

Perbandingan K-Means, DBSCAN, dan OPTICS untuk Klasterisasi Pasien Anemia Berdasarkan Parameter Hematologi

Amandasari Dinda Rabbani, Kartika Maulida Hindrayani, Muhammad Nasrudin

(Sains Data, Universitas Pembangunan Negeri Veteran Jawa Timur)

22083010093@student.upnjatim.ac.id

kartika.maulida.ds@upnjatim.ac.id

nasrudin.fasikom@upnjatim.ac.id

Corresponding author email: amanadin17@gmail.com

ABSTRAK

Anemia adalah penyakit akibat kekurangan sel darah merah yang masih menjadi masalah kesehatan global dengan prevalensi tinggi serta berdampak signifikansi terhadap kualitas hidup pasiennya. Variasi kondisi pasien anemia sering sekali disebabkan oleh faktor hematologi yang kompleks. Penelitian ini bertujuan untuk membandingkan ketiga metode klasterisasi, yaitu K-Means, DBSCAN, dan OPTICS, dalam mengelompokkan pasien anemia berdasarkan parameter hematologi. Data yang digunakan berasal dari data pasien anemia di salah satu rumah sakit di Surabaya pada tahun 2024 dengan variabel kode pasien, jenis kelamin, umur, hemoglobin, hematokrit, eritrosit dan leukosit. Lalu diolah melalui tahap pra-pemrosesan, normalisasi, dan dilakukan evaluasi hasil klasterisasi dengan metrik Silhouette Score, davies-Bouldin Index dan Calinski-Harabasz Score. Hasil menunjukkan bahwa metode OPTICS menghasilkan kualitas terbaik daripada kedua metode tersebut dengan hasilnya yaitu silhouette score = 0,23, DBI = 1,22, dan CH = 62,7. Dengan temuan ini menunjukkan bahwa OPTICS mampu menangkap pola distribusi data yang kompleks dibandingkan dengan kedua metode tersebut. Penelitian ini diharapkan dapat memberikan kontribusi dalam penerapan metode klasterisasi pada data medis, khususnya untuk membantu identifikasi kelompok pasien anemia berdasarkan karakteristik hematologinya secara lebih akurat dan informatif.

Kata Kunci: Anemia, DBSCAN, Hematologi, K-Means, OPTICS

I. PENDAHULUAN (10PT)

Anemia adalah salah satu penyakit yang memiliki masalah kesehatan global yang hingga kini masih menjadi perhatian yang mendalam, terutama di negara berkembang. Penyakit ini dilihat dari kondisi kadar hemoglobin dalam darah yang sangat rendah sehingga menghambat proses jalannya oksigen ke seluruh jaringan tubuh manusia[1]. Menurut laporan dari *World Health Organization (2019)*, 30% (571 juta) pada perempuan usia 15-49 tahun, 37% (32 juta) pada ibu hamil, dan 40% (269 juta) pada anak usia 6-59 bulan menderita anemia dengan wilayah Afrika dan Asia Tenggara yang terkena penyakit anemia[2]. Variasi dari tingkat keparahan anemia pada pasien yang terkena penyakit tersebut disebabkan dari faktor hematologisnya seperti kadar hemoglobin (Hb), hematokrit (HCT), eritrosit (RBC), dan leukosit (WBC) yang saling berkaitan. Data hematologi yang kompleks akan membuat identifikasi pola dan pengelompokan pasien menjadi tantangan tersendiri dalam analisis klinis.

Seiring berkembangnya ilmu data dan machine learning, pendekatan berbasis unsupervised learning seperti klasterisasi mulai banyak digunakan untuk mengelompokkan pasien berdasarkan kemiripan parameter fisiologis atau laboratorium.

Klasterisasi memungkinkan analisis yang lebih mendalam terhadap karakteristik pasien tanpa memerlukan label atau kategori yang telah diketahui sebelumnya. Penelitian oleh Qasrawi et al. (2022) menunjukkan bahwa metode K-Means efektif dalam mengidentifikasi pola nutrisi dan kadar hemoglobin yang berkaitan dengan risiko anemia pada populasi anak sekolah. Hasil penelitian tersebut memperlihatkan bahwa klasterisasi dapat membantu menemukan kelompok individu dengan pola hematologis dan gizi yang serupa, yang sebelumnya sulit terlihat dengan analisis konvensional[3].

Meskipun K-Means banyak digunakan karena kesederhanaan dan efisiensinya, metode ini memiliki keterbatasan ketika diterapkan pada data medis yang kompleks dan mengandung outlier. Untuk mengatasi hal tersebut, algoritma berbasis kepadatan seperti DBSCAN (Density-Based Spatial Clustering of Applications with Noise) diperkenalkan. DBSCAN mampu mendeteksi klaster dengan bentuk tidak beraturan serta mengidentifikasi data noise yang tidak termasuk dalam klaster mana pun[5]. Dalam penelitian oleh Mustikasari & Salman (2023) DBSCAN menunjukkan kinerja yang lebih baik dibanding K-Means ketika diterapkan pada data dengan variasi densitas tinggi. Studi ini menegaskan bahwa pendekatan berbasis kepadatan lebih adaptif untuk data klinis yang tidak berdistribusi normal, seperti data hematologi pasien anemia[4].

Selain DBSCAN, metode OPTICS (Ordering Points To Identify the Clustering Structure) dikembangkan untuk memperbaiki kelemahan DBSCAN, khususnya dalam penentuan parameter radius (eps). OPTICS memungkinkan identifikasi klaster dengan berbagai tingkat kepadatan dalam satu dataset[7]. Penelitian oleh Hajihosseini et al. (2024) di *Pattern Recognition Letters* melakukan evaluasi komprehensif terhadap OPTICS, GMM, dan K-Means, dan menemukan bahwa OPTICS lebih stabil terhadap variasi densitas serta lebih mampu mendeteksi struktur data kompleks. Temuan ini penting untuk konteks data hematologi, karena distribusi parameter darah pasien anemia sering kali bersifat heterogen dan tidak homogen secara statistik[6].

