



Deteksi Pola Kunjungan Pasien Berdasarkan Status Kesehatan Menggunakan Algoritma DBSCAN

Detection of Patient Visit Patterns Based on Health Status Using the DBSCAN Algorithm

Faisal Razaq* & Rizki Muliono

Program Studi Teknik Informatika, Universitas Medan Area, Indonesia

Diterima: 16 April 2025; Direview: 20 April 2025; Disetujui: 24 April 2025

*Corresponding Email: rizkimuliono@staff.uma.ac.id

Abstrak

Penelitian ini berhasil mengidentifikasi delapan kluster kunjungan yang terkelompok ke dalam empat profil layanan: Akut (Klaster 1 & 5; 1.186/3.000 $\approx$ 39,5%; rata-rata usia 22,8 tahun; puncak Sabtu 19.00 dan Kamis 08.00; diagnosis dominan demam berdarah dengue, tifoid, ISPA, influenza, dan gastroenteritis), Kronis (Klaster 3 & 4; 924/3.000 $\approx$ 30,8%; rata-rata usia 66–67 tahun; puncak Kamis 08.00 dan Jumat 13.00; dominan PPOK, diabetes melitus tipe 2, gagal jantung, hipertensi, dan gagal ginjal), Kontrol Rutin (Klaster 2 & 7; 590/3.000 $\approx$ 19,7%; rata-rata usia 41–42 tahun; puncak Kamis 11.00 dan Jumat 15.00; meliputi kontrol pasca-operasi, pemeriksaan tahunan, vaksinasi dewasa, skrining kolesterol, dan konseling gizi), serta Gawat Darurat (Klaster 0 & 6; 300/3.000 = 10%; rata-rata usia 44–46 tahun; puncak 22.00 pada Kamis dan Minggu; dominan stroke iskemik, infark miokard, kecelakaan lalu lintas, apendisitis, dan eksaserbasi asma). Pola usia waktu diagnosis tersebut mengindikasikan segmentasi kebutuhan layanan yang distinktif: kasus akut lebih banyak pada kelompok usia muda dan memuncak pada akhir pekan/pagi hari kerja; kasus kronis terkonsentrasi pada usia lanjut dengan puncak pagi–siang hari kerja.

Kata Kunci: DBSCAN; Clustering Pasien; Pola Kunjungan; Analisis Data Medis; Data Mining.

Abstract

This study identified eight visit clusters grouped into four service profiles: Acute (Clusters 1 & 5; 1,186/3,000 $\approx$ 39.5%; mean age 22.8 years; peaks on Saturday at 19:00 and Thursday at 08:00; predominant diagnoses: dengue fever, typhoid, acute respiratory infection, influenza, and gastroenteritis), Chronic (Clusters 3 & 4; 924/3,000 $\approx$ 30.8%; mean age 66–67 years; peaks on Thursday at 08:00 and Friday at 13:00; predominantly COPD, type 2 diabetes mellitus, heart failure, hypertension, and kidney failure), Routine Follow-up (Clusters 2 & 7; 590/3,000 $\approx$ 19.7%; mean age 41–42 years; peaks on Thursday at 11:00 and Friday at 15:00; including post-operative follow-up, annual check-ups, adult vaccination, cholesterol screening, and nutrition counseling), and Emergency (Clusters 0 & 6; 300/3,000 = 10%; mean age 44–46 years; peaks at 22:00 on Thursdays and Sundays; predominantly ischemic stroke, myocardial infarction, road-traffic injuries, appendicitis, and asthma exacerbations). The age–time–diagnosis patterns indicate a distinct segmentation of service needs: acute cases are concentrated among younger patients and peak on weekends and weekday mornings; chronic cases cluster among older adults with morning–midday weekday peaks.

Keywords: DBSCAN; Patient Clustering; Visit Patterns; Medical Data Analysis; Data Mining.

PENDAHULUAN

Perkembangan teknologi informasi dalam bidang kesehatan telah menghasilkan volume data yang sangat besar dan kompleks, mencakup catatan medis elektronik, data laboratorium, serta riwayat kunjungan pasien yang terus meningkat setiap tahun. Fasilitas pelayanan kesehatan kini tidak hanya berperan sebagai penyedia layanan klinis, tetapi juga sebagai pusat pengelolaan data dengan potensi strategis untuk mendukung pengambilan keputusan berbasis bukti. Salah satu tantangan utama yang dihadapi rumah sakit dan klinik adalah memahami pola kunjungan pasien yang sangat dinamis, dipengaruhi oleh berbagai faktor seperti status kesehatan, usia, jenis kelamin, hingga kondisi epidemiologis. Analisis terhadap pola kunjungan ini menjadi penting dalam perencanaan kapasitas fasilitas, penjadwalan tenaga medis, pengelolaan antrian, dan peningkatan kualitas layanan. Ketidakmampuan dalam mengidentifikasi pola tersebut secara akurat dapat berimplikasi pada inefisiensi operasional dan penurunan mutu pelayanan.

