasi berdasarkan kondisi kasus terkini sebagai dasar pengambilan keputusan yang efektif oleh Pemerintah Kota Bandung, khususnya dalam upaya peningkatan kualitas pelayanan Puskesmas. Solusi yang ditawarkan adalah penerapan algoritma *clustering* K-Means sebagai alat bantu analisis spasial. Solusi yang ditawarkan adalah penerapan algoritma *clustering* K-Means sebagai alat bantu analisis spasial. K-Means dipilih karena kemampuannya menghasilkan kluster yang mudah diinterpretasikan menjadi kategori risiko (rendah, sedang, tinggi) berdasarkan metrik beban kasus absolut, yang merupakan metrik paling langsung untuk menentukan prioritas alokasi sumber daya operasional di lapangan. Hasil *klastering* ini, yang kemudian divisualisasikan menggunakan Sistem Informasi Geografis (GIS), merupakan *state of the art* dari penelitian ini, menyediakan peta kerawanan yang intuitif dan mudah dipahami oleh pemangku kebijakan. Kontribusi utama dari penelitian ini adalah penyediaan model pengelompokan berbasis data yang valid, yang secara spesifik mengidentifikasi enam kecamatan berisiko tinggi (*hotspot*) untuk dijadikan acuan ilmiah bagi Dinas Kesehatan Kota Bandung dalam mengoptimalkan alokasi sumber daya dan merancang intervensi pencegahan pneumonia yang lebih efektif dan terarah.

2. METODOLOGI PENELITIAN

2.1 Kerangka Dasar Penelitian

2.1.1 Jenis dan Pendekatan Penelitian

Penelitian mengadopsi pendekatan kuantitatif sebagai fondasi utama analisis. Pendekatan ini berfokus pada proses ekstraksi pengetahuan tersembunyi dari kumpulan data besar secara otomatis atau semi-otomatis, dengan tujuan untuk mengklasifikasikan unit wilayah berdasarkan kesamaan tingkat beban kasus (Abdillah et al., 2023). Dalam ranah ilmu komputer, proses ini seringkali dikaitkan dengan disiplin *data mining* dan *machine learning*. Tujuannya adalah untuk mengungkap pola, tren, dan hubungan yang mungkin tidak terlihat melalui metode statistik deskriptif konvensional

Secara spesifik dalam konteks kesehatan, pendekatan kuantitatif ini memiliki manfaat yang sangat besar untuk menganalisis data klinis dan administratif dalam jumlah besar, seperti catatan kasus pneumonia. Hal ini dapat memberikan wawasan berharga untuk peningkatan kualitas layanan kesehatan dan perencanaan strategis (Akbar et al., 2025). Hal ini memungkinkan pengambil keputusan untuk bergerak dari sekadar pelaporan data menuju analisis yang lebih mendalam dan prediktif.

Pendekatan yang diaplikasikan dalam penelitian ini adalah *clustering* atau pengelompokan. *Clustering* merupakan salah satu teknik *unsupervised learning* dalam *machine learning*, yang sangat tepat digunakan ketika tidak ada label kelas atau kategori yang telah ditentukan sebelumnya pada data (Ariyanto, 2022). Metode ini dipilih karena relevansinya untuk mengidentifikasi daerah rawan (*hotspot*) risiko secara alami dari pola data yang ada. Ujian fundamental dari *clustering* adalah untuk mengelompokkan data ke dalam beberapa grup atau kluster, di mana objek data di dalam satu kluster memiliki tingkat kemiripan yang tinggi satu sama lain (kohesif), namun memiliki tingkat kemiripan yang rendah dengan objek data yang berada di kluster lain (separasi tinggi) (Ghazal et al., 2021). Dalam studi ini, *clustering* digunakan untuk mengelompokkan kecamatan-kecamatan di Kota Bandung berdasarkan metrik Beban Kasus Absolut pneumonia balita, sehingga dihasilkan kluster risiko Rendah, Sedang, dan Tinggi

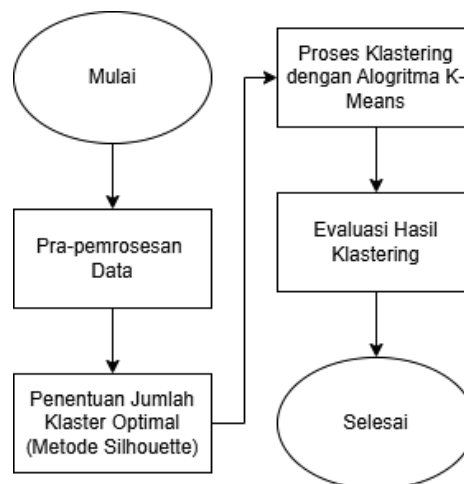
2.1.2 Sumber dan Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang secara spesifik diperoleh dari *Open Data Bandung*. Proses pengumpulan data sekunder ini dilakukan melalui akses terhadap sumber pelaporan reguler yang tersedia, seperti laporan-laporan yang berasal dari fasilitas kesehatan tingkat pertama, yaitu Pusat Kesehatan Masyarakat (Puskesmas), yang kemudian telah dikompilasi dan disiapkan oleh dinas terkait. *Dataset* yang dimanfaatkan ini mencakup informasi komprehensif mengenai jumlah balita penderita pneumonia di Kota Bandung pada tahun 2023. Informasi dalam *dataset* tersebut dirinci secara mendetail berdasarkan kecamatan, Unit Pelaksana Teknis (UPT) Puskesmas, dan jenis kelamin. Total data mentah yang berhasil dikumpulkan dan digunakan sebagai bahan analisis terdiri dari 160 baris data.