Lebih lanjut, penelitian Setiyawan et al. (2025) membandingkan metode DBSCAN, OPTICS, dan Agglomerative Clustering dalam pemetaan distribusi penyakit di wilayah Banjarnegara. Hasilnya menunjukkan bahwa OPTICS lebih unggul dalam mendeteksi klaster penyakit dengan tingkat kepadatan berbeda dan dalam mengidentifikasi area dengan persebaran tidak merata. Studi tersebut memberikan dasar metodologis yang kuat bagi penelitian ini untuk menerapkan pendekatan serupa pada data hematologi pasien anemia yang memiliki karakteristik non-linier dan kompleks[7].

Berdasarkan berbagai hasil penelitian tersebut, dapat disimpulkan bahwa setiap algoritma klasterisasi memiliki keunggulan dan kelemahan masing-masing. K-Means unggul dalam efisiensi dan interpretasi sederhana, namun kurang tangguh terhadap outlier dan bentuk klaster non-linier. DBSCAN dapat mengidentifikasi klaster dengan bentuk tak beraturan, tetapi sensitif terhadap pemilihan parameter. Sementara itu, OPTICS menawarkan solusi yang lebih fleksibel untuk mendeteksi klaster dengan variasi densitas dan kompleksitas tinggi. Oleh karena itu, penelitian ini berfokus untuk membandingkan performa ketiga algoritma tersebut (K-Means, DBSCAN, dan OPTICS) dalam mengelompokkan pasien anemia berdasarkan parameter hematologi, dengan tujuan menemukan metode yang paling sesuai dalam menangkap pola distribusi data medis yang kompleks serta memberikan interpretasi yang relevan secara klinis. berikan jurnalnya juga

II. METODOLOGI PENELITIAN (10 PT)

II.1 Variabel Penelitian

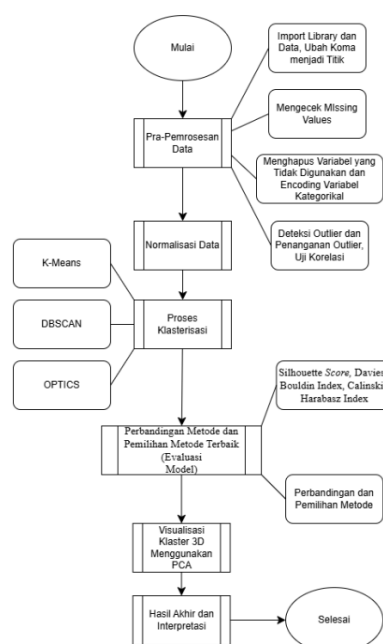
Dalam penelitian ini data yang digunakan dari salah satu rumah sakit di daerah Surabaya yang bernama Rumah Sakit Umum Daerah Haji, Provinsi Jawa Timur

pada tahun 2024. Analisis kluster ini akan dilakukan menggunakan beberapa variabel yang terdapat pada parameter hematologi dan variabel lainnya seperti jenis kelamin, umur, hemoglobin (g/dl), lekosit (mm³), hematokrit (%), dan trombosit (mm³). Dan tabel berikut ini memberikan daftar variabel yang digunakan beserta tipe datanya dan satuannya:

Tabel 2.1 Variabel Dataset

Variabel	Nama Variabel	Keterangan	Satuan	Tipe Data
X_1	Jenis Kelamin	Menunjukkan jenis kelamin pasien, yaitu laki-laki atau perempuan berfungsi untuk melihat perbedaan karakteristik pasien berdasarkan gender	Tidak Ada	Katagorikal
X_2	Umur	Usia pasien pada saat dirawat di rumah sakit berfungsi untuk melihat dapat memengaruhi kondisi anemia dan respons terhadap perawatan	Tahun	Numerik
X_3	Hemoglobin	Protein dalam sel darah merah yang berfungsi mengikat dan mengangkut oksigen ke seluruh tubuh	g/dl	Numerik
X_4	Lekosit	Jumlah sel darah putih, berfungsi mencerminkan kondisi sistem imun pasien dan kemungkinan infeksi	mm ³	Numerik
X_5	Hematokrit	Persentase volume sel darah merah dari total volume darah, berfungsi mencerminkan kapasitas pengangkutan oksigen	%	Numerik
X_6	Trombosit	Jumlah sel darah kecil yang berperan dalam proses pembekuan darah untuk menghentikan perdarahan	mm ³	Numerik

II.2 Analisis Kluster



Gambar 2.1 Diagram Alir Penelitian

Analisis kluster merupakan metode unsupervised learning yang bertujuan untuk mengelompokkan sekumpulan objek berdasarkan tingkat kemiripan karakteristiknya, sehingga objek yang berada dalam satu kelompok memiliki pola yang relatif serupa dibandingkan dengan kelompok lainnya[6]. Dalam penelitian ini, analisis kluster diterapkan untuk mengelompokkan pasien anemia berdasarkan parameter hematologi seperti kadar hemoglobin, hematokrit, leukosit, trombosit, serta variabel demografis seperti umur dan jenis kelamin. Proses ini dilakukan dengan bahasa pemrograman *Python* menggunakan pustaka *scikit-learn*, *NumPy*, *Pandas*, dan *Matplotlib*, yang memungkinkan analisis numerik dan visualisasi data secara efisien. Berikut tahapan untuk melakukan analisis kluster yaitu:

1. Pra-Pemrosesan Data

Tahap awal penelitian adalah pra-pemrosesan data yang bertujuan menyiapkan dataset agar siap digunakan dalam analisis kluster. Data pasien yang diambil dari sumber sekunder dibersihkan dari nilai yang tidak konsisten, contohnya seperti format desimal yang menggunakan koma diganti menjadi titik agar dapat diproses sebagai numerik. Mengecek *missing values* setiap variabelnya. Kolom kode pasien dihapus sementara agar tidak ikut di analisis klasternya, lalu menghapus kolom yang tidak relevan seperti kondisi pulang, rawat inap atau rawat jalan dan lama rawat (hari) dan variabel kategorikal seperti jenis kelamin diubah menjadi bentuk numerik melalui Label Encoding. Langkah ini penting agar seluruh variabel dapat diproses oleh algoritma statistik dan tidak menyebabkan kesalahan dalam perhitungan jarak antar objek. Kemudian dilakukan deteksi dan penanganan outlier untuk melihat apakah data ini sudah bagus untuk dilakukan proses klasterisasi atau tidak, jika ada outlier lalu dibuat penanganan outlier berupa winsorizing kemudian jika sudah dilakukan penanganan outlier lalu dilakukan uji korelasinya agar bisa melihat variabel mana yang bisa dilakukan ketahap selanjutnya.

2. Normalisasi Data

Tahapan berikutnya adalah normalisasi data agar semua fitur memiliki skala yang sebanding. Normalisasi dilakukan dengan menggunakan metode *StandardScaler*, yang mentransformasikan data menjadi distribusi dengan rata-rata nol dan simpangan baku satu:

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

Keterangan:

x : nilai asli

μ : rata-rata

σ : simpangan baku

Langkah ini memastikan bahwa variabel seperti kadar hemoglobin dan jumlah leukosit memiliki kontribusi yang seimbang terhadap pembentukan kluster[6]. Tanpa normalisasi, fitur dengan skala besar dapat mendominasi hasil perhitungan jarak Euclidean.

3. Proses Klasterisasi

Tahap inti dari penelitian ini adalah pembentukan kluster. Tiga algoritma digunakan, yaitu K-Means, DBSCAN, dan OPTICS. Ketiganya bekerja dengan prinsip yang berbeda dan memiliki dasar matematis tersendiri.

a. K-Means

K-Means merupakan metode berbasis centroid yang membagi data menjadi k kluster dengan meminimalkan jumlah jarak kuadrat antara titik data dan pusat kluster. Fungsi objektif yang diminimalkan adalah:

$$J = \sum_{i=1}^k \sum_{x_j \in C_i} \|x_j - \mu_i\|^2 \quad (2)$$

Keterangan:

J : total variansi dalam klaster

x_j : titik data ke- j

C_i : klaster ke- i

μ_i : centroid klaster ke- i

Proses iteratif dilakukan hingga posisi centroid tidak lagi berubah secara signifikan. Nilai k optimal ditentukan dengan metode Elbow, di mana grafik *inertia* terhadap jumlah klaster membentuk titik siku yang menandakan keseimbangan antara variansi dan kompleksitas[9].

b. DBSCAN

DBSCAN membentuk klaster berdasarkan kepadatan titik data di sekitar area tertentu. Dua parameter penting digunakan yaitu ε (epsilon) untuk menentukan radius pencarian maksimum dari suatu titik, dan minPts untuk menghitung jumlah minimum titik yang harus ada dalam radius ε . Rumus jarak antar-titik ditentukan dengan jarak Euclidean:

$$\text{dist}(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (3)$$

Jika jumlah titik dalam radius memenuhi akan seperti ini $|N_\varepsilon(p)| \geq \text{minPts}$

Keterangan:

$\text{dist}(p, q)$: jarak antara dua titik p dan q ,

$N_\varepsilon(p)$: himpunan titik dalam radius epsilon dari p .

maka titik p disebut core point dan membentuk klaster dengan semua titik disekitarnya. Titik yang tidak memiliki cukup tetangga disebut noise dan tidak masuk ke dalam klaster manapun[8]. Metode ini mampu mendeteksi klaster berbentuk arbitrer dan lebih tahan terhadap *outlier* dibandingkan K-Means.

c. OPTICS

OPTICS merupakan pengembangan dari DBSCAN yang mampu menangani data dengan variasi kepadatan berbeda. OPTICS menghasilkan urutan titik berdasarkan reachability distance, yang menunjukkan seberapa mudah suatu titik dijangkau oleh titik lain dalam konteks kepadatan lokal. Rumus yang digunakan adalah:

$$\text{reachability_dist}(p, o) = \max(\text{core_dist}(p), \text{dist}(p, o)) \quad (4)$$

$$\text{core_dist}(p) = \min_{\text{dist}(p, o)} \{ \text{dist}(p, o) \mid |N_\varepsilon(p)| \geq \text{minPts} \} \quad (5)$$

Keterangan:

$\text{re_dist}(p)$: jarak minimum agar titik p dianggap sebagai core point,

$\text{dist}(p, o)$: jarak Euclidean antara titik p dan o ,

$\text{reachability_dist}(p, o)$: jarak keterjangkauan antara dua titik dalam konteks kepadatan lokal.

Nilai *reachability distance* kecil menandakan titik-titik padat, sedangkan nilai tinggi menunjukkan wilayah jarang atau batas klaster [6]. OPTICS menghasilkan *reachability plot* yang memperlihatkan struktur hierarkis klaster dan membantu memilih jumlah kelompok optimal tanpa menentukan parameter ε secara tegas.