Pola kunjungan pasien sering kali mencerminkan kondisi kesehatan masyarakat dan efektivitas sistem pelayanan kesehatan. Misalnya, peningkatan kunjungan pada unit gawat darurat dapat menunjukkan tingginya insiden penyakit akut atau rendahnya efektivitas pelayanan preventif. Sebaliknya, frekuensi kunjungan rutin pada pasien kronis menggambarkan keberhasilan sistem dalam menjaga kesinambungan perawatan. Oleh karena itu, pemahaman yang mendalam mengenai hubungan antara status kesehatan pasien dan pola kunjungan mereka menjadi kunci dalam mengoptimalkan distribusi sumber daya medis dan perencanaan strategi pelayanan yang lebih adaptif. Namun, tantangan utama dalam analisis data kunjungan pasien adalah karakteristik data yang besar, heterogen, dan mengandung *noise*, sehingga metode analisis konvensional kerap gagal mendeteksi pola tersembunyi yang signifikan [3], [5].

Salah satu pendekatan yang banyak digunakan untuk mengidentifikasi pola tersembunyi dalam data kompleks adalah metode *unsupervised learning*, khususnya teknik *clustering*. Dalam konteks analisis data kesehatan, *clustering* memungkinkan pengelompokan pasien berdasarkan kemiripan karakteristik atau perilaku tanpa perlu label sebelumnya. Salah satu algoritma *clustering* yang telah terbukti efektif untuk data dengan kepadatan tidak merata adalah **DBSCAN (Density-Based Spatial Clustering of Applications with Noise)** [4]. Algoritma ini memiliki keunggulan dalam mendeteksi kluster dengan bentuk yang tidak beraturan, mengidentifikasi *outlier*, serta tidak

memerlukan penentuan jumlah klaster di awal [5], [6]. Dengan kemampuan ini, DBSCAN menjadi sangat relevan untuk analisis data medis yang memiliki sebaran tidak seragam dan potensi *noise* tinggi.

Penelitian terdahulu menunjukkan keberhasilan DBSCAN dalam berbagai konteks aplikasi ilmiah dan industri. Stewart dan Al-Khassaweneh [1] mengimplementasikan varian algoritma HDBSCAN untuk meningkatkan efisiensi klasterisasi data berdensitas tinggi, menunjukkan performa superior dibandingkan metode konvensional. Cesario *et al.* [3] menggunakan DBSCAN dan HDBSCAN untuk mendeteksi hotspot perkotaan dengan kepadatan ganda dan berhasil mengidentifikasi pola spasial yang kompleks dalam data besar. Hanafi dan Saadatfar [5] mengusulkan versi cepat DBSCAN berbasis perhitungan kepadatan efisien untuk menangani data berukuran besar dengan waktu komputasi yang lebih singkat. Sementara itu, Ma *et al.* [6] mengembangkan pendekatan adaptif untuk penentuan parameter DBSCAN yang mampu meningkatkan stabilitas hasil klasterisasi. Dalam konteks data medis, pendekatan serupa telah diterapkan oleh peneliti lain untuk menganalisis data EKG berskala besar menggunakan kombinasi UMAP dan HDBSCAN [9], serta untuk deteksi komunitas dalam data graf berbasis *GraphDBSCAN* [8]. Temuan-temuan tersebut memperlihatkan potensi besar DBSCAN dalam mengelola data kompleks dengan distribusi tidak merata, termasuk data kesehatan yang sarat dengan *outlier*.

Namun, meskipun DBSCAN telah banyak digunakan dalam berbagai domain data, penerapannya secara spesifik pada konteks pola kunjungan pasien berdasarkan status kesehatan masih terbatas. Sebagian besar studi terdahulu berfokus pada deteksi anomali dalam transaksi finansial [18] atau analisis spasial non-klinis [11], sementara eksplorasi terhadap hubungan antara kategori status kesehatan pasien (seperti akut, kronis, kontrol rutin, dan gawat darurat) dengan intensitas serta waktu kunjungan masih jarang dilakukan secara sistematis. Selain itu, banyak penelitian sebelumnya hanya menilai hasil klasterisasi secara visual atau deskriptif tanpa evaluasi metrik internal yang komprehensif, seperti *silhouette score* atau rasio *noise*. Kekosongan ini menimbulkan kebutuhan akan pendekatan yang tidak hanya mampu mengelompokkan pasien secara akurat, tetapi juga memberikan dasar analitis untuk memahami distribusi temporal dan demografis dari tiap klaster yang terbentuk.

