

Comparative Analysis of DBSCAN, OPTICS, and Agglomerative Clustering Methods for Identifying Disease Distribution Patterns in Banjarnegara Community Health Centers

Dillyana Tugan Setiyawan^{*1}, Berlilana², Azhari Shouni Barkah³

¹ Information System, Computer Science Faculty, Universitas Amikom Purwokerto, Indonesia

^{2,3} Magister of Computer Science, Computer Science Faculty, Universitas Amikom Purwokerto, Indonesia

Email: ¹polibara@gmail.com

Received : Apr 9, 2025; Revised : May 14, 2025; Accepted : May 15, 2025; Published : Jun 10, 2025

Abstract

The variation in disease distribution patterns across community health centers in Banjarnegara Regency necessitates a precise segmentation analysis to support effective allocation of healthcare resources. This study aims to compare the effectiveness of three clustering methods DBSCAN, OPTICS, and Agglomerative Clustering in grouping Puskesmas based on the type and number of diseases they manage. The evaluation methods used include the Silhouette Score and the Davies-Bouldin Index, which assess the quality of the clustering results. The analysis indicates that Agglomerative Clustering produces the most stable cluster structures, reflected in its highest Silhouette Score, compared to DBSCAN and OPTICS, which tend to yield more noise and less optimal clustering quality. These findings suggest that hierarchical clustering approaches are more effective in the context of healthcare service distribution data at the primary care level. The results of this study are expected to serve as a foundation for the formulation of data-driven and region-based health policies, particularly in designing more targeted interventions and optimizing the distribution of healthcare services.

Keywords: DBSCAN, OPTICS, Agglomerative Clustering, Silhouette Score, Disease Distribution

This work is an open access article and licensed under a Creative Commons Attribution-Non Commercial 4.0 International License



1. PENDAHULUAN

Kesehatan masyarakat merupakan aspek fundamental dalam menunjang kualitas hidup dan pembangunan suatu wilayah. Ketimpangan dalam distribusi penyakit yang tidak terdeteksi secara tepat dapat berdampak besar terhadap peningkatan angka kesakitan, penurunan produktivitas penduduk, dan pembebanan layanan fasilitas kesehatan, khususnya di daerah yang memiliki keterbatasan akses dan sumber daya [1],[2]. Kabupaten Banjarnegara merupakan salah satu wilayah dengan jumlah penduduk lebih dari satu juta jiwa dengan 35 Puskesmas yang tersebar di berbagai kecamatan, menjadi wilayah yang memerlukan pendekatan sistematis untuk menganalisis penyebaran penyakit secara menyeluruh [3].

Data kunjungan pasien ke Puskesmas, jika dianalisis dengan pendekatan yang tepat, dapat menjadi sumber informasi strategis dalam melihat tren penyakit, pola sebaran geografis, dan beban layanan di masing-masing fasilitas kesehatan [4], [5]. Metode clustering atau pengelompokan data menjadi solusi yang potensial, dalam mengelompokkan wilayah berdasarkan kemiripan kasus penyakit yang ditangani. Metode ini telah banyak digunakan dalam studi kesehatan untuk memahami segmentasi wilayah, perilaku pasien, hingga perencanaan distribusi sumber daya [6], [7],[8].

Beberapa algoritma clustering yang banyak diaplikasikan dalam bidang epidemiologi dan pelayanan kesehatan antara lain DBSCAN (*Density Based Spatial Clustering of Applications with Noise*), OPTICS (*Ordering Points To Identify the Clustering Structure*), dan Agglomerative Clustering. DBSCAN dan OPTICS merupakan metode berbasis kepadatan yang efektif dalam mendeteksi cluster dan outlier pada data yang tidak seragam distribusinya [9], [10],[11] DBSCAN bekerja berdasarkan dua parameter utama, epsilon dan MinPts, yang menentukan jarak antar titik dan jumlah minimum dalam membentuk cluster akan tetapi performanya sangat dipengaruhi oleh pemilihan parameter [12], [13]. OPTICS hadir sebagai penyempurna dari DBSCAN, karena mampu mengenali cluster dengan tingkat kepadatan yang bervariasi tanpa perlu menetapkan nilai epsilon secara eksplisit [14].

Agglomerative Clustering merupakan metode hirarkis yang menyusun cluster berdasarkan kedekatan data secara bertahap. Keunggulannya terletak pada stabilitas hasil pengelompokan dan kemampuannya dalam divisualisasikan melalui dendrogram, yang memudahkan dalam menentukan jumlah cluster optimal [15], [16], [17]. Agglomerative Clustering tidak bergantung pada parameter kepadatan, sehingga sangat sesuai untuk data kesehatan dengan variasi antar wilayah [18], [19].

Penelitian dengan menggabungkan ketiga metode clustering DBSCAN, OPTICS, dan Agglomerative Clustering untuk menganalisis pola penyebaran penyakit di tingkat layanan primer, khususnya di Puskesmas, masih sangat jarang ditemukan di Indonesia. Sebagian besar studi sebelumnya cenderung hanya menggunakan satu metode saja, seperti k-means atau fuzzy c-means, dalam mengevaluasi beban penyakit tidak menular [20], [21]. Penerapan beberapa metode dalam suatu kasus dapat memberikan gambaran yang lebih menyeluruh mengenai variasi distribusi penyakit antar wilayah, serta membantu mengidentifikasi area yang memiliki kebutuhan layanan kesehatan lebih tinggi dan mendesak.

Berdasarkan hal tersebut, penelitian ini bertujuan untuk membandingkan efektivitas metode DBSCAN, OPTICS, dan Agglomerative Clustering dalam mengelompokkan berdasarkan karakteristik penyakit di Puskesmas Kabupaten Banjarnegara. Diharapkan dengan pendekatan ini dapat ditemukan metode yang paling optimal untuk menganalisis pola penyebaran penyakit, serta memberikan rekomendasi bagi pengambil kebijakan dalam merumuskan strategi distribusi tenaga kesehatan, alokasi sumber daya, dan intervensi berbasis wilayah secara lebih tepat sasaran [22], [23],[24].