4. Perbandingan Metode dan Pemilihan Metode Terbaik (Evaluasi Model)

Setelah klaster terbentuk, hasilnya dievaluasi menggunakan tiga metrik internal, yaitu Silhouette Score, Davies–Bouldin Index (DBI), dan Calinski–Harabasz Index (CH).

a. Silhouette Score

Silhouette Score merupakan salah satu metode evaluasi dalam analisis klusterisasi yang digunakan untuk mengukur kualitas pembentukan cluster. Metode ini menilai sejauh mana sebuah objek mirip dengan anggota cluster-nya sendiri dibandingkan dengan cluster lain. Dengan demikian, Silhouette Score tidak hanya mengukur tingkat kepadatan cluster, tetapi juga keterpisahan antar cluster [11,12]. Berikut rumus untuk silhouette score sebagai berikut:

$$s(i) = \frac{b(i)-a(i)}{\max(a(i),b(i))} \quad (6)$$

Keterangan:

$a(i)$: jarak rata-rata antar-poin dalam satu klaster

$b(i)$: jarak rata-rata ke klaster terdekat

nilai $s(i)$ mendekati 1 menunjukkan kualitas klaster yang baik[

b. Davies–Bouldin Index (DBI)

Davies-Bouldin Index (DBI) merupakan salah satu metode evaluasi internal dalam klusterisasi yang digunakan untuk menilai kualitas hasil pengelompokan data. DBI menilai seberapa baik cluster yang terbentuk dengan mempertimbangkan kompaksi (compactness) dalam satu cluster dan separasi (separation) antar cluster[14]. Berikut rumus untuk Davies-Bouldin Index sebagai berikut:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \frac{S_i + S_j}{d_{ij}} \quad (7)$$

Keterangan:

S_i : rata-rata jarak antar-anggota dalam klaster i

d_{ij} : jarak antar-centroid klaster i dan j

c. Calinski–Harabasz Index (CH)

$$CH = \frac{Tr(B_k)/(k-1)}{Tr(W_k)/(n-k)} \quad (8)$$

Keterangan:

B_k : variansi antar-klaster

W_k : variansi dalam klaster

Nilai CH tinggi menunjukkan struktur klaster yang optimal[8].

Setelah mendapatkan hasilnya kemudian dibandingkan ketiga metode tersebut lalu dipilih mana metode terbaiknya.

5. Visualisasi Klaster 3D Menggunakan PCA

Setelah klusterisasi, dilakukan Principal Component Analysis (PCA) untuk menurunkan dimensi data menjadi tiga komponen utama dan memvisualisasikan hasil klaster dalam ruang 3D. Transformasi PCA dinyatakan sebagai:

$$Z=XW \quad (9)$$

Keterangan:

Z : matriks hasil proyeksi data baru,

X : data terstandarisasi

W : matriks eigenvector dari matriks kovarians X

Visualisasi 3D memperlihatkan distribusi tiap klaster dengan warna berbeda, yang menunjukkan keterpisahan kelompok pasien anemia berat, sedang, dan ringan[9].

6. Hasil Akhir dan Interpretasinya

Berdasarkan perhitungan menggunakan data penelitian ini, berikut hasil evaluasi

klaster dari metode K-Means, DBSCAN dan OPTICS.

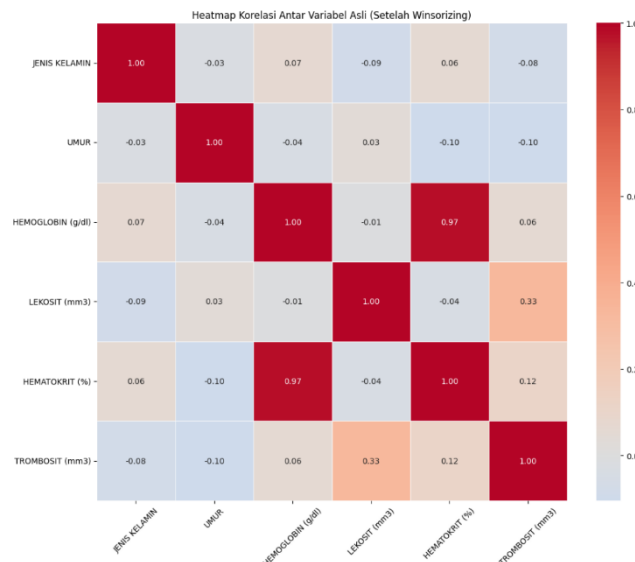
Tabel 2.2 Evaluasi Klaster

Metode	Silhouette Score	Davies–Bouldin Index	Calinski–Harabasz Index
K-Means	0.201455	1.623474	101.822232
DBSCAN	0.190627	1.241163	39.477873
OPTICS	0.230618	1.222243	62.729296

III. HASIL DAN ANALISIS (10 PT)

Pada tahap ini adalah tahap hasil dan analisis untuk menyajikan hasil pengolahan data beserta interpretasi yang didapatkan pada proses analisis. Data pasien penyakit anemia yang sudah dikumpulkan dan dilakukan pengolahan menggunakan 3 metode klasterisasi untuk melakukan pasien anemia berdasarkan parameter hematologi. Tahapan yang dilakukan yaitu pra-pemrosesan data, normalisasi data, proses klasterisasi, perbandingan metode dan pemilihan metode terbaik, visualisasi klaster 3D menggunakan PCA dan hasil akhir beserta interpretasi.

Pada tahapan pra-pemrosesan data dilakukan pengubahan tanda koma menjadi titik pada kolom hemoglobin dan hematokrit lalu dicek missing value setiap variabelnya, kemudian dihapus kolom yang tidak digunakan seperti kondisi pulang, rawat inap atau rawat jalan, dan lama rawat kemudian kolom yang tipenya kategorikal di encoding agar menjadi tipe numerik kemudian di deteksi outlier dan pada variabel trombosit terdapat 7 outlier, hemoglobin terdapat 3 outlier, hematokrit terdapat 3 outlier dan lekosit terdapat 28 outlier, setelah mendapatkan hasil outlier kemudian ditanganin outlier ini menggunakan winsorizing dimana hasilnya sudah dibatasi dengan cara nilai-nilai yang terlalu kecil atau besar ke batas tertentu.