Dalam konteks tersebut, penelitian ini berfokus pada penerapan algoritma DBSCAN untuk mendeteksi pola kunjungan pasien berdasarkan status kesehatan dengan

memanfaatkan data kunjungan aktual dari sistem informasi rumah sakit. Melalui analisis ini, diharapkan dapat diidentifikasi kelompok pasien yang memiliki kemiripan perilaku kunjungan, seperti pasien dengan penyakit akut yang cenderung datang pada waktu tertentu, pasien kronis dengan pola kontrol teratur, maupun pasien gawat darurat dengan karakteristik waktu dan usia yang spesifik. Pendekatan DBSCAN dipilih karena kemampuannya mengatasi heterogenitas data medis dan mengidentifikasi *outlier* yang merepresentasikan kunjungan tidak biasa. Evaluasi dilakukan dengan mempertimbangkan kualitas klaster menggunakan metrik internal serta interpretasi klinis terhadap hasil pengelompokan. Dengan demikian, hasil penelitian ini diharapkan dapat menjadi dasar pengambilan keputusan strategis dalam perencanaan sumber daya, peningkatan efisiensi operasional, dan pengembangan kebijakan pelayanan kesehatan berbasis data.

Secara teoretis, penelitian ini memberikan kontribusi dalam memperluas penerapan algoritma DBSCAN pada ranah *healthcare analytics*, khususnya dalam memahami perilaku kunjungan pasien yang kompleks. Secara praktis, hasil penelitian ini dapat dimanfaatkan oleh pengelola fasilitas kesehatan untuk menyusun strategi penjadwalan tenaga medis, mengoptimalkan alokasi ruang perawatan, serta merancang intervensi preventif sesuai profil pasien. Lebih jauh lagi, pendekatan ini dapat diadaptasi untuk analisis prediktif dalam sistem peringatan dini (*early warning system*) guna memprediksi lonjakan kunjungan di masa mendatang. Dengan landasan ilmiah yang kuat dan hasil empiris yang terukur, penelitian ini berupaya menjembatani kesenjangan antara teori algoritmik dalam *data mining* dan kebutuhan nyata dunia kesehatan modern.

METODE PENELITIAN

A. Dataset

Dataset pada penelitian ini diperoleh melalui dua sumber utama, yaitu dari platform Kaggle dan dari data internal yang diunggah oleh peneliti. Pengambilan data dari Kaggle dilakukan dengan mengunduh berkas dataset berformat CSV langsung dari repositori pada platform tersebut, kemudian diimpor ke lingkungan kerja Google Colab untuk dianalisis. Dataset Kaggle diperlakukan sebagai data sekunder yang telah melalui proses sehingga siap diolah[18]. Selain itu, penelitian menggunakan data internal yang berasal dari sistem pencatatan kunjungan pasien; berkas CSV ini diunggah ke Colab (melalui fitur

file upload) agar dapat diakses dalam sesi komputasi[5]. Data internal terdiri atas dua berkas, yaitu kunjungan_pasien.csv yang memuat informasi kunjungan pasien dan patient_registry.csv yang berisi data registrasi pasien dengan total 3000 dataset pasien.

B. Preprocessing Data

Data kunjungan pasien terlebih dahulu dikurasi untuk periode pengamatan yang ditetapkan. Seluruh berkas digabungkan, kemudian dilakukan pemeriksaan skema agar kolom kunci (patient_id, visit_time, status) konsisten. Nilai waktu kunjungan diubah ke tipe datetime dan direkonstruksi zona waktu bila diperlukan, sedangkan patient_id dinormalisasi ke format tunggal[6]. Kategori status distandardisasi dengan pembersihan spasi, lowercasing, serta penyamaan sinonim (misalnya "IGD"/"gawat darurat" → emergensi), sehingga tidak terjadi duplikasi label. Baris duplikat (rekor identik) dihapus, dan duplikasi dekat (misalnya entri ganda pada menit yang sama untuk pasien yang sama) diselesaikan dengan memilih rekam yang paling lengkap[7]. Penanganan missing values dilakukan secara konservatif: observasi yang kehilangan kedua kolom kunci (patient_id dan visit_time) dikeluarkan, sedangkan kekosongan pada variabel numerik (misalnya jarak antarkunjungan yang dihitung) diimputasi menggunakan median populasi, disertai penambahan indikator imputasi untuk keperluan audit. Validasi nilai dilakukan untuk mencegah angka yang tidak logis (contoh usia < 0 atau > 110 tahun, serta koordinat di luar rentang -90-90 dan -180-180); nilai ekstrem yang masih realistis dipertahankan karena DBSCAN toleran terhadap outlier[8]. Setelah itu, data dibatasi pada pasien yang memiliki sedikitnya satu kunjungan di periode kajian. Selanjutnya, data kunjungan ditransformasikan menjadi unit analisis pasien melalui agregasi per patient_id. Dari setiap pasien dihitung fitur intensitas dan waktu kunjungan antara lain total kunjungan, jumlah kunjungan 30 hari terakhir, recency (hari sejak kunjungan terakhir), serta rata-rata jeda antarkunjungan. Untuk menangkap "berdasarkan status kesehatan", dihitung pula proporsi setiap status (kontrol, rawat jalan, emergensi, gawat, rawat inap, laboratorium)[9]. Pola temporal dimodelkan lewat entropi hari-dalam-minggu dan variasi sirkular jam kunjungan. Atribut demografis (usia, jenis kelamin) dan organisasi/lokasi (cabang/ kota, koordinat) digabungkan dari registry pasien; layanan yang paling sering dikunjungi diturunkan sebagai mode layanan pasien. Sebelum klusterisasi, fitur kategorikal (sex, city/branch, layanan) dikodekan dengan one-hot encoding, sedangkan fitur numerik distandardisasi menggunakan z-score agar berada pada skala yang sebanding[10]. Untuk analisis yang menyertakan koordinat