Variabel utama yang digunakan dalam pemodelan *clustering* ini adalah *Beban Kasus Absolut* (jumlah total kasus pneumonia balita) per kecamatan. Pemilihan variabel tunggal ini didasarkan pada tujuan penelitian yang bersifat taktis dan operasional, yaitu untuk mengidentifikasi wilayah dengan prioritas intervensi dan alokasi sumber daya segera berdasarkan jumlah kasus yang dikonfirmasi. Dengan memfokuskan pada beban kasus absolut, hasil pengelompokan secara langsung mencerminkan *immediate need* dan tingkat kerawanan yang paling mendesak di lapangan, yang menjadi acuan langsung untuk peningkatan kualitas pelayanan Puskesmas dan respons Dinas Kesehatan (Ramadhan et al., 2021). Variabel pemicu atau faktor risiko lain (seperti kepadatan penduduk, persentase BBLR, atau status gizi) tidak diikutsertakan sebagai input *clustering* karena tujuan penelitian ini adalah memetakan *outcome* (jumlah kasus) untuk keperluan manajemen kesehatan dan bukan untuk pemodelan epidemiologis multivariat.

2.2 Tahapan Penelitian

Proses penelitian ini disusun secara sistematis melalui empat tahapan utama. Setiap fase dirancang untuk memastikan integritas data dan validitas model, mulai dari persiapan data hingga evaluasi kualitas kluster. Rangkaian tahapan ini merupakan kerangka kerja yang esensial untuk mencapai luaran akhir, yaitu model klustering yang akurat dan bermakna. Tahapan-tahapan penelitian secara rinci disajikan pada Gambar 1, sebagai berikut:



Gambar 1. Alur Diagram Tahapan Penelitian

2.2.1 Pra-pemrosesan Data

Tahap pra-pemrosesan data merupakan langkah krusial dalam untuk memastikan kualitas data sebelum masuk ke tahap pemodelan. Berdasarkan alur kerja pada , tahap ini meliputi beberapa kegiatan utama:

- a. Seleksi Fitur (Feature Selection): Tujuan dari seleksi fitur adalah untuk memilih atribut-atribut dari *dataset* awal yang relevan dan esensial untuk mencapai tujuan penelitian.
 1. Atribut yang Dihilangkan: Kolom-kolom yang tidak memiliki kontribusi informatif untuk klustering, seperti *id*, *kode_provinsi*, *nama_provinsi*, *satuan*, dan *tahun*, dihilangkan. Penghapusan ini bertujuan untuk menyederhanakan analisis dan mengurangi dimensi data.
 2. Atribut Inti yang Dipertahankan: Penelitian ini memfokuskan pada variabel inti: *kemendagri_nama_kecamatan*, *jenis_kelamin*, dan *jumlah_balita*.
 3. Variabel Utama *Clustering*: Variabel utama yang akan digunakan dalam pemodelan klustering adalah *Beban Kasus Absolut* (jumlah total kasus pneumonia balita) per kecamatan. Pemilihan variabel tunggal ini bersifat taktis dan operasional, karena secara langsung mencerminkan tingkat kerawanan yang paling mendesak di lapangan, yang menjadi acuan segera untuk peningkatan kualitas pelayanan Puskesmas dan respons Dinas Kesehatan.
- b. Pembersihan Data (Data Cleaning): Pembersihan data dilakukan untuk meningkatkan konsistensi dan akurasi hasil analisis.
 1. Pemeriksaan *Missing Values*: Data diperiksa secara teliti untuk mengidentifikasi dan menangani *missing values* (data yang hilang).
 2. Pemeriksaan Data Duplikat: Pemeriksaan juga dilakukan untuk mengidentifikasi adanya data duplikat.
- c. Transformasi dan Agregasi Data: Tahap ini mengubah data mentah menjadi format yang sesuai dan siap untuk dimasukkan ke dalam algoritma K-Means.
 1. Data Awal: Data sekunder yang dikumpulkan dari Open Data Bandung pada tahun 2023 berisi 160 baris data. Data ini terperinci berdasarkan jenis kelamin dan Unit Pelaksana Teknis (UPT) Puskesmas.
 2. Proses Agregasi: Data yang terperinci tersebut kemudian diagregasi untuk mendapatkan total beban kasus absolut (jumlah total balita penderita pneumonia) per *kemendagri_nama_kecamatan*.