Penelitian ini membandingkan tiga metode clustering pada layanan kesehatan di Puskesmas dalam satu wilayah Kabupaten. Pendekatan ini tidak hanya mengisi celah riset yang ada, tetapi juga memberikan kontribusi nyata dalam pengembangan sistem informasi kesehatan berbasis data.

2. METODE PENELITIAN

2.1. Data

Penelitian ini menggunakan data dari 35 Puskesmas yang diambil dari dinas kesehatan Kabupaten Banjarnegara. Populasi berupa pasien yang melakukan pemeriksaan di seluruh Puskesmas di wilayah Kabupaten Banjarnegara sejumlah 467.440. Sampel penelitian ini menggunakan seluruh pasien yang berobat di 35 puskesmas.

2.2. Parameter Tuning pada DBSCAN

Proses tuning parameter DBSCAN dilakukan secara sistematis untuk memperoleh kombinasi terbaik antara nilai epsilon (ϵ) dan MinPts (Minimum Points) yang menghasilkan kualitas clustering optimal. Pemilihan parameter ini sangat penting karena berpengaruh langsung terhadap hasil pengelompokan serta kemampuan model dalam mendeteksi noise dan membentuk cluster yang bermakna [15]. Berbeda dengan metode clustering lainnya, DBSCAN tidak memerlukan jumlah cluster yang ditentukan sebelumnya, tetapi bergantung pada dua parameter utama, yaitu Epsilon (eps) yang

menentukan jarak maksimum antara titik yang masih dianggap dalam satu cluster, serta MinPts yang menentukan jumlah minimum titik dalam radius eps agar membentuk cluster [5]. Keunggulan utama DBSCAN adalah kemampuannya dalam mendeteksi outlier (*noise*), yang sangat berguna dalam analisis pola penyebaran penyakit di Puskesmas Kabupaten Banjarnegara. Metode ini mengidentifikasi wilayah dengan tingkat kepadatan penyakit tinggi, sekaligus mengenali Puskesmas yang memiliki pola penyakit berbeda dari mayoritas. Berikut adalah langkah-langkah dalam menerapkan DBSCAN [18].

1. Rentang nilai Epsilon (ϵ):

Nilai ϵ ditentukan berdasarkan k-distance graph menggunakan metode elbow/knee point.

Rentang nilai yang diuji adalah dari 1.5 hingga 3.5 dengan interval kenaikan 0.5, yakni: $\epsilon \in \{1.5, 2.0, 2.5, 3.0, 3.5\}$

2. Rentang nilai MinPts:

Nilai MinPts yang diuji bervariasi dari 3 hingga 6 berdasarkan referensi literatur yang menyatakan bahwa MinPts optimal umumnya berkisar antara $2 \times \text{dimensi data}$ hingga $2 \times \text{dimensi} + 1$.

3. Alasan pemilihan rentang:

Nilai-nilai tersebut dipilih karena mewakili kombinasi yang umum digunakan dalam data kesehatan masyarakat dengan struktur yang tidak terlalu kompleks tetapi tetap memiliki kemungkinan noise tinggi.

4. Hasil terbaik:

Berdasarkan hasil uji Silhouette Score, kombinasi parameter terbaik ditemukan pada $\epsilon = 3.5$ dan MinPts = 3, yang menghasilkan nilai Silhouette Score tertinggi (0.381).

5. Justifikasi Metrik Evaluasi

Evaluasi kualitas hasil clustering, digunakan dua metrik evaluasi yang saling melengkapi, yaitu Silhouette Score dan Davies Bouldin Index (DBI).

a. Silhouette Score

Silhouette Score mengukur seberapa mirip suatu objek dengan cluster-nya sendiri dibandingkan dengan cluster lain. Nilai berkisar antara -1 hingga 1.

b. Davies Bouldin Index (DBI)

Davies-Bouldin Index mengukur rasio antara jarak dalam cluster (*intra cluster*) dan jarak antar cluster (*inter cluster*). Nilai yang lebih kecil menunjukkan kualitas cluster yang lebih baik. DBI efektif dalam menilai seberapa rapat titik dalam satu cluster dan seberapa jauh cluster tersebut dari cluster lainnya. DBI berguna untuk mengevaluasi model clustering berbasis kepadatan yang cenderung menghasilkan bentuk cluster tidak beraturan seperti DBSCAN dan OPTICS.

2.3. Parameter Tuning dan Evaluasi OPTICS

OPTICS (*Ordering Points To Identify the Clustering Structure*) merupakan algoritma clustering berbasis kepadatan yang dikembangkan untuk mengatasi kelemahan DBSCAN dalam menentukan nilai parameter ϵ (Eps). OPTICS memungkinkan pengelompokan data dengan tingkat kepadatan yang bervariasi tanpa perlu menentukan nilai ϵ secara eksplisit [14]

Konsep utama OPTICS adalah membangun urutan (*ordering*) dari titik-titik berdasarkan nilai core distance dan reachability distance. OPTICS kemudian menggunakan informasi dari urutan tersebut untuk membentuk cluster dengan kepadatan yang berbeda [17]. Langkah-langkah Algoritma OPTICS [8]

1. Parameter kunci

a. MinPts: Menentukan jumlah minimum tetangga untuk mengidentifikasi core point. Dalam penelitian ini diuji dalam rentang 3 hingga 6.