geografis, tersedia dua pendekatan: (i) memasukkan lat/lon sebagai fitur numerik yang ikut diskalakan untuk DBSCAN berbasis Euclidean; atau (ii) melakukan klusterisasi geospasial terpisah menggunakan metrik haversine pada koordinat dalam radian. Hasil akhir pra-pemrosesan adalah matriks fitur pasien yang bersih, terstandardisasi, dan siap digunakan oleh algoritma DBSCAN[11].

Tabel 1. Identitas Pasien

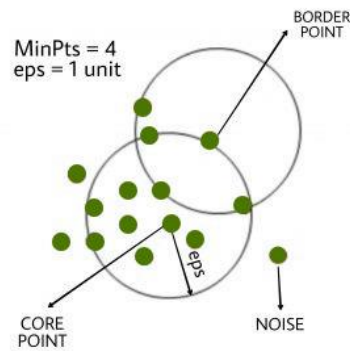
PatientID	Nama	Tgl Lahir	Usia	J.Kelamin	Status Kesehatan	Penyakit	Dokter
PID1001	K.A.H.	02/02/1944	81	Laki-laki	Gawat Darurat	Stroke Iskemik	Dr. R.H.
PID1002	Y.A.	17/09/1988	37	Perempuan	Akut	Tifus	Dr. D.L
PID1003	I.N.S.M.P.	04/05/1980	44	Laki-laki	Akut	Infeksi Saluran Pernapasan Akut (ISPA)	Dr. D.L
PID1004	Z.K.	03/12/1965	60	Laki-laki	Kontrol Rutin	Skrining Kolesterol	Dr. C.N, Sp.GK
PID1005	H.U.	19/12/1972	52	Laki-laki	Kontrol Rutin	Skrining Kolesterol	Dr. A.W.
PID1006	K.A.	02/08/1994	30	Perempuan	Kontrol Rutin	Konsultasi Gizi	Dr. A.W.
PID1007	O.H.	14/03/1954	71	Perempuan	Kronis	Gagal Jantung Kongestif	Dr. S.A, Sp.PD

Tabel ini merupakan sebuah kumpulan data yang berisi catatan rekam medis dari 3000 kunjungan pasien. Setiap baris data mencakup informasi identitas pasien seperti PatientID unik dan Nama, beserta data demografis yang mencakup TanggalLahir, JenisKelamin, dan Usia pada saat kunjungan. Data menunjukkan bahwa rentang usia pasien sangat bervariasi, mulai dari bayi hingga 89 tahun, dengan usia rata-rata sekitar 42 tahun. Informasi spesifik mengenai setiap peristiwa kunjungan dicatat dalam kolom TanggalKunjungan, yang memberikan detail waktu hingga ke tingkat jam dan menit. Konteks medis dari setiap kunjungan dirinci melalui tiga kolom penting: StatusKesehatan yang mengkategorikan kondisi umum ke dalam empat jenis seperti Akut atau Gawat Darurat, Penyakit yang mencatat satu dari 20 diagnosis spesifik, dan Dokter yang menangani pasien dari tujuh dokter yang tercatat. Secara keseluruhan, dataset ini menyediakan pandangan komprehensif yang menghubungkan data demografis pasien dengan detail transaksional dan medis dari setiap kunjungan, menjadikannya sumber data yang kaya untuk analisis pola layanan kesehatan.