3. Dataset Akhir: Proses agregasi ini menghasilkan *dataset* akhir yang siap untuk dianalisis, di mana setiap baris data kini secara tunggal merepresentasikan satu kecamatan dengan total beban kasus pneumonia-nya. Inilah variabel Beban Kasus Absolut yang menjadi input pemodelan.

2.2.2 Penentuan Jumlah Kluster Optimal (Metode Silhouette)

Salah satu tantangan utama dalam algoritma K-Means adalah kebutuhan untuk menentukan jumlah kluster (k) secara apriori (Ariyanto, 2022b). Keterbatasan ini menuntut adanya mekanisme validasi internal yang objektif untuk memilih nilai k yang paling representatif. Untuk mengatasi hal ini, penelitian ini menggunakan Metode Silhouette untuk menemukan nilai k yang paling optimal. *Silhouette Score* berfungsi sebagai metrik yang kuat untuk mengukur seberapa baik sebuah objek data ditempatkan dalam klusternya dibandingkan dengan kluster lain. Skor ini dihitung secara individual untuk setiap sampel data dan memiliki rentang nilai terstandarisasi dari -1 hingga +1. Nilai *Silhouette Score* yang mendekati +1 merupakan indikasi kualitas kluster yang sangat baik, yang berarti objek data tersebut sangat cocok (kohesif) dengan klusternya sendiri dan memiliki pemisahan yang jauh (separasi tinggi) dari kluster-kluster tetangga.

Sebaliknya, nilai yang mendekati -1 menunjukkan bahwa objek tersebut mungkin salah ditempatkan (Ghazal et al., 2021b). Dalam proses implementasinya, algoritma K-Means dijalankan secara berulang dengan jumlah kluster yang bervariasi, mulai dari $k=2$ hingga $k=10$. Nilai k yang menghasilkan *Silhouette Score* tertinggi kemudian secara definitif dipilih sebagai jumlah kluster optimal yang akan digunakan dalam pemodelan akhir. Pemilihan ini memastikan bahwa kluster yang dihasilkan memiliki kohesi dan separasi yang maksimal, yang pada akhirnya memberikan dasar yang kuat untuk interpretasi kebijakan kesehatan.

2.2.3 Proses Klastering dengan Algoritma K-Means

Setelah jumlah kluster optimal ditentukan, proses pengelompokan dilakukan menggunakan algoritma K-Means. K-Means adalah salah satu algoritma partisi yang paling populer karena performanya yang baik pada dataset besar (Ghazal et al., 2021b). Algoritma ini bekerja secara iteratif untuk mempartisi data ke dalam k kluster yang telah ditentukan (Ariyanto, 2022b). Langkah-langkah dalam algoritma K-Means yang diterapkan adalah sebagai berikut:

- a. Inisialisasi Centroid: Tahap awal ini adalah krusial dalam algoritma partisi. Proses ini melibatkan pemilihan k titik data secara acak dari *dataset* untuk dijadikan sebagai pusat kluster (*centroid*) awal. Titik-titik yang dipilih ini akan berfungsi sebagai representasi pusat kluster yang menjadi acuan awal untuk proses iterasi
- b. Penugasan Objek: Setelah *centroid* diinisialisasi, langkah berikutnya adalah menghitung jarak antara setiap objek data (yaitu, setiap kecamatan) dengan masing-masing *centroid* yang ada. Metrik jarak yang digunakan dalam penelitian ini adalah Euclidean Distance, yang merupakan jarak garis lurus antara dua titik dalam ruang Euklides. (Ghazal et al., 2021b). Berdasarkan hasil perhitungan jarak tersebut, setiap objek data kemudian ditugaskan ke kluster dengan *centroid* terdekat dengannya
- c. Pembaruan Centroid: Setelah semua objek data berhasil ditugaskan ke kluster terdekat, posisi *centroid* untuk setiap kluster harus dihitung ulang. *Centroid* baru ditetapkan dengan mengambil nilai rata-rata (*mean*) dari semua objek data yang menjadi anggota kluster tersebut. Pembaruan ini bertujuan untuk meminimalkan varians internal kluster dan membuat *centroid* lebih akurat merepresentasikan anggota kluster yang baru.
- d. Iterasi: Langkah Penugasan Objek (B) dan Pembaruan Centroid (C) diulang secara terus-menerus. Proses iterasi berlanjut hingga kondisi Konvergensi tercapai, yaitu ketika tidak ada lagi objek data yang berpindah kluster, atau posisi *centroid* tidak lagi berubah secara signifikan. Pada titik ini, kluster dianggap telah stabil dan proses *clustering* selesai.