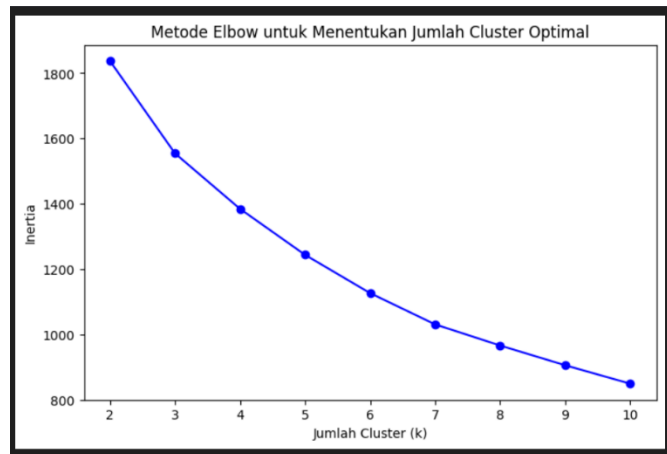
Gambar 3.1 Heatmap Uji Korelasi

Selanjutnya ke uji korelasi berdasarkan hasil heatmap dan tabel korelasi setelah proses *winsorizing*, sebagian besar variabel menunjukkan tidak ada hubungan yang signifikan satu sama lain (nilai korelasi mendekati 0). Hanya terdapat satu korelasi kuat, yaitu antara HEMOGLOBIN (g/dl) dan HEMATOKRIT (%) dengan $r = 0.97$, yang menandakan hubungan positif sangat erat dan wajar secara fisiologis karena keduanya sama-sama mencerminkan kadar sel darah merah. Selain itu, terdapat korelasi lemah

positif antara LEKOSIT (mm^3) dan TROMBOSIT (mm^3) dengan $r = 0.33$, sedangkan variabel lain seperti UMUR dan JENIS KELAMIN tidak berkorelasi berarti dengan variabel darah. Secara keseluruhan, dataset ini memiliki multikolinearitas rendah, sehingga cocok untuk analisis lanjutan seperti *clustering*.

III.1 Menentukan Jumlah Kluster Optimal Untuk Metode K-Means

Setelah melakukan uji korelasi dilanjutkan ke tahapan normalisasi data lalu dilanjut ke proses klasterisasi berikut proses klasterisasi untuk menentukan kluster optimal pada metode K-Means, berikut gambar dari metode elbow.



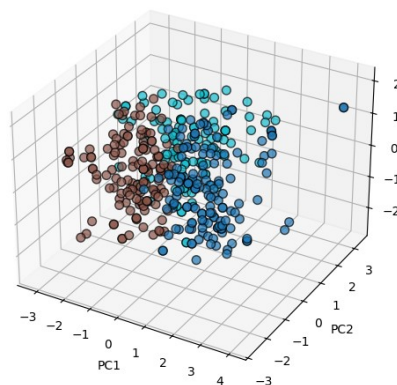
Gambar 3.2 Metode Elbow

Berdasarkan grafik metode Elbow, penurunan nilai *inertia* paling tajam terjadi hingga $k = 3$, kemudian mulai melandai. Hal ini menunjukkan bahwa jumlah cluster optimal berada pada $k = 3$, karena setelah titik tersebut penambahan cluster tidak memberikan peningkatan signifikan dalam kualitas pemodelan.

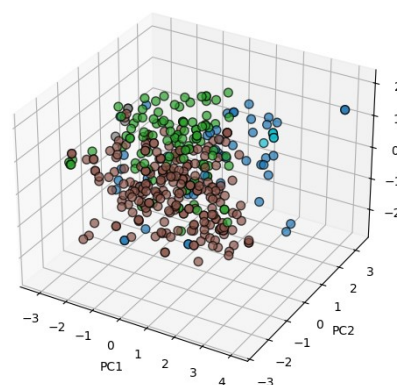
III.2 Visualisasi Kluster

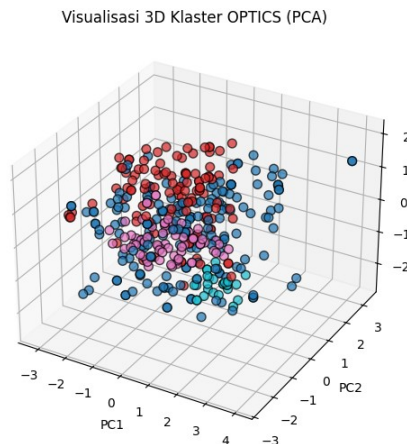
Pada penelitian ini dilakukan proses klasterisasi terhadap data pasien anemia dengan menggunakan tiga metode, yaitu K-Means, DBSCAN, dan OPTICS. Tujuan dari proses klasterisasi ini adalah untuk mengelompokkan pasien berdasarkan kesamaan karakteristik yang dimiliki, sehingga dapat diketahui pola distribusi pasien anemia secara lebih terstruktur. Hasil visualisasi klasterisasi tiga dimensi (3D) menggunakan metode Principal Component Analysis (PCA) menunjukkan perbedaan pola pembentukan kluster pada masing-masing algoritma. Berikut hasil gambar visualisasi klasternya setiap 3 metode.

Visualisasi 3D Kluster K-Means (PCA)



Visualisasi 3D Kluster DBSCAN (PCA)





Gambar 3.3 Visualisasi Hasil Kluster dari 3 Metode

Pada metode K-Means, terbentuk tiga kluster utama dengan jumlah anggota yang relatif seimbang, yaitu kluster 0 sebanyak 147 pasien, kluster 1 sebanyak 137 pasien, dan kluster 2 sebanyak 110 pasien. Hasil visualisasi memperlihatkan bahwa ketiga kluster tersebut cukup terpisah, meskipun masih terdapat sedikit tumpang tindih antar kelompok. Kondisi ini menunjukkan bahwa metode K-Means mampu membentuk kluster dengan baik, namun batas antar kluster belum sepenuhnya tegas.