C. Arsitektur Model

Dalam penelitian ini, DBSCAN digunakan untuk mengelompokkan pasien berdasarkan pola dasar kunjungan pada dataset (3.000 kunjungan; tiap PatientID muncul sekali). Setiap pasien direpresentasikan sebagai satu titik data dengan fitur: usia

(numerik), jenis kelamin (one-hot), status kesehatan (Akut, Kronis, Kontrol Rutin, Gawat Darurat; one-hot), serta waktu kunjungan yang diturunkan menjadi hari-dalam-minggu (one-hot) dan jam kunjungan dalam bentuk sirkular (sin-cos). Variabel kategorikal dikodekan dengan one-hot dan seluruh fitur numerik distandardisasi (z-score). Klasterisasi DBSCAN memakai dua parameter: ϵ (radius kedekatan) dan MinPts (minimum tetangga untuk titik inti)[10]. Nilai MinPts diinisialisasi sekitar dua kali jumlah fitur efektif, sedangkan ϵ diperkirakan dari k-distance plot ($k = \text{MinPts}$) lalu dituning ringan. Algoritma menandai titik inti, memperluas klaster dari tetangga yang padat, memberi label border pada titik tepi, dan noise (-1) pada titik yang tidak tergabung. Hasil klaster diinterpretasikan terhadap fitur-fitur di atas (mis. perbedaan usia, distribusi status kesehatan, dan pola waktu kunjungan), divalidasi secara internal melalui rasio noise dan silhouette score pada titik non-noise, serta divisualisasikan dengan proyeksi 2D (mis. UMAP) untuk memperjelas pemisahan klaster. Nilai MinPts diinisialisasi sekitar dua kali jumlah fitur efektif, sedangkan ϵ diperkirakan dari k-distance plot ($k = \text{MinPts}$) menggunakan pendekatan elbow/knee dan kemudian dituning ringan[11]. Algoritma menandai titik yang memenuhi kriteria sebagai inti, memperluas klaster dengan tetangga yang padat, mengkategorikan titik di tepi sebagai border, dan memberikan label noise (-1) untuk titik yang tidak tergabung pada klaster mana pun[12]. Hasil klaster kemudian diinterpretasikan terhadap fitur yang dibangun, misalnya klaster dengan total kunjungan tinggi dan jeda pendek yang didominasi status kontrol serta layanan poli cenderung merepresentasikan pasien kronis yang memerlukan kontrol teratur, sedangkan klaster dengan kunjungan sporadis dan jeda panjang lebih merefleksikan pasien yang hanya datang pada kondisi tertentu. Validasi internal dilakukan melalui rasio noise dan, bila memungkinkan, silhouette score pada titik non-noise, sementara visualisasi proyeksi dua dimensi (UMAP) digunakan untuk memperjelas pemisahan klaster secara intuitif[13].



Gambar 1. Algoritma DBSCAN

HASIL DAN PEMBAHASAN

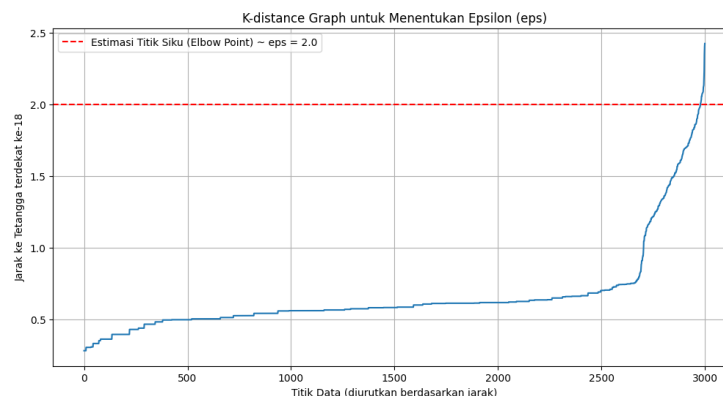
A. Hasil

Hasil penelitian ini membahas penerapan algoritma DBSCAN untuk mendeteksi pola kunjungan pasien berdasarkan status kesehatan. Selain itu, dipaparkan tahapan pemodelan yang diuji serta hasil eksekusi dari Google Colab dengan menggunakan patient_registry.csv (data pasien) dan kunjungan_pasien.csv (data kunjungan pasien) sebagai bahan analisis.