2.2.4 Evaluasi Hasil Klastering (Indeks Davies-Bouldin)

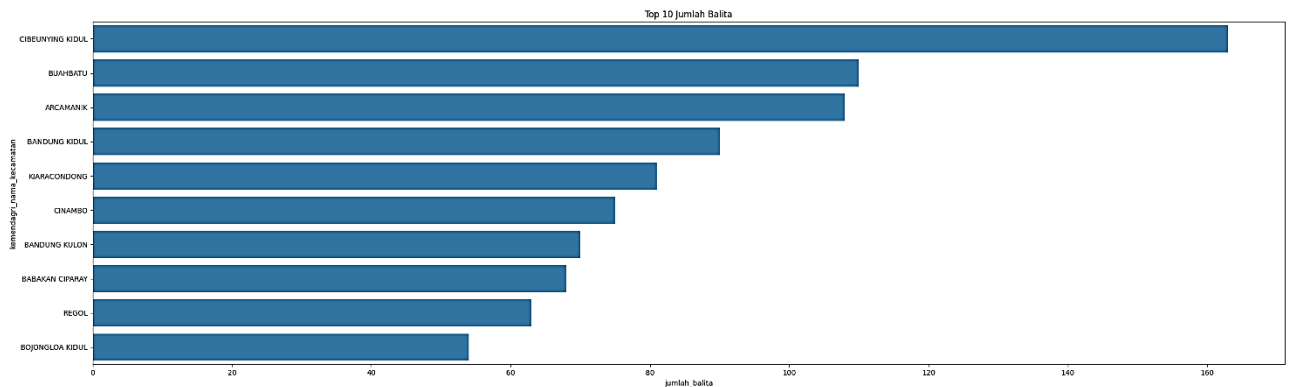
Untuk memvalidasi dan mengukur kualitas internal struktur kluster yang dihasilkan oleh implementasi algoritma K-Means, tahapan evaluasi merupakan proses yang krusial untuk memastikan bahwa hasil pengelompokan memiliki dasar ilmiah yang kuat dan dapat diandalkan. Evaluasi klastering adalah tahap penting dalam metode *unsupervised learning* yang bertujuan untuk mengukur kualitas dan validitas dari kluster yang telah terbentuk. Proses ini sangat penting untuk memverifikasi bahwa pengelompokan yang dihasilkan memiliki makna statistik dan dapat diinterpretasikan secara efektif dalam konteks masalah yang diteliti. Tujuannya adalah untuk memastikan kluster memiliki homogenitas internal yang tinggi (kohesi) dan heterogenitas eksternal yang tinggi (separasi).

Dalam penelitian ini, validasi dilakukan menggunakan metrik eksternal Indeks Davies-Bouldin (DBI). DBI diakui sebagai metrik evaluasi kluster yang efektif karena mampu mengukur rasio antara sebaran intra-kluster (jarak rata-rata antara setiap titik data dengan *centroid*-nya) dengan sebaran antar-kluster (jarak antara *centroid-centroid* kluster). Metrik DBI digunakan untuk mengevaluasi hasil *clustering*, di mana semakin kecil nilai yang diperoleh (non-negatif ≥ 0), maka semakin baik pula nilai kluster yang didapatkan (Purwaningsih & Nurelasari, 2023). Dengan kata lain, DBI secara simultan menilai kohesi internal kluster dan separasi antar-kluster. Nilai DBI merupakan kriteria yang jelas: nilai DBI yang lebih rendah menunjukkan kualitas klastering yang lebih baik. Hal ini tercapai karena nilai yang rendah mengindikasikan bahwa kluster yang terbentuk bersifat padat (kohesif) secara internal dan terpisah dengan baik (separasi tinggi) dari kluster lainnya (Ghazal et al., 2021). Oleh karena itu, hasil evaluasi ini memberikan keyakinan bahwa model klastering yang dibangun memiliki struktur yang valid dan bermakna, yang merupakan prasyarat penting sebelum hasil ini dapat diinterpretasikan untuk mendukung kebijakan kesehatan berbasis data.