Sementara itu, metode DBSCAN menghasilkan pembagian kluster yang tidak merata, dengan kluster 1 sebanyak 229 pasien, kluster 0 sebanyak 105 pasien, kluster 2 sebanyak 4 pasien, kluster 3 sebanyak 6 pasien, serta 50 data yang dikategorikan sebagai *noise*. Hal ini menunjukkan bahwa metode DBSCAN cukup sensitif terhadap parameter yang digunakan, sehingga masih terdapat sejumlah data yang tidak dapat dikelompokkan ke dalam kluster mana pun.

Adapun metode OPTICS menghasilkan struktur kluster yang lebih alami dengan distribusi yang bervariasi, yaitu kluster 0 sebanyak 111 pasien, kluster 1 sebanyak 74 pasien, dan kluster 2 sebanyak 27 pasien, serta 182 data yang terdeteksi sebagai *noise*. Meskipun jumlah *noise* pada metode ini cukup tinggi, hasil visualisasi memperlihatkan bahwa OPTICS mampu mengidentifikasi kluster dengan kepadatan berbeda serta menangkap struktur data yang lebih kompleks dibandingkan dua metode lainnya.

III.3 Perbandingan Kualitas Kluster

Tahapan ini dilakukan untuk membandingkan metode klusterisasi dan menampilkan metode mana yang terbaik, seperti bisa dilihat pada tabel 2.2 terdapat evaluasi kluster dan selanjutnya dilakukan pemilihan metode yang terbaik, seperti berikut ini.

Tabel 3.1 Ranking Metode Clustering Berdasarkan Semua Metrik

Metode	Rank	Silhouette	Rank DBI	Rank CH	Total Rank
K-Means		2.0	3.0	1.0	6.0
DBSCAN		3.0	2.0	3.0	8.0
OPTICS		1.0	1.0	2.0	4.0

Tabel 3.2 Model Terbaik

Metode	OPTICS
Silhouette	0.230618
Davies-Bouldin	1.222243
Calinski-Harabasz	62.729296

Rank Silhouette	1.0
Rank DBI	1.0
Rank CH	2.0
Total Rank	4.0

Berdasarkan hasil evaluasi performa dari ketiga metode klusterisasi yang digunakan, yaitu K-Means, DBSCAN, dan OPTICS, dilakukan perbandingan menggunakan tiga metrik evaluasi utama, yaitu Silhouette Score, Davies-Bouldin Index, dan Calinski-Harabasz Index. Hasil pengujian menunjukkan bahwa metode OPTICS memperoleh nilai Silhouette Score tertinggi (0.230618), Davies-Bouldin Index terendah (1.222243), serta nilai Calinski-Harabasz Index sebesar 62.729296, yang mengindikasikan kualitas kluster yang baik. Selain itu, berdasarkan hasil perankingan dari ketiga metrik tersebut, metode OPTICS memiliki total peringkat terbaik dengan nilai Total Rank = 4, dibandingkan K-Means = 6, dan DBSCAN = 8.

Berdasarkan hasil evaluasi tersebut serta didukung oleh visualisasi yang dihasilkan, dapat disimpulkan bahwa metode OPTICS merupakan metode yang paling baik digunakan untuk mengelompokkan pasien anemia. Metode ini mampu menangani variasi kepadatan data, memisahkan pasien dengan karakteristik yang berbeda secara lebih akurat, serta mendeteksi pasien dengan pola yang tidak umum (*outlier*). Dengan demikian, OPTICS memberikan hasil kluster yang paling representatif terhadap kondisi nyata distribusi pasien anemia dan dapat menjadi metode yang efektif dalam analisis data medis untuk mendukung pengambilan keputusan terkait penanganan pasien anemia.

III.4 Hasil Akhir Kluster

Pada tahap ini adalah tahapan terakhir dari penelitian ini, tahapan yang dilakukan yaitu hasil pengelompokan seluruh variabel dengan menggunakan 3 metode tersebut, namun saya memberikan hasil kluster yang menggunakan metode OPTICS dikarenakan metode tersebut adalah metode yang terbaik dari kedua metode K-Means dan DBSCAN. Hasil klusterisasi yang menggunakan metode OPTICS menunjukkan bahwa pasien penyakit anemia terbagi menjadi 4 kluster yang dimana kondisi atau kategorinya sama tetapi interpretasinya berbeda. Berikut adalah rincian klusterisasi menggunakan metode OPTICS yang ditunjukkan pada tabel dibawah ini.

Tabel 3.3 Hasil Akhir Kluster Metode OPTICS

Kluster	Kode Pasien dari Metode OPTICS	Rata-Rata Variabel	Total Kluster	Kategori
-1	A003, A006, A007, A009, B010, B012, B013, B015, B016, B018, B021, B022, B024, B030, B032, B033, B034, C035, C043, C044, C045, C046, C047, C052, C053, C055, C056, C057, C061, C063, C064, C072, C073, C074, C077, C081, C083, C085, C086, C087, C089, C090, C093, C094, C095, C096, C098, C099, C101, C102, C104, D106, D110, D113, D116, D118, D120, D121, D131, D135, D137, D139, D141, D144, D148, D150, D151, D153, D156, D157, D158, D159, D160, D161, D163, D164, D165, D166, D167, D168,	JENIS KELAMIN_scaled = 0.430154 UMUR_scaled = -0.048726 HEMOGLOBIN (g/dl)_scaled = 0.007870 LEKOSIT (mm3)_scaled = 0.314303 HEMATOKRIT (%)_scaled = -0.000218 TROMBOSIT (mm3)_scaled = 0.197485	182	Anemia Sedang