- b. ξ (ξ): Parameter tambahan dalam metode cluster extraction OPTICS, yang merepresentasikan penurunan jarak yang signifikan di reachability plot. Pada penelitian ini nilai ξ ditetapkan pada 0.05 mengikuti standar literatur.
2. Alasan pemilihan OPTICS:
OPTICS dipilih karena mampu mengidentifikasi cluster dengan kepadatan yang bervariasi, yang umum dijumpai dalam data kesehatan masyarakat seperti distribusi penyakit antar Puskesmas.
3. Evaluasi Kualitas Kluster:
OPTICS menghasilkan urutan titik dan bukan langsung kluster eksplisit, hasil kluster diekstrak menggunakan pendekatan berbasis reachability distance dan kemudian dievaluasi menggunakan Silhouette Score dan Davies Bouldin Index. Dua metrik ini digunakan untuk menilai kualitas hasil kluster final yang dihasilkan oleh OPTICS setelah proses ekstraksi kluster dilakukan.

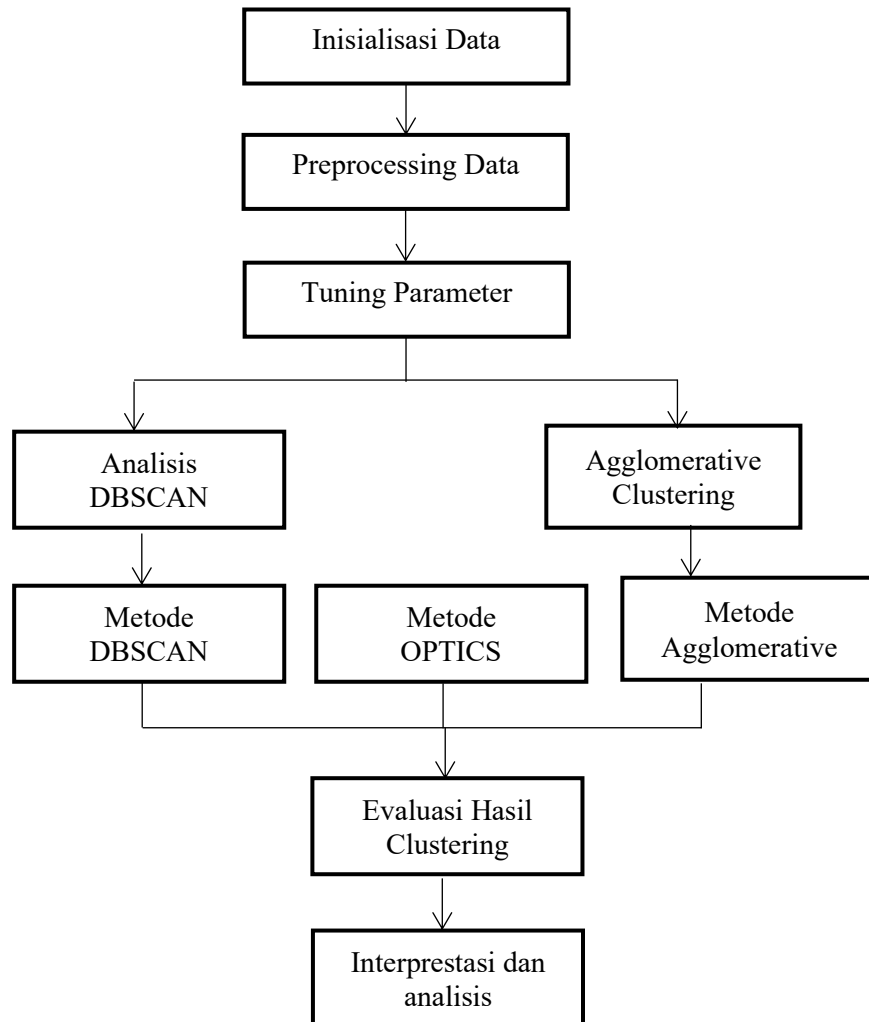
2.4. Parameter Tuning dan Evaluasi Agglomerative Clustering

Agglomerative Clustering merupakan salah satu metode Hierarchical Clustering yang bekerja dengan pendekatan bottom-up, di mana setiap data awalnya dianggap sebagai cluster tersendiri, kemudian secara bertahap digabungkan dengan cluster lain yang memiliki kesamaan hingga terbentuk jumlah cluster yang optimal [20].

Agglomerative Clustering tidak memerlukan parameter Epsilon (ϵ) dan MinPts, sehingga lebih stabil dalam membentuk kelompok data yang jelas, penerapan dendrogram memungkinkan penentuan jumlah cluster terbaik berdasarkan hubungan hierarkis antar data [4]. Berikut adalah langkah-langkah dalam penerapan Agglomerative Clustering [19]:

1. Parameter kunci
 - a. Linkage method: Menentukan bagaimana jarak antar kluster dihitung. Tiga metode yang diuji dalam penelitian ini adalah:
 - 1) Ward linkage: Meminimalkan variansi total dalam kluster (paling cocok untuk data spasial).
 - 2) Average linkage: Menggunakan rata-rata jarak antar anggota dua kluster.
 - 3) Complete linkage: Menggunakan jarak terjauh antar titik dari dua kluster.
 - b. Jumlah kluster ($n_clusters$): Rentang nilai yang diuji adalah 2 hingga 6.
2. Alasan pemilihan metode ini:
Agglomerative Clustering sangat sesuai untuk eksplorasi struktur hirarkis dari data, terutama ketika data memiliki pola bertingkat atau bertahap seperti penyebaran penyakit antar wilayah.
3. Evaluasi Kualitas Kluster:
Silhouette Score digunakan untuk menilai kepadatan dan keterpisahan kluster. Davies Bouldin Index digunakan untuk mengevaluasi rasio antara jarak intra dan antar kluster.