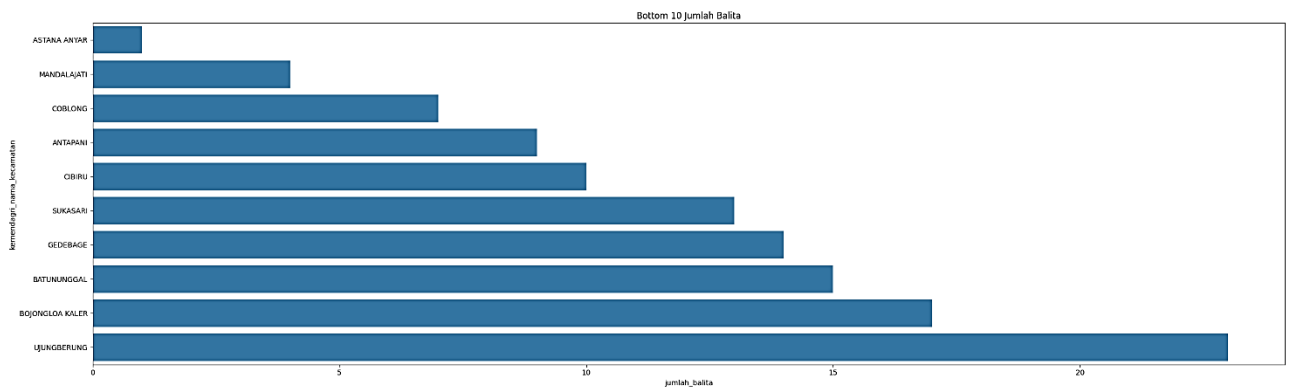
3. HASIL DAN PEMBAHASAN

3.1 Deskripsi Data Awal

Data yang digunakan dalam tahap analisis ini merupakan data jumlah penderita pneumonia pada kelompok balita tahun 2023. *Dataset* ini adalah hasil akhir dari serangkaian proses pra-pemrosesan data yang ketat, memastikan bahwa data tersebut telah bersih, teragregasi berdasarkan kecamatan, dan siap untuk dianalisis lebih lanjut. Sebagai langkah awal, dilakukan analisis eksplorasi (*Exploratory Data Analysis*) untuk mengidentifikasi distribusi kasus secara umum, khususnya untuk menentukan kecamatan dengan jumlah kasus tertinggi dan terendah. Hasil dari analisis ini menggarisbawahi adanya variasi yang signifikan dalam persebaran beban kasus pneumonia di seluruh wilayah Kota Bandung. Distribusi yang heterogen ini diperkuat oleh visualisasi data awal, seperti yang disajikan pada Gambar 2 (10 Kecamatan Kasus Tertinggi) dan Gambar 3 (10 Kecamatan Kasus Terendah).



Gambar 2. 10 Kecamatan dengan Jumlah Kasus Pneumonia Tertinggi



Gambar 3. 10 Kecamatan dengan Jumlah Kasus Pneumonia Terendah

Dari visualisasi pada Gambar 2 dan Gambar 3, terlihat bahwa Kecamatan Cibeunying Kidul dan Bandung Kulon memiliki beban kasus tertinggi, sementara Kecamatan Astana Anyar mencatatkan kasus terendah. Distribusi yang tidak merata ini memperkuat justifikasi untuk melakukan pengelompokan wilayah menggunakan *clustering* untuk mengidentifikasi pola persebaran.

3.2 Hasil Penentuan Klaster Optimal

Penentuan jumlah klaster (k) yang optimal merupakan tahap krusial untuk memastikan bahwa hasil klastering yang diperoleh memiliki validitas struktural yang tinggi. Dalam penelitian ini, penentuan nilai k dilakukan menggunakan Metode Silhouette¹. Evaluasi dilakukan secara iteratif dengan menguji nilai k yang bervariasi, yaitu dari $k=2$ hingga $k=10$. Hasil perhitungan *Silhouette Score* untuk setiap nilai k disajikan pada Gambar 4.

```
Silhouette Score dengan n: 2 adalah 0.6338500144119079
Silhouette Score dengan n: 3 adalah 0.6085321129533344
Silhouette Score dengan n: 4 adalah 0.5931027279867744
Silhouette Score dengan n: 5 adalah 0.582570781847149
Silhouette Score dengan n: 6 adalah 0.5923107599244701
Silhouette Score dengan n: 7 adalah 0.6036181479398575
Silhouette Score dengan n: 8 adalah 0.6084949702007756
Silhouette Score dengan n: 9 adalah 0.6126197460489813
Silhouette Score dengan n: 10 adalah 0.6060587566372929
Best cluster: 2, Silhouette Coefficient: 0.6338500144119079
```

Gambar 4. Hasil Perhitungan Silhouette Score

Berdasarkan hasil yang disajikan pada Gambar 4, secara teknis, nilai *Silhouette Score* tertinggi diperoleh saat $k = 2$, dengan skor sebesar 0.633. Meskipun demikian, untuk konteks analisis kebijakan kesehatan publik dan kebutuhan manajemen intervensi, pengelompokan menjadi hanya dua kategori (misalnya, risiko rendah dan risiko tinggi) seringkali dianggap kurang memberikan gradasi yang memadai untuk penentuan prioritas sumber daya. Oleh karena pertimbangan interpretasi dan implementasi kebijakan yang lebih praktis, penelitian ini memutuskan untuk tetap menggunakan $k = 3$, yang juga menunjukkan skor *Silhouette* yang memadai, yaitu 0.608. Keputusan untuk menggunakan tiga kluster ini sejalan dengan penelitian sebelumnya di bidang kesehatan yang juga membentuk tiga kategori untuk merepresentasikan tingkat risiko rendah, sedang, dan tinggi, sehingga hasilnya menjadi lebih mudah diinterpretasikan dan diimplementasikan oleh pemangku kebijakan (Vitalaya & Prasetyo, 2020).