	D170, D171, D172, D173, D174, D175, D177, D178, D179, D181, D182, D184, D185, D187, D188, D190, D193, D194, D196, D197, D198, D206, D208, D213, D216, D218, D219, D222, D223, D224, D226, D227, D228, D230, D232, D234, D235, D237, D241, D242, D243, D244, D250, D251, D253, D255, D256, D260, D263, D265, D267, D268, D274, D277, D278, D279, D280, D282, D285, D287, D290, D291, D297, D299, D302, D303, D304, D307, D308, D317, D319, D321, D322, D326, D327, D330, D331, D333, D334, D335, D336, D339, E340, E349, E350, E352, F355, F357, F360, F362, F365, F366, F368, F369, F370, F371, F372, F373, F375, F376, F379, F381.			
0	A001, A002, A005, A008, B011, B020, B028, B029, B031, C037, C039, C041, C042, C048, C051, C054, C062, C065, C066, C067, C068, C075, C076, C078, C079, C080, D111, D115, D119, D123, D125, D126, D127, D128, D130, D136, D140, D142, D146, D149, D162, D176, D180, D183, D189, D191, D192, D199, D200, D202, D203, D205, D210, D211, D212, D217, D220, D221, D229, D231, D233, D238, D246, D249, D252, D254, D257, D259, D264, D266, D269, D270, D275, D281, D283, D284, D286, D288, D289, D294, D300, D305, D306, D311, D312, D313, D314, D315, D318, D320, D323, D337, D338, E341, E342, E343, E345, E346, E347, E351, F359, F361, F363, F367, F377, F380, F383, F384, F386, F390, F394.	JENIS KELAMIN_scaled = - 1.369640 UMUR_scaled = 0.163967 HEMOGLOBIN (g/dl)_scaled = -0.315474 LEKOSIT (mm3)_scaled = - 0.061282 HEMATOKRIT (%)_scaled = - 0.297477 TROMBOSIT (mm3)_scaled = - 0.071573	111	Anemia Berat
1	B019, B025, B027, C040, C049, C050, C059, C069, C070, C071, C082, C084, C088, C092, C097, C100, C103, D105, D107, D109, D117, D124, D129, D133, D134, D138, D143, D147, D152, D155, D169, D195, D201, D204, D207, D209, D215, D225, D236, D239, D240, D245, D247, D248, D258, D262, D271, D272, D273, D276, D296, D298, D301, D310, D316, D324, D325, D328, D329, D332.	JENIS KELAMIN_scaled = 0.730119 UMUR_scaled = 0.321964 HEMOGLOBIN (g/dl)_scaled = 0.149371 LEKOSIT (mm3)_scaled = - 0.525912 HEMATOKRIT (%)_scaled = 0.099345 TROMBOSIT (mm3)_scaled = -	74	Anemia Ringan

	E344, E348, E353, F354, F358, F364, F374, F378, F382, F387, F389, F391, F392, F393.	0.447594	
2	A004, B014, B017, B023, B026, C036, C038, C058, C060, C091, D108, D112, D114, D122, D132, D145, D154, D186, D214, D261, D292, D293, D295, D309, F356, F385, F388.	JENIS KELAMIN_scaled = 0.730119 UMUR_scaled = -1.228057 HEMOGLOBIN (g/dl)_scaled = 0.834514 LEKOSIT (mm3)_scaled = -0.425311 HEMATOKRIT (%)_scaled = 0.952155 TROMBOSIT (mm3)_scaled = 0.189785	Tidak Anemia

Berdasarkan hasil klasterisasi menggunakan metode OPTICS, diperoleh empat kelompok pasien dengan karakteristik hematologi yang berbeda, mencerminkan variasi tingkat keparahan anemia.

Klaster -1 mencakup 182 pasien dengan rata-rata kadar hemoglobin dan hematokrit mendekati nilai normal, disertai sedikit peningkatan pada jumlah leukosit. Kondisi ini menunjukkan adanya kecenderungan anemia sedang, di mana kadar eritrosit masih cukup untuk mengangkut oksigen, namun mulai menunjukkan penurunan efisiensi darah. Klaster 0 terdiri dari 111 pasien yang memiliki nilai hemoglobin dan hematokrit terendah dibanding klaster lainnya, menandakan kondisi anemia berat. Hal ini menunjukkan bahwa pasien dalam klaster ini kemungkinan mengalami penurunan signifikan dalam kapasitas pengangkutan oksigen, yang dapat berdampak pada fungsi organ vital.

Klaster 1 berisi 74 pasien dengan kadar hemoglobin dan hematokrit yang sedikit lebih tinggi dibanding klaster 0, namun masih berada di bawah rentang normal. Kondisi ini merepresentasikan anemia sedang, di mana pasien menunjukkan perbaikan relatif tetapi masih memerlukan pemantauan medis lanjutan. Sementara itu, Klaster 2 yang terdiri dari 27 pasien menunjukkan kadar hemoglobin dan hematokrit paling tinggi di antara seluruh klaster, serta kestabilan pada variabel hematologi lainnya. Oleh karena itu, klaster ini dikategorikan sebagai tidak anemia, menandakan kondisi fisiologis darah yang sehat dan normal.

Secara keseluruhan, metode OPTICS mampu mengelompokkan pasien secara efektif berdasarkan parameter hematologi, sehingga memudahkan interpretasi klinis terhadap tingkat keparahan anemia mulai dari tidak anemia hingga anemia berat.