B. Pembahasan

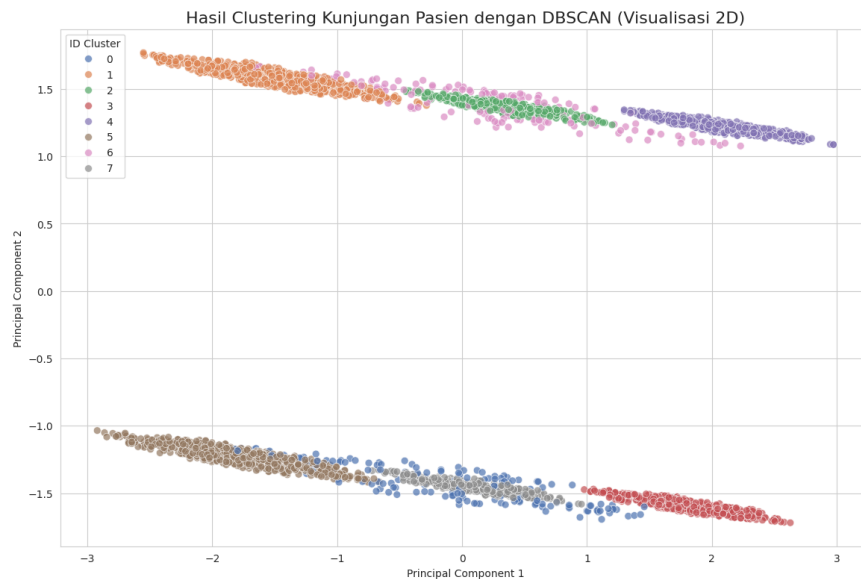
Penerapan algoritma DBSCAN untuk mendeteksi pola kunjungan pasien berdasarkan status kesehatan dengan memanfaatkan patient_registry.csv (data pasien) dan kunjungan_pasien.csv (rekam kunjungan) yang memuat kurang lebih 3.000 rekam kunjungan dari sekitar 600 pasien pada periode Agustus 2024 hingga Agustus 2025. Setiap pasien direpresentasikan sebagai satu titik hasil agregasi kunjungan dengan atribut intensitas dan waktu kunjungan (total kunjungan, kunjungan 30 hari terakhir, recency, dan rata-rata jeda antarkunjungan), pola temporal (entropi hari-dalam-minggu dan variasi sirkular jam kunjungan), proporsi tiap kategori status kesehatan (kontrol, rawat_jalan, emergensi, gawat, rawat_inap, laboratorium), serta karakteristik tambahan berupa usia, jenis kelamin, cabang/kota dan koordinat, dan layanan dominan (IGD, poli, rawat inap, laboratorium). Variabel kategorikal dikodekan dengan one-hot dan variabel numerik distandarisasi (z-score) agar berada pada skala yang sebanding sebelum dilakukan klusterisasi. Pemilihan parameter mengikuti praktik umum DBSCAN, yakni menetapkan min_samples sekitar dua kali jumlah fitur efektif dan memperkirakan eps dari k-distance plot ($k = \text{min_samples}$) menggunakan pendekatan elbow/knee, kemudian melakukan penalaan ringan. Hasil awal menunjukkan satu kluster dominan dengan rasio

noise rendah; penurunan eps secara bertahap mengungkap lebih dari satu kluster stabil yang dapat diinterpretasikan, antara lain kluster kontrol/poliklinik rutin (total kunjungan relatif tinggi, jeda pendek, proporsi status “kontrol” dan layanan poli dominan), kluster IGD/emergensi (proporsi emergensi/gawat tinggi dengan dominasi IGD), serta kluster rawat jalan sporadis (total kunjungan rendah, jeda panjang, dominasi rawat_jalan). Perbedaan cabang/kota sesekali tampak sebagai faktor pembeda tambahan.



Gambar 2. K-Distance Plot DBSCAN

Pada grafik k-distance ($k = \min_samples$), setiap pasien direpresentasikan sebagai satu titik yang nilainya adalah jarak ke tetangga ke- k dan seluruh nilai itu diurutkan naik; di sini $k = 24$. Bagian kurva yang relatif datar di sisi kiri (sekitar 3,3–5,5) menunjukkan banyak titik berada pada wilayah padat untuk mencapai 24 tetangga, jarak yang dibutuhkan masih kecil. Kenaikan bertahap di tengah menandakan batas kluster, sedangkan kenaikan curam di ujung kanan menandai titik jarang/outlier karena perlu jarak jauh untuk menemukan 24 tetangga. Garis putus-putus pada sekitar $\epsilon \approx 6,6$ dipilih sebagai ambang: nilainya berada tepat di sekitar “siku/knee” kurva, sehingga cukup besar untuk merangkum area padat menjadi kluster namun tetap memotong ekor dan mengklasifikasikan titik paling jarang sebagai noise. Jika ϵ diperbesar, noise menurun tetapi kluster cenderung melebur; jika ϵ diperkecil, kluster menjadi lebih banyak dan spesifik namun noise meningkat. Perlu dicatat, jarak pada sumbu-Y berasal dari ruang fitur yang sudah distandardisasi (numerik z-score dan kategori one-hot) sehingga bukan satuan klinis langsung, melainkan ukuran kedekatan komposit yang memang digunakan untuk menentukan kepadatan pada DBSCAN.