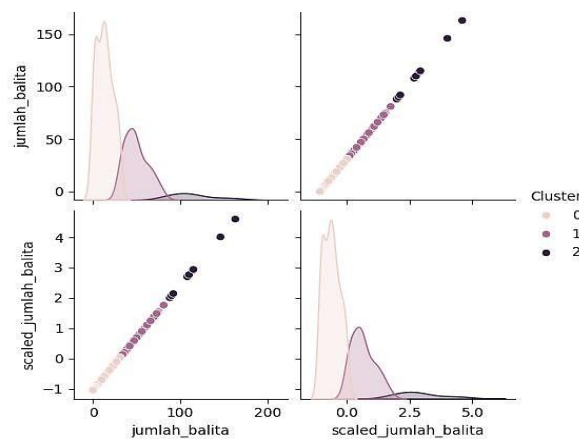
3.3 Hasil Penentuan Kluster Optimal

Dengan menetapkan jumlah kluster sebanyak 3, algoritma K-Means diimplementasikan pada dataset. Proses iteratif K-Means menghasilkan tiga kluster yang stabil dengan titik pusat (centroid) yang merepresentasikan nilai rata-rata jumlah kasus untuk setiap kluster. Hasil centroid akhir disajikan pada Tabel 1.

Tabel 1. Nilai Centroid Akhir untuk Setiap Kluster

Kluster	Kategori	Nilai Centroid
0	Rendah	12.22
1	Sedang	47.13
2	Tinggi	110.70

Berdasarkan nilai centroid tersebut, setiap kecamatan di Kota Bandung dikelompokkan ke dalam salah satu dari tiga kluster. Visualisasi persebaran data pada setiap kluster dapat dilihat pada Gambar 5, yang menunjukkan adanya pemisahan yang jelas antara ketiga kelompok.



Gambar 5. Pairplot Hasil Klustering

Distribusi anggota untuk setiap kluster adalah sebagai berikut:

- Kluster 0 (Rendah):** Terdiri dari 12 kecamatan dengan rata-rata kasus sekitar 12 balita.
Anggota: Antapani, Astana Anyar, Bandung Wetan, Batununggal, Bojongloa Kaler, Cibeunying Kaler, Cicendo, Gedebage, Lengkong, Panyileukan, Regol, Sumur Bandung.
- Kluster 1 (Sedang):** Terdiri dari 12 kecamatan dengan rata-rata kasus sekitar 47 balita.
Anggota: Andir, Babakan Ciparay, Bojongloa Kidul, Cibiru, Cidadap, Cinambo, Coblong, Mandalajati, Rancasari, Sukajadi, Sukasari, Ujungberung.
- Kluster 2 (Tinggi):** Terdiri dari 6 kecamatan dengan rata-rata kasus sekitar 111 balita.
Anggota: Arcamanik, Bandung Kidul, Bandung Kulon, Buahbatu, Cibeunying Kidul, Kiaracondong.

3.4 Evaluasi dan Validasi Hasil

Validasi kualitas hasil pengelompokan (clustering) merupakan tahapan krusial untuk memastikan bahwa struktur kluster yang dihasilkan bermakna dan dapat diandalkan. Dalam penelitian ini, evaluasi dilakukan menggunakan Davies-Bouldin Index (DBI), di mana hasil perhitungan menunjukkan nilai DBI sebesar 0.4877 (atau 0,487). Nilai DBI adalah metrik yang mengukur rasio antara sebaran intra-kluster (kepadatan di dalam kluster) dengan jarak antar-kluster (pemisahan antar kluster). Nilai DBI yang lebih rendah mengindikasikan kualitas klustering yang lebih baik, karena ini berarti kluster yang terbentuk padat (kohesif) dan terpisah dengan baik (separasi tinggi) dari kluster lainnya. Oleh karena itu, perolehan nilai DBI 0.4877 menunjukkan bahwa model pengelompokan yang membagi 30 kecamatan di Kota Bandung menjadi tiga kluster risiko (Rendah, Sedang, Tinggi) memiliki struktur yang cukup valid dan dapat diandalkan.