IV. KESIMPULAN

Berdasarkan hasil penelitian mengenai perbandingan metode K-Means, DBSCAN, dan OPTICS dalam klasterisasi pasien anemia berdasarkan parameter hematologi, diperoleh bahwa metode OPTICS memberikan hasil terbaik dengan nilai Silhouette Score sebesar 0,2306, Davies–Bouldin Index sebesar 1,2222, dan Calinski–Harabasz Index sebesar 62,7293. Hasil tersebut menunjukkan bahwa metode OPTICS memiliki kualitas klasterisasi yang lebih baik dibandingkan K-Means (Silhouette = 0,2015; DBI = 1,6234; CH = 101,8222) dan DBSCAN (Silhouette = 0,1906; DBI = 1,2411; CH = 39,4778).

Metode OPTICS juga mampu membentuk kluster dengan distribusi dan kepadatan data yang lebih bervariasi, sehingga dapat memisahkan pasien berdasarkan tingkat keparahan anemia dengan lebih akurat. Berdasarkan hasil analisis, pasien dikelompokkan menjadi beberapa kategori, yaitu tidak anemia, anemia ringan, anemia sedang, dan anemia berat. Hal ini menunjukkan bahwa metode OPTICS efektif dalam mengidentifikasi pola klinis pasien anemia dan berpotensi menjadi alat bantu analisis dalam mendukung diagnosis awal berdasarkan data hematologi.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Rumah Sakit Umum Daerah Haji Provinsi Jawa Timur atas dukungan dan penyediaan data penelitian, serta kepada dosen pembimbing dan Program Studi Sains Data, Universitas Pembangunan Nasional Veteran Jawa Timur, yang telah memberikan arahan, bimbingan dan dukungan selama proses penelitian ini berlangsung. Dan saya juga berterimakasih kepada tim Discover yang telah memberikan kesempatan, dukungan, serta fasilitas dalam penyusunan penelitian ini.

REFERENSI

- [1] W. M. Gardner et al., "Prevalence, years lived with disability, and trends in anaemia burden by severity and cause, 1990–2021: findings from the Global Burden of Disease Study 2021," *Lancet Haematol.*, vol. 10, no. 9, pp. e713–e734, 2023.
- [2] World Health Organization, *Guideline on haemoglobin cutoffs to define anaemia in individuals and populations*. Geneva: World Health Organization, 2024. ISBN 978-92-4-008854-2. Available: <https://iris.who.int/handle/10665/379899>
- [3] R. Qasrawi and D. Abu Al-Halawa, "Cluster Analysis and Classification Model of Nutritional Anemia Associated Risk Factors Among Palestinian Schoolchildren, 2014," *Front. Nutr.*, vol. 9, no. May, pp. 1–11, 2022.
- [4] M. Mustikasari and N. Salman, "Density-Based Clustering Analysis with DBSCAN and OPTICS," *Jurnal INSYPRO (Information System and Processing)*, vol. 8, no. 1, pp. 1–8, 2023, doi: 10.24252/insypro.v8i1.36347.
- [5] A. R. Pratama, B. S. Dewantara, D. M. Sari, and D. Pramadihanto, "Improvement of DBSCAN Algorithm Involving Automatic Parameter Estimation and Curvature Analysis in 3D Point Cloud of Piled Pipe," *Journal of Image and Graphics*, vol. 12, no. 2, pp. 175–185, 2024, doi: 10.18178/joig.12.2.175-185.
- [6] M. Hajihosseini, A. Maghsoudi, and R. Ghezlbash, "A comprehensive evaluation of OPTICS, GMM and K-means clustering methodologies for geochemical anomaly detection connected with sample catchment basins," *Geochemistry*, vol. 84, no. 2, p. 126094, 2024.
- [7] D. T. Setiawan, B. Berlilana, and A. S. Barkah, "Comparative Analysis of DBSCAN, OPTICS, and Agglomerative Clustering Methods for Identifying Disease Distribution Patterns in Banjarnegara Community Health Centers," *J. Tek. Inform.*, vol. 6, no. 3, pp. 1229–1240, 2025.
- [8] S. H. Hastuti, A. Septiani, H. Hendrayani, and W. P. Nurmawati, "Penerapan Metode OPTICS dan ST-DBSCAN untuk Klasterisasi Data Kesehatan," *Edumatic J. Pendidik. Inform.*, vol. 8, no. 1, pp. 252–261, 2024.
- [9] A. A. Wani, "Comprehensive analysis of clustering algorithms: exploring

- limitations and innovative solutions,” *PeerJ Computer Science*, vol. 10, p. e2286, Aug. 2024, doi: [10.7717/peerj-cs.2286](https://doi.org/10.7717/peerj-cs.2286)
- [10] M. A. Ahmed, H. Baharin, and P. N. E. Nohuddin, “Analysis of K-means, DBSCAN and OPTICS cluster algorithms on Al-Quran verses,” *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 11, no. 8, pp. 248–253, 2020, doi: 10.14569/IJACSA.2020.0110831.
- [11] Muhammad Raqib Syahkur, D. Hartama, and S. Solikhun, “Evaluasi Jumlah Cluster pada Algoritma K-Means++ Menggunakan Silhouette dan Elbow dengan Validasi Nilai DBI dalam Mengelompokkan Gizi Balita,” *JST (Jurnal Sains dan Teknol.)*, vol. 13, no. 3, pp. 487–496, 2024.
- [12] J. M. Guntur, “Algoritma K-Means untuk Meningkatkan Silhouette Score pada Pengelompokan Data Stok Bahan Manufaktur di PT. XYZ Kabupaten Majalengka,” *Multinetics*, vol. 11, no. 1, pp. 11–21, 2025.
- [13] Fathurrahman, H. Sri, and R. Kusumawati, “EVALUASI CLUSTERING K-MEANS DAN K-MEDOID PADA PERSEBARAN COVID-19 DI INDONESIA DENGAN METODE DAVIES-BOULDIN INDEX (DBI),” *J. Mnemon.*, vol. 6, no. 2, pp. 117–128, 2023.