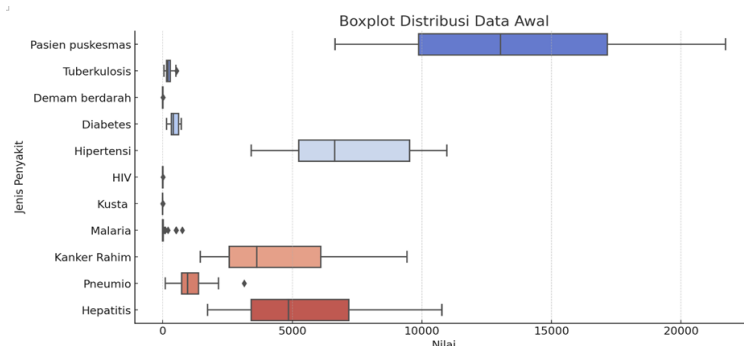
2.5. Diagram Penelitian



Gambar 1. Diagram penelitian

3. HASIL DAN PEMBAHASAN

3.1. Tampilan Data



Gambar 2. Visualisasi Boxplot sebaran penyakit

Gambar 2 menunjukkan Boxplot distribusi data awal berupa sebaran jumlah pasien di Puskesmas Kabupaten Banjarnegara berdasarkan berbagai jenis penyakit. Visualisasi boxplot distribusi data awal memperlihatkan pola distribusi yang bervariasi antar jenis penyakit. Penyakit seperti Hipertensi dan Kanker Rahim menunjukkan rentang distribusi yang luas, yang menunjukkan ketidakmerataan kasus pasien di berbagai Puskesmas. Data menggambarkan terdapat outlier pada beberapa penyakit, seperti Tuberkulosis dan Diabetes menunjukkan terdapat Puskesmas dengan prevalensi penyakit yang sangat tinggi atau rendah. Penelitian sebelumnya menggarisbawahi pentingnya menerapkan deteksi outlier untuk mengidentifikasi kasus - kasus yang tidak biasa, sehingga dapat mempengaruhi tentang distribusi kasus penyakit dalam populasi [13]. Keberadaan outlier dapat memunculkan pola yang tidak biasa yang perlu diperhatikan untuk analisis lebih lanjut yang meliputi keterkaitan antara penyakit dan faktor risiko yang beragam, [12].

Hasil sebaran data menunjukkan bahwa penyakit dengan distribusi lebih kecil seperti Kusta dan Demam Berdarah cenderung lebih merata di seluruh Puskesmas, yang dapat diasosiasikan dengan angka kejadian yang lebih rendah [11]. Distribusi penyakit tidak hanya ditentukan oleh tingkat prevalensinya, tetapi juga dipengaruhi oleh faktor-faktor demografis serta struktur dan kapasitas organisasi pelayanan kesehatan setempat, yang memengaruhi pola pencarian dan akses pasien terhadap layanan medis [7]. Teknik clustering mampu memberikan wawasan yang lebih mendalam terkait pola distribusi penyakit yang bervariasi di setiap wilayah [25]. Secara keseluruhan, analisis boxplot berperan penting dalam membangun pemahaman awal mengenai bagaimana variasi tersebut dapat memengaruhi perencanaan dan pengelolaan layanan kesehatan di tingkat Puskesmas. Temuan ini membuka peluang bagi penerapan intervensi yang lebih tepat sasaran serta pengelolaan sumber daya kesehatan yang lebih efisien, dengan pendekatan yang berfokus pada pencegahan dan pengendalian penyakit [26].

3.2. Hasil DBSCAN, OPTICS, dan Agglomerative

Hasil kualitas klusterisasi dari metode DBSCAN, OPTICS, dan Agglomerative Clustering, dengan masing-masing kombinasi Parameter 1 dan Parameter 2 yang mewakili konfigurasi spesifik algoritma, Epsilon dan MinPts untuk metode *density based*. Penilaian dilakukan menggunakan dua metrik utama, yaitu Silhouette Score dan Davies Bouldin Index. Perbandingan ini memberikan dasar untuk menilai efektivitas relatif dari setiap metode dalam mengelompokkan data penyakit di wilayah penelitian, hasil sebagai berikut :

Tabel 1. Gabungan Nilai Indeks Silhouette dan Davies Bouldin Index

Metode	Parameter 1	Parameter 2	Silhouette Score	Davies-Bouldin Index
DBSCAN	$\hat{\mu} = 2.0$	MinPts = 3	0.281	0.812
DBSCAN	$\hat{\mu} = 2.5$	MinPts = 3	0.298	0.776
DBSCAN	$\hat{\mu} = 3.0$	MinPts = 3	0.348	0.744
DBSCAN	$\hat{\mu} = 3.5$	MinPts = 3	0.381	0.698
OPTICS	MinPts = 3	$\hat{\epsilon} = 0.05$	0.291	0.812
OPTICS	MinPts = 4	$\hat{\epsilon} = 0.05$	0.305	0.795
OPTICS	MinPts = 5	$\hat{\epsilon} = 0.05$	0.318	0.781
OPTICS	MinPts = 6	$\hat{\epsilon} = 0.05$	0.327	0.769
Agglomerative	Linkage = Ward	Cluster = 3	0.362	0.642
Agglomerative	Linkage = Average	Cluster = 3	0.341	0.683
Agglomerative	Linkage = Complete	Cluster = 3	0.328	0.701

Tabel 1 menunjukkan bahwa metode DBSCAN dengan kombinasi parameter $\epsilon = 3.5$ dan MinPts = 3 menghasilkan nilai Silhouette Score tertinggi, yaitu 0,381, serta nilai Davies Bouldin Index (DBI)