$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left(\frac{s_i + s_j}{d_{ij}} \right) \quad (1)$$

Di mana k adalah jumlah kluster, s_i dan s_j adalah ukuran dispersi (scatter) kluster i dan j (yang merepresentasikan kepadatan internal), dan d_{ij} adalah jarak antara pusat (centroid) kluster i dan kluster j (yang merepresentasikan pemisahan antar kluster). Karena nilai 0.4877 yang diperoleh mendekati nol dan lebih rendah dibandingkan dengan nilai DBI dari beberapa studi sebelumnya, hal ini memberikan keyakinan bahwa model yang dibangun berhasil memisahkan kelompok risiko pneumonia secara jelas dan terstruktur.

3.5 Pembahasan

Hasil penelitian ini berhasil memetakan 30 kecamatan di Kota Bandung ke dalam tiga kluster kerawanan pneumonia pada balita. Kluster 2 (Tinggi), yang terdiri dari Arcamanik, Bandung Kidul, Bandung Kulon, Buahbatu, Cibeunying Kidul, dan Kiaracondong, secara jelas teridentifikasi sebagai *hotspot* atau wilayah prioritas utama yang memerlukan perhatian khusus dari Dinas Kesehatan. Temuan ini memberikan bukti empiris yang kuat untuk pengambilan keputusan berbasis data, yang merupakan tujuan utama dari pemanfaatan teknologi dalam analisis data kesehatan (Abdillah et al., 2023).

Identifikasi wilayah-wilayah dengan risiko tinggi ini sangat krusial, karena berbagai penelitian epidemiologi menunjukkan bahwa kejadian pneumonia pada balita sangat dipengaruhi oleh faktor-faktor lingkungan dan sosial yang seringkali terkonsentrasi secara geografis. Faktor risiko seperti kepadatan hunian yang tinggi, ventilasi rumah yang buruk, status gizi, dan kelengkapan imunisasi terbukti memiliki hubungan signifikan dengan kejadian pneumonia (Hariyanto, 2020; Husna M et al., 2022). Meskipun data demografis tersebut tidak dimasukkan sebagai variabel dalam model *clustering* ini, hasil pengelompokan dapat menjadi dasar untuk penelitian lanjutan yang mengkaji korelasi antara kluster risiko dengan faktor-faktor tersebut di tingkat kecamatan (Hidayani & Km, 2020).

Implikasi praktis dari temuan ini adalah Dinas Kesehatan Kota Bandung dapat merancang intervensi yang lebih terfokus. Untuk kluster tinggi, program seperti penyuluhan intensif mengenai rumah sehat, peningkatan cakupan imunisasi, dan surveilans aktif oleh kader kesehatan dapat diprioritaskan (Hidayani & Km, 2020). Temuan ini juga menggarisbawahi pentingnya analisis data kesehatan yang tidak hanya berhenti pada pelaporan, tetapi juga dilanjutkan ke tahap analisis prediktif untuk mendukung perencanaan strategis (Akbar et al., 2025).

Penelitian ini memiliki keterbatasan, yaitu model klustering hanya didasarkan pada satu variabel utama (jumlah kasus). Penelitian di masa depan dapat dikembangkan dengan memasukkan variabel-variabel lain yang relevan, seperti jumlah balita per kecamatan (untuk menghitung *incidence rate*), tingkat kepadatan penduduk, dan data status gizi atau kelengkapan imunisasi, agar dapat menghasilkan model pengelompokan yang lebih komprehensif.

4. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma K-Means untuk mengelompokkan kasus pneumonia pada balita di Kota Bandung berdasarkan Beban Kasus Absolut tahun 2023. Hasil klustering membagi 30 kecamatan ke dalam tiga kategori risiko: rendah, sedang, dan tinggi. Kluster risiko tinggi (*hotspot*) secara spesifik diidentifikasi meliputi Arcamanik, Bandung Kidul, Bandung Kulon, Buahbatu, Cibeunying Kidul, dan Kiaracondong, yang membutuhkan prioritas intervensi kesehatan segera. Validitas model dikonfirmasi melalui Evaluasi Indeks Davies-Bouldin (DBI), yang menghasilkan nilai 0,487, menunjukkan kualitas kluster yang baik dengan pemisahan yang jelas antar kelompok. Temuan ini memberikan dasar ilmiah bagi Dinas Kesehatan Kota Bandung untuk merancang strategi pencegahan dan penanggulangan pneumonia yang lebih terarah, seperti peningkatan cakupan imunisasi, penyuluhan lingkungan rumah sehat, dan surveilans aktif di wilayah risiko tinggi. Dengan demikian, teknik *clustering* terbukti dapat digunakan sebagai alat bantu analisis spasial, dan jurnal ini diharapkan dapat menjadi jurnal pendukung serta acuan untuk diberlakukannya riset kesehatan dan epidemiologi lebih lanjut, khususnya dalam mengkaji korelasi antara kluster risiko dengan faktor-faktor pemicu pneumonia pada balita.