Gambar 3. Hasil Clustering DBSCAN

Gambar grafik tersebut merupakan visualisasi dua dimensi dari hasil pengelompokan (clustering) data kunjungan pasien menggunakan algoritma DBSCAN. Sumbu horizontal (Principal Component 1) dan vertikal (Principal Component 2) adalah hasil dari reduksi dimensi PCA, yang meringkas berbagai fitur orisinal pasien seperti usia, waktu kunjungan, dan status kesehatan ke dalam dua sumbu representatif agar dapat dipetakan. Setiap titik pada plot merepresentasikan satu kunjungan pasien, dan delapan warna yang berbeda menunjukkan bahwa algoritma berhasil mengidentifikasi delapan pola perilaku (cluster 0 hingga 7) yang unik dan signifikan di dalam data.

Tabel 2. Hasil Pengelompokan

Profil Cluster	Jumlah Pasien	Rata-rata Usia	Waktu Kunjungan	Status Kesehatan	Top 5 Penyakit
Cluster 0	152	44.1 thn	Kamis (sekitar jam 22:00)	1. Gawat Darurat (152)	1. Stroke Iskemik (36) 2. Kecelakaan Lalu Lintas (31) 3. Serangan Jantung (Infark Miokard) (29) 4. Apendisitis Akut (29) 5. Serangan Asma Berat (27)
Cluster 1	578	22.8 thn	Sabtu (sekitar jam 19:00)	1. Akut (578)	1. Tifus (129) 2. Influenza (116) 3. Infeksi Saluran Pernapasan Akut (ISPA) (111) 4. Gastroenteritis Akut (111) 5. Demam Berdarah Dengue (DBD) (111)
Cluster 2	295	41.6 thn	Jumat (sekitar jam 15:00)	1. Kontrol Rutin (295)	1. Kontrol Pasca Operasi (69) 2. Pemeriksaan Tahunan (59) 3. Vaksinasi Dewasa (57) 4. Skrining Kolesterol (55) 5. Konsultasi Gizi (55)
Cluster 3	446	66.4 thn	Kamis (sekitar jam 8:00)	1. Kronis (446)	1. Penyakit Paru Obstruktif Kronis (PPOK) (105) 2. Diabetes Melitus Tipe 2 (95) 3. Gagal Jantung Kongestif (84) 4. Hipertensi Esensial (84) 5. Gagal Ginjal Kronis (78)
Cluster 4	478	66.9 thn	Jumat (sekitar jam 13:00)	1. Kronis (478)	1. Gagal Ginjal Kronis (109) 2. Gagal Jantung Kongestif

					(102) 3. Penyakit Paru Obstruktif Kronis (PPOK) (92) 4. Diabetes Melitus Tipe 2 (89) 5. Hipertensi Esensial (86)
Cluster 5	608	22.8 thn	Kamis (sekitar jam 8:00)	1. Akut (608)	1. Demam Berdarah Dengue (DBD) (127) 2. Tifus (126) 3. Infeksi Saluran Pernapasan Akut (ISPA) (120) 4. Influenza (118) 5. Gastroenteritis Akut (117)
Cluster 6	148	45.7 thn	Minggu (sekitar jam 22:00)	1. Gawat Darurat (148)	1. Serangan Jantung (Infark Miokard) (38) 2. Stroke Iskemik (35) 3. Apendisitis Akut (26) 4. Serangan Asma Berat (25) 5. Kecelakaan Lalu Lintas (24)
Cluster 7	295	41.4 thn	Kamis (sekitar jam 11:00)	1. Kontrol Rutin (295)	1. Skrining Kolesterol (68) 2. Konsultasi Gizi (60) 3. Kontrol Pasca Operasi (57) 4. Vaksinasi Dewasa (56) 5. Pemeriksaan Tahunan (54)

SIMPULAN

Berdasarkan Tabel 2 menunjukkan delapan klaster yang secara substantif mengerucut menjadi empat profil layanan Akut, Kronis, Kontrol Rutin, dan Gawat Darurat dengan proporsi terbesar pada kasus Akut ($\approx 39,5\%$), disusul Kronis ($\approx 30,8\%$), Kontrol Rutin ($\approx 19,7\%$), dan Gawat Darurat (10%). Setiap profil menampilkan konsistensi karakteristik demografis, ritme waktu kunjungan, dan spektrum diagnosis: Akut didominasi pasien usia muda dengan puncak akhir pekan/pagi hari kerja; Kronis terpusat pada kelompok usia lanjut dengan puncak pagi-siang hari kerja; Kontrol Rutin terakumulasi pada pertengahan jam kerja; sedangkan Gawat Darurat meningkat pada malam hari. Pola ini memberi landasan empiris untuk penjadwalan layanan yang terdiferensiasi, alokasi tenaga dan sarana yang proporsional, penyiapan logistik klinis (obat, alat, dan protokol) sesuai jam sibuk, serta perancangan intervensi preventif dan edukasi yang terarah per profil. Secara keseluruhan, temuan menegaskan pentingnya perencanaan kapasitas berbasis data dan kesiapsiagaan klinis berorientasi waktu guna meningkatkan mutu, keselamatan, dan efisiensi pelayanan.

