SPEKTRUM AKADEMIKA
Volume. 1, Nomor. 2, 2026

SEGMENTASI PROFIL BIOMARKER MEDIS DAN TINGKAT KEAKTIFAN PENDONOR DARAH DI UDD PMI KOTA SURABAYA MENGGUNAKAN K-MEANS CLUSTERING DAN PCA

Khairunisa Olive Ektha[1], Nada Amalia Choiro[2], Kartika Maulida Hindrayani[3], Aviolla Terza Damaliana[4]

[1,2,3,4] Program Studi Sains Data, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jawa Timur

24083010120@student.upnjatim.ac.id[1], 24083010066@student.upnjatim.ac.id[2], kartika.maulida.sada@upnjatim.ac.id[3], aviolla.terza.sada@upnjatim.ac.id[4]

Abstract

Maintaining a stable and safe blood supply requires comprehensive insight into both the physiological characteristics and donation retention patterns of active blood donors. At the Blood Transfusion Unit of the Indonesian Red Cross in Surabaya City, donation records have primarily been summarized through basic demographic counts without integrating clinical screening indicators and repeat donation behavior into unified donor profiles. This study aims to segment active blood donors based on medical screening biomarkers and donation frequency to support targeted donor retention strategies. Using observational transaction records collected from January to October 2026, clinical data cleaning and donor-level aggregation transformed 136,031 initial records into 134,980 validated donation transactions representing 78,074 unique donors. Seven standardized variables comprising average age, donation frequency, hemoglobin level, systolic blood pressure, diastolic blood pressure, pulse rate, and body weight were clustered using the k-means algorithm, validated via silhouette evaluation and the non-parametric Kruskal-Wallis test, and projected using principal component analysis. Four distinct donor segments emerged with an optimal silhouette score of 0.2036 and statistically significant differences across all indicators: young adults with lighter body weight representing 47.31 percent of donors with low donation frequency, mature routine donors representing 19.39 percent with the highest average frequency of 3.54 donations within ten months, middle-aged donors with higher blood pressure and body weight representing 32.53 percent, and a minor elevated-pulse group representing 0.77 percent. These findings provide an empirical foundation for automated post-donation reminders aligned with the mandatory sixty-one-day medical recovery interval.

Keyword: blood donor segmentation; k-means clustering; medical biomarkers; principal component analysis; retention analytics

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Abstrak

Menjaga ketersediaan stok darah yang aman dan berkelanjutan memerlukan pemahaman komprehensif terhadap karakteristik fisiologis serta pola keaktifan pendonor darah. Pada Unit Donor Darah Palang Merah Indonesia Kota Surabaya, rekapitulasi data pelayanan darah selama ini masih berfokus pada statistik demografi deskriptif tanpa mengintegrasikan indikator pemeriksaan medis dan perilaku donor ulang ke dalam satu profil terpadu. Penelitian ini bertujuan mengelompokkan populasi pendonor aktif berdasarkan biomarker pemeriksaan klinis dan frekuensi donor guna mendukung strategi peningkatan retensi pendonor. Menggunakan data transaksi donor periode Januari hingga Oktober 2026, tahapan pembersihan anomali klinis dan agregasi tingkat individu menghasilkan 134.980 transaksi valid yang merepresentasikan 78.074 pendonor unik. Tujuh variabel terstandardisasi yang meliputi rata-rata usia, frekuensi donor, kadar hemoglobin, tekanan darah sistolik, tekanan darah diastolik, denyut nadi, dan berat badan dikelompokkan menggunakan algoritma k-means, dievaluasi melalui skor silhouette serta uji non-parametrik Kruskal-Wallis, dan diproyeksikan menggunakan analisis komponen utama. Hasil pemodelan membentuk empat segmen pendonor dengan skor silhouette optimal sebesar 0,2036 dan perbedaan signifikan pada seluruh indikator, yakni kelompok dewasa muda berbobot ringan sebesar 47,31 persen dengan frekuensi donor rendah, kelompok rutin matang sebesar 19,39 persen dengan rata-rata frekuensi tertinggi mencapai 3,54 kali dalam sepuluh bulan, kelompok tekanan darah dan berat badan menengah atas sebesar 32,54 persen, serta kelompok minoritas nadi tinggi sebesar 0,77 persen. Temuan ini menjadi dasar empiris perancangan pengingat donor ulang otomatis sesuai aturan jeda medis minimal enam puluh satu hari.