REFERENCES

- Abdillah, N., Susilo, H., & Ihksan, M. (2023). Sosialisasi Pemanfaatan Teknologi Data Mining Untuk Analisis Data Kesehatan Di Klinik Amanah. *Jurnal Abdimas Saintika*, 5(1), 181–186. <https://doi.org/http://dx.doi.org/10.30633/jas.v5i1.1940>
- Abimayu, A. T., & Rahmawati, N. D. (2023). Analisis Faktor Risiko Kejadian Stunted, Underweight, dan Wasted Pada Balita di Wilayah Kerja Puskesmas Rangkapan Jaya, Kota Depok, Jawa Barat Tahun 2022. *Jurnal Biostatistik, Kependudukan, Dan Informatika Kesehatan*, 3(2)(2), 88–101. <https://doi.org/10.51181/BIKFOKES.V3I2.6820>
- Akbar, I., Supriadi, F., & Junaedi, D. I. (2025a). Pemanfaatan Machine Learning Di Bidang Kesehatan. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(1), 1744–1749. <https://doi.org/10.36040/JATI.V9I1.12663>
- Ariyanto, D. (2022). Data Mining Menggunakan Algoritma K-Means untuk Klasifikasi Penyakit Infeksi Saluran Pernafasan Akut. *Jurnal Sistim Informasi Dan Teknologi*, 4(1), 13–18. <https://doi.org/10.37034/jsisfotek.v4i1.117>



- Chotimah S. (2022). Implementasi Sistem Informasi Kesehatan di Fasilitas Pelayanan Kesehatan Indonesia: Literature Review. *JURMIK (Jurnal RekamMedis Dan Manajemen Informasi Kesehatan)*, 2(1), 8–13. <https://doi.org/10.53416/JURMIK.V2I1.67>
- Ghazal, T. M., Hussain, M. Z., Said, R. A., Nadeem, A., Hasan, M. K., Ahmad, M., Khan, M. A., & Naseem, M. T. (2021). Performances of k-means clustering algorithm with different distance metrics. *Intelligent Automation and Soft Computing*, 30(2), 735–742. <https://doi.org/10.32604/iasc.2021.019067>
- Hariyanto, H. (2020). Kejadian Pneumonia pada Anak Usia 12-59 Bulan. *HIGEIA (Journal of Public Health Research and Development)*, 4(Special 3), 549–560. <https://doi.org/https://doi.org/10.15294/higeia.v4iSpecial%203.40524>
- Hidayani, W. R., & Km, S. (2020). *Pneumonia : Epidemiologi, Faktor Risiko Pada Balita* (1st ed., Vol. 1). Pena Persada.
- Husna M, Pertiwi F, & Nasution A. (2022). Faktor-Faktor Yang Berhubungan Dengan Kejadian Pneumonia Pada Balita Di Puskesmas Semplak Kota Bogor 2020. *PROMOTOR*, 5(3), 273–280. <https://doi.org/10.32832/PRO.V5I3.6168>
- Purwaningsih, E., & Nurelasari, E. (2023). Implementasi Metode K-Means Clustering Dengan Davies Bouldin Index Pada Analisis Faktor Penyebab Perceraian. *Information Management For Educators And Professionals : Journal of Information Management*, 7(2), 143. <https://doi.org/10.51211/IMBI.V7I2.2307>
- Rakuasa, H., Tambunan, M. P., & Tambunan, R. P. (2021). Analisis Sebaran Spasial Tingkat Kejadian Kasus Covid-19 Dengan Metode Kernel Density di Kota Ambon. *Jurnal Geografi : Media Informasi Pengembangan Dan Profesi Kegeografian*, 18(2), 76–82. <https://doi.org/10.15294/jg.v18i2.28234>
- Ramadhan, F., Muhafidin, D., & Miradhia, D. (2021). Kualitas Pelayanan Kesehatan Puskesmas Ibum Kabupaten Bandung. *JANE (Jurnal Administrasi Negara)*, 12(2), 58–63. <https://jurnal.unpad.ac.id/jane/article/view/28684>
- Rosiana, P. S., Apriliansyah Mohsa, A., Fadila, M. A., Jaman, J. H., Karawang, U. S., Ronggo Waluyo, J. H., & Timur, T. (2023). Visualisasi Data Tindak Kejahatan Berdasarkan Jenis Kriminalitas Di Kabupaten Karawang Dengan Menggunakan Algoritma Clustering K-Means. *Jurnal Informatika Dan Teknik Elektro Terapan*, 11(3s1), 822–828. <https://doi.org/https://doi.org/10.23960/jitet.v11i3s1.3347>
- Vitalaya, N., & Prasetio, R. T. (2020a). Implementasi Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Pneumonia Pada Balita Di Kota Bandung. *EProsiding Sistem Informasi (POTENSI)*, 1(1), 108–116. <http://eprosiding.ars.ac.id/index.php/psi>