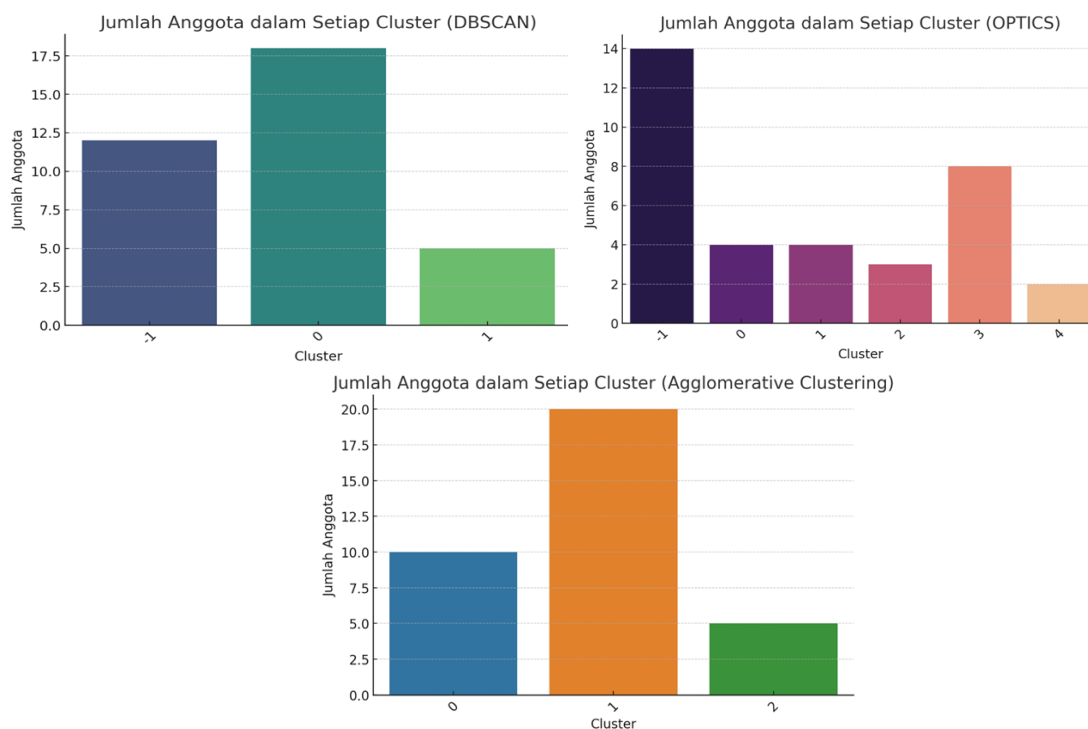
sebesar 0,698. Nilai ini mengindikasikan struktur kluster yang relatif baik dalam hal pemisahan antar kelompok data. DBSCAN juga menghasilkan sejumlah besar data yang dikategorikan sebagai noise, yang mencerminkan keterbatasan dalam mengelompokkan data secara konsisten. Hasil dari metode DBSCAN sangat bergantung pada pemilihan parameter, khususnya nilai epsilon dan MinPts, yang berperan krusial dalam menentukan kualitas hasil klusterisasi.

Metode OPTICS dengan konfigurasi parameter MinPts = 6 dan xi = 0.05 menghasilkan nilai Silhouette Score tertinggi sebesar 0,327 dan Davies-Bouldin Index sebesar 0,769. Hasil ini menunjukkan bahwa kluster yang terbentuk oleh OPTICS memiliki tingkat separabilitas yang rendah, dengan batas antar kluster yang cenderung tumpang tindih. Keunggulan utama OPTICS terletak pada kemampuannya menangani data dengan kepadatan yang bervariasi serta mendeteksi struktur kluster yang kompleks. Pada penelitian ini performa OPTICS masih menunjukkan stabilitas yang lebih rendah.

Metode Agglomerative Clustering dengan pendekatan linkage Ward dan jumlah kluster sebanyak tiga menghasilkan Silhouette Score sebesar 0,362 dan Davies Bouldin Index sebesar 0,642. Nilai ini mencerminkan keseimbangan yang optimal antara kepadatan internal dan pemisahan antar kluster, serta menunjukkan struktur kluster yang lebih stabil dan terdefinisi dengan baik. Segmentasi yang dihasilkan tampak jelas, dengan minimnya data yang tidak terkelompok (*noise*), sehingga mengindikasikan kualitas klusterisasi yang unggul. Jika dibandingkan dengan DBSCAN dan OPTICS, metode Agglomerative tidak hanya mampu membentuk kluster yang lebih terpisah secara efektif, tetapi juga menunjukkan konsistensi dalam representasi struktur hierarkis data yang dianalisis.

3.3. Visualisasi Hasil Clustering

Berikut adalah visualisasi clustering DBSCAN, OPTICS dan Agglomerative yang menggambarkan distribusi penyakit di wilayah Puskesmas Kabupaten Banjarnegara.



Gambar 3. Diagram Cluster DBSCAN, OPTICS dan Agglomerative

Gambar 3 menunjukkan perbandingan visual hasil clustering menggunakan tiga metode berbeda, yakni DBSCAN, OPTICS, dan Agglomerative Clustering, untuk mengelompokkan Puskesmas di

Kabupaten Banjarnegara berdasarkan pola penyakit. Masing-masing metode menghasilkan konfigurasi klaster yang mencerminkan karakteristik algoritmanya.

Visualisasi hasil DBSCAN teridentifikasi tiga klaster utama dan sejumlah data yang diklasifikasikan sebagai noise (berlabel -1). Klaster 0 merupakan klaster dengan jumlah anggota terbanyak, yang menunjukkan bahwa sebagian besar Puskesmas memiliki pola penyakit yang serupa. Klaster 1 memiliki jumlah anggota yang lebih sedikit, mencerminkan keberadaan Puskesmas dengan karakteristik penyakit yang berbeda dari mayoritas populasi. Didapati variasi signifikan dalam distribusi penyakit antar fasilitas layanan kesehatan, serta pentingnya pendekatan klasterisasi dalam memahami heterogenitas secara spasial. Hasil clustering dapat berubah drastis jika terjadi sedikit pergeseran pada parameter, tanpa prosedur tuning parameter yang sistematis, DBSCAN berpotensi menghasilkan klaster yang tidak stabil dan sulit direplikasi dalam analisis spasial epidemiologi [11].