Kata Kunci: analisis retensi; biomarker medis; k-means clustering; principal component analysis; segmentasi pendonor darah

1. PENDAHULUAN

Ketersediaan darah yang aman, mencukupi, dan berkesinambungan merupakan komponen krusial dalam sistem pelayanan kesehatan darurat maupun perawatan medis klinis. Unit Donor Darah Palang Merah Indonesia (UDD PMI) Kota Surabaya sebagai pusat penyedia layanan transfusi darah utama di wilayah Surabaya dan sekitarnya melayani puluhan ribu pendonor sukarela setiap tahunnya. Berdasarkan catatan transaksi pelayanan darah periode Januari hingga Oktober 2026, tercatat lebih dari 136 ribu aktivitas pengambilan darah (*aftap*) berhasil yang berasal dari puluhan ribu individu pendonor dengan latar belakang demografi, wilayah domisili, serta kondisi fisiologis yang beragam. Meskipun volume transaksi tersebut tergolong tinggi, komposisi keaktifan pendonor menunjukkan ketimpangan yang perlu dicermati, yakni 60,36% pendonor hanya tercatat menyumbangkan darah sebanyak satu kali selama sepuluh bulan pengamatan, sedangkan hanya 39,64% yang berstatus sebagai pendonor rutin dengan frekuensi dua kali atau lebih.

Fenomena rendahnya proporsi pendonor rutin tersebut berkaitan erat dengan kesiapan kondisi fisik calon pendonor saat pemeriksaan penapisan awal serta kepatuhan terhadap jadwal donor ulang. Secara medis sesuai standar pelayanan transfusi darah di Indonesia, setiap pendonor darah lengkap (*whole blood*) wajib memenuhi ambang batas parameter klinis, meliputi kadar hemoglobin (HB), tekanan darah sistolik dan diastolik, denyut nadi, serta berat badan (BB), dan harus menjalani masa pemulihan minimal 61 hari (sekitar dua bulan) sebelum diperbolehkan mendonor kembali (Armayanti & Astuti, 2023). Kajian empiris pada unit donor

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

darah di Indonesia menunjukkan bahwa ketidaksiapan fisik seperti kadar HB rendah maupun tekanan darah di luar rentang normal menjadi penyebab utama penangguhan sementara calon pendonor sehingga menghambat kesinambungan donasi ulang (Armayanti & Astuti, 2023; Pratiwi & Prasetyo, 2026). Selama ini, pelaporan data di UDD PMI Kota Surabaya masih disajikan dalam bentuk rekapitulasi agregat per kategori demografi secara terpisah sehingga belum mampu memetakan bagaimana kombinasi profil kesehatan fisik berinteraksi dengan tingkat loyalitas dan frekuensi kedatangan pendonor.

Pemanfaatan pendekatan sains data dan pembelajaran mesin tanpa pengawasan (*unsupervised machine learning*) telah berkembang pesat dalam lima tahun terakhir untuk menggali struktur kelompok tersembunyi pada data pelayanan darah. Pada konteks lokal Indonesia, penerapan algoritma *K-Means Clustering* mulai dimanfaatkan di sejumlah unit Palang Merah Indonesia untuk memetakan profil pendonor berdasarkan usia, golongan darah, kelayakan seleksi, maupun hasil skrining infeksi menular lewat transfusi darah (Andini & Farokhah, 2022; Ramadhan, 2023; Ningsih et al., 2025; Pratiwi & Prasetyo, 2026). Sejalan dengan studi analitik bank darah di tingkat internasional oleh Twumasi dan Twumasi (2022) serta Ben Elmir et al. (2023), pemodelan berbasis riwayat transaksi donor terbukti efektif mendukung stabilitas rantai pasok darah. Dari sisi metodologi komputasi, kajian tinjauan komprehensif oleh Ezugwu et al. (2022) dan Ikotun et al. (2023) menegaskan bahwa *K-Means Clustering* tetap unggul dalam skalabilitas untuk mempartisi data multivariat berukuran puluhan ribu baris, sementara Gewers et al. (2021) dan Greenacre et al. (2022) membuktikan keandalan *Principal Component Analysis* (PCA) dalam memproyeksikan struktur dimensi tinggi ke bidang visual dua dimensi, serta Shutaywi dan Kachouie (2021) menekankan peran indeks *silhouette* dalam evaluasi jumlah klaster optimal.