DAFTAR PUSTAKA

- [1] Stewart, N., & Al-Khassaweneh, M. (2022). An implementation of the HDBSCAN* clustering algorithm. *Applied Sciences*, 12(5), 2405. <https://doi.org/10.3390/app12052405>
- [2] Vijayan, D., & Aziz, I. (2023). Adaptive hierarchical density-based spatial clustering algorithm for streaming applications. *Telecom*, 4(1), 1–14. <https://doi.org/10.3390/telecom4010001>
- [3] Cesario, E., Lindia, P., & Vinci, A. (2023). Detecting multi-density urban hotspots: A DBSCAN and HDBSCAN approach. *Big Data and Cognitive Computing*, 7(1), 29. <https://doi.org/10.3390/bdcc7010029>

- [4] Estro, T., Antunes, M., Bhandari, P., Gandhi, A., Kuenning, G., Liu, Y., Waldspurger, C., Wildani, A., & Zadok, E. (2024). Accelerating multi-tier storage cache simulations using knee detection. *Performance Evaluation*, 164, 102410. <https://doi.org/10.1016/j.peva.2024.102410>
- [5] Hanafi, N., & Saadatfar, H. (2022). A fast DBSCAN algorithm for big data based on efficient density calculation. *Expert Systems with Applications*, 203, 117501. <https://doi.org/10.1016/j.eswa.2022.117501>
- [6] Ma, B., Yang, C., Li, A., Chi, Y., & Chen, L. (2023). A faster DBSCAN algorithm based on self-adaptive determination of parameters. *Procedia Computer Science*, 221, 2901–2910. <https://doi.org/10.1016/j.procs.2023.07.017>
- [7] Fok, T. Y., & Ye, N. (2025). Kneeliverse: A universal knee-detection library. *SoftwareX*, 26, 102161. <https://doi.org/10.1016/j.softx.2025.102161>
- [8] (2025). GraphDBSCAN: Optimized DBSCAN for noise-resistant community detection in graph clustering. *Future Internet*, 17(4), 150. <https://doi.org/10.3390/fi17040150>
- [9] (2025). Scalable clustering of complex ECG health data: Big data clustering analysis with UMAP and HDBSCAN. *Computation*, 13(6), 144. <https://doi.org/10.3390/computation13060144>
- [10] (2023). Improvement of DBSCAN algorithm based on K-Dist graph for adaptive determining parameters. *Electronics*, 12(15), 3213. <https://doi.org/10.3390/electronics12153213>
- [11] (2023). Fast multidimensional scaling on big geospatial data using neural networks and DBScan clustering. *Earth Science Informatics*, 16, 4015–4031. <https://doi.org/10.1007/s12145-023-01004-9>
- [12] Blanco-Portals, J., Peiró, F., & Estradé, S. (2022). Strategies for EELS data analysis: Introducing UMAP and HDBSCAN for dimensionality reduction and clustering. *Microscopy and Microanalysis*, 28(S1), 109–122. <https://doi.org/10.1017/S1431927621013696>
- [13] Wongoutong, C. (2024). The impact of neglecting feature scaling in k-means clustering. *PLOS ONE*, 19(12), e0310839. <https://doi.org/10.1371/journal.pone.0310839>
- [14] Zhao, Z., Zhou, W., Liang, D., Liu, W., Lee, X., et al. (2024). Denoising method based on improved DBSCAN for LiDAR point cloud. *IEEE Access*, 12, 3464611. <https://doi.org/10.1109/ACCESS.2024.3464611>
- [15] Wang, Z., Shi, H., Sun, Y., Liu, H., & Chen, X. (2025). Application of three-dimensional HDBSCAN for identifying abnormal working areas of ships based on fish migration. *Applied Sciences*, 15(5), 2621. <https://doi.org/10.3390/app15052621>
- [16] (2024). Constrained density-based spatial clustering of applications with noise and multiple objectives. *Knowledge-Based Systems*, 296, 111070. <https://doi.org/10.1016/j.knosys.2024.111070>
- [17] Khairina, N., Sibarani, T. T. S., Muliono, R., Sembiring, Z., & Muhathir, M. (2022). Identification of pneumonia using the K-nearest neighbors method using HOG feature extraction. *JITE (Journal of Informatics and Telecommunication Engineering)*, 5(2), 563–569. <https://doi.org/10.31289/jite.v5i2.6216>
- [18] Alwi, B., & Muliono, R. (2025). Analisis Klustering Menggunakan Algoritma DBSCAN untuk Deteksi Anomali dalam Data Transaksi Keuangan. *INCODING: Journal of Informatic and Computer Science Engineering*, 3(2). <https://doi.org/10.34007/incoding.v5i1.827>