Hasil clustering menggunakan metode OPTICS terhadap data Puskesmas di Kabupaten Banjarnegara diperoleh struktur kepadatan data yang bervariasi. Klaster -1 mencakup 14 Puskesmas, merupakan kelompok dengan jumlah anggota terbanyak dan merepresentasikan data yang diklasifikasikan sebagai outlier serta memiliki keragaman signifikan dalam pola penyebaran penyakit antar wilayah. Klaster 3 diperoleh klaster valid terbesar dengan 8 Puskesmas dengan menggambarkan kesamaan pola penyakit antar fasilitas kesehatan yang memungkinkan terbentuknya pengelompokan yang kuat. Hasil ini menegaskan bahwa metode OPTICS efektif dalam mengidentifikasi struktur kepadatan data, khususnya dalam mengelompokkan entitas dengan karakteristik penyakit yang seragam. Pendekatan berbasis kepadatan, seperti OPTICS, sangat berguna dalam klasifikasi data medis dan pemetaan penyakit, karena mampu secara simultan mengidentifikasi klaster dan mendeteksi outlier [27].

"Klaster 0, 1, dan 2 masing-masing terdiri dari 4 dan 3 Puskesmas, yang mengindikasikan keberadaan kelompok-kelompok kecil dengan karakteristik penyakit yang lebih spesifik atau menyimpang dari pola umum. Hal tersebut menunjukkan bahwa sebagian Puskesmas menghadapi tantangan kesehatan yang khas, yang memerlukan pendekatan pelayanan yang lebih terarah dan berbasis lokal. Penelitian sebelumnya juga mengungkapkan bahwa outlier dalam analisis klaster sering kali mencerminkan anomali penting yang dapat memberikan wawasan berharga bagi peningkatan sistem pelayanan kesehatan [28]. Dominasi titik data yang dikategorikan sebagai outlier (-1) menandakan bahwa distribusi penyakit di wilayah studi tidak homogen dan cukup kompleks. Penggunaan metode OPTICS mampu menangkap dimensi tersembunyi dari struktur data dan memperkaya pemahaman terhadap penyebaran penyakit. Hasil ini mendukung perlunya strategi intervensi yang lebih presisi dan berbasis bukti dalam pengelolaan kesehatan masyarakat [22].

Hasil Agglomerative Clustering terdapat tiga klaster terhadap pola distribusi layanan kesehatan di Puskesmas Kabupaten Banjarnegara. Klaster 1 mencakup jumlah anggota terbanyak, menunjukkan bahwa mayoritas Puskesmas memiliki kesamaan dalam pola penyakit yang ditangani dan dapat dikategorikan dalam satu kelompok dominan. Klaster 0 diperoleh dengan jumlah anggota menengah, menunjukkan keberadaan kelompok Puskesmas yang memiliki karakteristik berbeda dari klaster dominan. Klaster 2 terdiri dari jumlah anggota paling sedikit, yang mengindikasikan keberadaan kelompok Puskesmas dengan karakteristik pola penyakit yang unik. Analisis klaster dapat mengungkap perbedaan penting dalam distribusi penyakit dan karakteristik wilayah layanan yang memerlukan strategi pengelolaan yang disesuaikan [27].

Distribusi hasil klasterisasi diperoleh mayoritas Puskesmas memiliki karakteristik yang serupa dan dapat dikelompokkan dalam satu kategori yang homogen, terdapat pula sejumlah Puskesmas dengan pola yang khas dan menyimpang dari mayoritas. Kondisi ini menunjukkan perlunya pendekatan diferensial dalam perencanaan layanan kesehatan. Sehingga penerapan metode clustering hierarkis tidak hanya berfungsi sebagai alat pemetaan struktur layanan, tetapi juga menjadi instrumen strategis untuk

merancang intervensi yang lebih adaptif, kontekstual, dan berbasis pada karakteristik data yang sebenarnya.

4. DISKUSI

Hasil analisis clustering dengan metode DBSCAN, OPTICS, dan Agglomerative Clustering menunjukkan perbedaan dalam pengelompokan Puskesmas berdasarkan pola penyakit yang ada di Kabupaten Banjarnegara. Masing-masing metode memiliki keunggulan dan keterbatasan dalam mengidentifikasi cluster yang terbentuk berdasarkan distribusi data pasien.

4.1. Hasil perbandingan Clustering

Hasil analisis clustering pada penelitian ini menunjukkan bahwa masing-masing metode—DBSCAN, OPTICS, dan Agglomerative Clustering memiliki kekuatan dan kelemahan tersendiri dalam mengelompokkan Puskesmas berdasarkan pola penyakit. Agglomerative Clustering terbukti menghasilkan struktur klaster yang paling stabil dan dapat diinterpretasikan secara praktis, sebagaimana terlihat dari nilai Silhouette Score (0.362) yang relatif tinggi dan DBI (0.642) yang rendah, menandakan bahwa klaster yang terbentuk memiliki kompaksi internal dan pemisahan antar cluster yang baik.

Hasil ini konsisten dengan penelitian sebelumnya bahwa pendekatan agglomerative dalam pemetaan capaian vaksinasi COVID-19 di Puskesmas menghasilkan segmentasi wilayah yang lebih representative dengan metode hirarkis [8]. Penerapan k-means pada data penyakit tidak menular memiliki keterbatasan dalam menangani variasi spasial dan distribusi data yang kompleks. Sebaliknya, metode agglomerative lebih unggul dalam menangkap struktur hierarki dan variasi bertahap yang umum terjadi dalam data layanan kesehatan [20].

Pada metode DBSCAN memiliki keunggulan dalam mendeteksi outlier serta sangat sensitif terhadap pemilihan parameter ϵ dan MinPts. Pada konfigurasi optimal ($\epsilon = 3.5$, MinPts = 3), DBSCAN mampu menghasilkan Silhouette Score tertinggi (0.381) dan disertai menghasilkan banyak noise. Hasil clustering DBSCAN menjadi tidak stabil ketika terjadi variasi kecil dalam parameter, yang berpotensi menghasilkan jumlah cluster dan komposisi yang sangat berbeda. DBSCAN tanpa penyesuaian adaptif parameter berisiko menghasilkan segmentasi yang misleading pada data spasial kesehatan [11].

Metode OPTICS yang dirancang dalam mengatasi keterbatasan DBSCAN untuk menangani variasi kepadatan, memberikan hasil yang kurang optimal, ditandai dengan Silhouette Score yang relatif rendah (maksimum 0.327) dan bahkan negatif (-0.0025) pada konfigurasi tertentu. Terdapat outlier (-1) dalam hasil OPTICS menunjukkan bahwa distribusi penyakit di Puskesmas memiliki heterogenitas tinggi serta memiliki nilai dalam konteks eksplorasi reachability dan mendeteksi substruktur lokal yang tersembunyi.

4.2. Implikasi Hasil

Hasil penelitian ini memiliki implikasi penting bagi perencanaan dan pengambilan keputusan strategis di sektor kesehatan daerah, khususnya oleh Dinas Kesehatan Kabupaten Banjarnegara. Temuan bahwa metode Agglomerative Clustering menghasilkan segmentasi yang lebih stabil dan representatif memberikan dasar ilmiah yang kuat untuk menyusun kebijakan berbasis klaster. Dinas Kesehatan dapat mengembangkan intervensi yang lebih terarah dan efisien, baik dari sisi penyuluhan, distribusi logistik, maupun penempatan tenaga medis dengan mengelompokkan Puskesmas berdasarkan kesamaan pola penyakit yang ditangani,

Pengidentifikasian klaster ekstrem atau outlier melalui metode DBSCAN dan OPTICS dapat dimanfaatkan untuk merancang pendekatan khusus pada wilayah yang memiliki beban penyakit yang memerlukan perhatian lebih. Pendekatan data driven berpotensi mempercepat respons pelayanan kesehatan primer dan meningkatkan efisiensi penggunaan sumber daya di tengah keterbatasan anggaran

dan kapasitas pelayanan. Integrasi hasil clustering ke dalam sistem perencanaan kesehatan daerah mampu memperkuat tata kelola kesehatan yang lebih adaptif, responsif, dan berbasis kebutuhan riil masyarakat.

5. KESIMPULAN

Penelitian ini menunjukkan bahwa metode Agglomerative Clustering paling efektif dalam mengelompokkan Puskesmas berdasarkan distribusi penyakit, menghasilkan struktur klaster yang stabil dan informatif untuk mendukung perencanaan layanan kesehatan. Sementara DBSCAN dan OPTICS memberikan keunggulan dalam mendeteksi noise dan variasi lokal, performa keduanya cenderung dipengaruhi oleh sensitivitas parameter dan karakteristik distribusi data.

Penelitian ini memiliki keterbatasan, terutama pada penggunaan dataset yang hanya mencakup satu kabupaten dalam rentang waktu tertentu, sehingga hasilnya belum dapat digeneralisasi secara luas. Diperlukan penerapan metode clustering yang diuji pada dataset multi tahun dan lintas wilayah dalam mengevaluasi konsistensi pola klaster, dengan penambahan integrasi variabel spasial, demografis, dan sumber daya kesehatan. Rekomendasi ini penting sebagai langkah lanjutan dalam membangun sistem pengambilan keputusan kesehatan yang berbasis data dan responsif terhadap dinamika penyakit di tingkat layanan primer.

DAFTAR PUSTAKA

- [1] N. G. F. M. Beukers, N. Su, G. J. M. G. v. d. Heijden, and B. G. Loos, "Periodontitis is associated with multimorbidity in a large dental school population," *J. Clin. Periodontol.*, vol. 50, no. 12, pp. 1621–1632, 2023, doi: 10.1111/jcpe.13870.
- [2] M. T. Kumbeni, P. A. Apanga, M. W. Chanase, J. N. Alem, and N. Mireku-Gyimah, "The role of the public and private health sectors on factors associated with early essential newborn care practices among institutional deliveries in Ghana," *BMC Health Serv. Res.*, vol. 21, no. 1, 2021, doi: 10.1186/s12913-021-06665-0.
- [3] Kementerian Kesehatan RI, "Profil kesehatan Indonesia Tahun 2023," 2023, [Online]. Available: <https://www.kemkes.go.id/id/profil-kesehatan-indonesia-2023>
- [4] F. Jaotombo *et al.*, "Machine-learning prediction for hospital length of stay using a French medico-administrative database," *J. Mark. Access Heal. Policy*, vol. 11, no. 1, 2022, doi: 10.1080/20016689.2022.2149318.
- [5] K. Davagdorj, J. Lee, K. H. Park, and K. H. Ryu, "Cost-sensitive neural network for prediction of hypertension using class imbalance dataset," in *Advances in Intelligent Information Hiding and Multimedia Signal Processing*, 2021, pp. 44–51. doi: 10.1007/978-981-33-6757-9_6.
- [6] J. Fang *et al.*, "A new particle swarm optimization algorithm for outlier detection: industrial data clustering in wire arc additive manufacturing," *IEEE Trans. Autom. Sci. Eng.*, vol. 21, no. 2, pp. 1244–1257, 2024, doi: 10.1109/tase.2022.3230080.
- [7] J. Rustamov, Z. Rustamov, N. Badawi, R. Damseh, F. Jasmi, and N. Zaki, "Benchmarking outlier detection algorithms in healthcare: an analysis across diverse medical datasets," 2024, doi: 10.21203/rs.3.rs-5242162/v1.
- [8] D. P. Sari and N. Sumita, "Achievement cluster of COVID-19 vaccination at the South Bengkulu Health Center using agglomerative hierarchical clustering," *J. Stat. Data Sci.*, vol. 1, no. 2, pp. 10–16, 2022, doi: 10.33369/jsds.v1i2.24082.
- [9] C. Tang, H. Wang, Z. Wang, X. Zeng, H. Yan, and Y. Xiao, "An improved optics clustering algorithm for discovering clusters with uneven densities," *Intell. Data Anal.*, vol. 25, no. 6, pp. 1453–1471, 2021, doi: 10.3233/ida-205497.
- [10] X. Han, C. Armenakis, and M. Jadidi, "Dbscan optimization for improving marine trajectory clustering and anomaly detection," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XLIII-B4-2, pp. 455–461, 2020, doi: 10.5194/isprs-archives-xliii-b4-2020-455-202.
- [11] X. Zhang and S. Zhou, "Woa-dbscan: application of whale optimization algorithm in dbscan parameter adaption," *IEEE Access*, vol. 11, pp. 91861–91878, 2023, doi:

- 10.1109/access.2023.3307412.
- [12] L. Chen, C. Gu, S. Zheng, and Y. Wang, "A method for identifying gross errors in dam monitoring data," *Water*, vol. 16, no. 7, p. 978, 2024, doi: 10.3390/w16070978.
 - [13] S. Li, "An improved dbSCAN algorithm based on the neighbor similarity and fast nearest neighbor query," *IEEE Access*, vol. 8, pp. 47468–47476, 2020, doi: 10.1109/access.2020.2972034.
 - [14] J. Lin, C. Shang, and Q. Shen, "Towards dynamic fuzzy rule interpolation via density-based rule promotion from interpolated outcomes," *Mathematics*, vol. 12, no. 3, p. 402, 2024, doi: 10.3390/math12030402.
 - [15] S. Kossieris, M. Asgarimehr, and J. Wickert, "Unsupervised machine learning for gnss reflectometry inland water body detection," *Remote Sens.*, vol. 15, no. 12, p. 3206, 2023, doi: 10.3390/rs15123206.
 - [16] F. T. T. Lai *et al.*, "Comparing multimorbidity patterns among discharged middle-aged and older inpatients between Hong Kong and Zurich: a hierarchical agglomerative clustering analysis of routine hospital records," *Front. Med.*, vol. 8, 2021, doi: 10.3389/fmed.2021.651925.
 - [17] J. Dufek, X. Xiao, and R. R. Murphy, "Best viewpoints for external robots or sensors assisting other robots," *IEEE Trans. Human-Machine Syst.*, vol. 51, no. 4, pp. 324–334, 2021, doi: 10.1109/thms.2021.3090765.
 - [18] A. Rashid *et al.*, "Geochemical modeling source provenance, public health exposure, and evaluating potentially harmful elements in groundwater: statistical and human health risk assessment (hhra)," *Int. J. Environ. Res. Public Health*, vol. 19, no. 11, p. 6472, 2022, doi: 10.3390/ijerph19116472.
 - [19] R. Vatambeti and V. K. Damera, "Intelligent diagnosis of obstetric diseases using hgs-aoa based extreme learning machine," *Acadlore Trans. AI Mach. Learn.*, vol. 2, no. 1, pp. 21–32, 2023, doi: 10.56578/ataiml020103.
 - [20] T. A. Munandar and A. Y. Y. Pratama, "Regional clustering based on types of non-communicable diseases using k-means algorithm," *MATRIK J. Manajemen, Tek. Inform. Dan Rekayasa Komput.*, vol. 23, no. 2, pp. 285–296, 2024, doi: 10.30812/matrik.v23i2.3352.
 - [21] E. W. Ambarsari, N. Dwitianti, N. Selvia, W. N. Cholifah, and P. D. Mardika, "Comparison approaches of the fuzzy c-means and gaussian mixture model in clustering the welfare of the Indonesian people," *KnE Soc. Sci.*, 2023, doi: 10.18502/kss.v8i9.13315.
 - [22] H. Xu, L. Zhang, P. Li, and F. Zhu, "Outlier detection algorithm based on k-nearest neighbors-local outlier factor," *J. Algorithm. Comput. Technol.*, vol. 16, 2022, doi: 10.1177/17483026221078111.
 - [23] E. O. Amoako *et al.*, "Examining facilitative services for entry into substance use disorder treatment: a cluster analysis of treatment facilities," *PLoS One*, vol. 19, no. 5, p. e0304094, 2024, doi: 10.1371/journal.pone.0304094.
 - [24] J. Carmona-Pérez *et al.*, "Multimorbidity clusters in patients with chronic obstructive airway diseases in the epicron cohort," *Sci. Rep.*, vol. 11, no. 1, 2021, doi: 10.1038/s41598-021-83964-w.
 - [25] M. Science, E. Division, and S. Arabia, "Bayesian spatial functional data clustering: applications in disease surveillance b," pp. 1–19, 2019, [Online]. Available: <https://doi.org/10.48550/arXiv.2407.12633>
 - [26] M. F. Guagliardo, K. A. Jablonski, J. G. Joseph, and D. C. Goodman, "Do pediatric hospitalizations have a unique geography?," vol. 9, pp. 1–9, 2004.
 - [27] V. N. Vakharia, S. Toescu, A. J. Copp, and D. Thompson, "A topographical analysis of encephalocele locations: generation of a standardised atlas and cluster analysis," *Child's Nerv. Syst.*, vol. 39, no. 7, pp. 1911–1920, 2023, doi: 10.1007/s00381-023-05883-7.
 - [28] M. B. Hossen, "A new statistical approach for the identification of outlier genes in cancer microarray data," *Asian J. Med. Biol. Res.*, vol. 6, no. 4, pp. 795–801, 2021, doi: 10.3329/ajmbr.v6i4.51248.