Meskipun demikian, kajian terdahulu pada layanan transfusi darah umumnya masih memisahkan analisis perilaku retensi dari analisis biomarker fisiologis hasil penapisan medis atau berfokus pada peramalan permintaan agregat semata. Melalui program magang sains data yang dilaksanakan secara kolaboratif bersama petugas teknis dan manajemen pelayanan darah di UDD PMI Kota Surabaya, studi ini berinteraksi langsung dengan tim pengelola data lapangan untuk memvalidasi batasan parameter medis dan aturan jeda donor 61 hari. Penelitian ini bertujuan memetakan tipologi profil biomarker medis dan tingkat keaktifan dari 78.074 pendonor unik menggunakan *K-Means Clustering* dan PCA, memvalidasi signifikansi perbedaan antar segmen melalui uji non-parametrik Kruskal-Wallis, serta merumuskan strategi pengingat donor ulang dan edukasi pra-donor yang tepat sasaran bagi UDD PMI Kota Surabaya.

2. METODE

Penelitian ini menerapkan pendekatan analisis kuantitatif eksploratif dan pemodelan klastering berbasis data observasional pelayanan darah di UDD PMI Kota Surabaya. Lokasi kegiatan bertempat di UDD PMI Kota Surabaya dengan cakupan data transaksi donor berhasil selama sepuluh bulan, terhitung sejak 1 Januari hingga 31 Oktober 2026 atau mencakup rentang waktu pengamatan selama 304 hari kalender, sedangkan tahapan pembersihan data, pemodelan, dan validasi operasional bersama pembimbing lapangan dilaksanakan pada periode Agustus hingga November 2026. Pada kondisi awal sebelum pemodelan

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

diimplementasikan, pemantauan profil pendonor di instansi mitra masih mengandalkan tabel distribusi frekuensi tunggal berdasarkan jenis kelamin, golongan darah, dan kelompok usia tanpa adanya pengelompokan multivariat yang menggabungkan biomarker klinis dengan riwayat keaktifan individu. Luaran yang diharapkan dari implementasi metode ini adalah terbentuknya peta segmentasi pendonor aktif yang tervalidasi secara statistik beserta rekomendasi sistem pengingat donor ulang berbasis jadwal kelayakan medis.

Sumber data utama diperoleh dari basis data sistem informasi pelayanan donor UDD PMI Kota Surabaya yang terdiri atas 136.031 baris catatan transaksi pengambilan darah berhasil. Teknik pemilihan sampel menggunakan pendekatan *total sampling* terhadap seluruh catatan transaksi valid yang memenuhi standar kelayakan fisiologis penapisan medis donor darah. Tahap prapemrosesan data dilakukan untuk menjamin validitas dan reliabilitas indikator dengan menyaring nilai hilang (*missing values*) serta mengeluarkan pencilan ekstrem akibat kesalahan pencatatan input di luar rentang klinis wajar, meliputi kadar HB di luar rentang 12,0–18,0 g/dL, tekanan darah sistolik (TENSI1) di luar 80–200 mmHg, tekanan darah diastolik (TENSI2) di luar 50–120 mmHg, denyut nadi (NADI) di luar 50–120 denyut per menit (bpm), BB di luar 45–150 kg, serta usia pendonor di luar 17–70 tahun. Proses penyaringan ketat tersebut mempertahankan 134.980 baris transaksi valid atau setara dengan 99,23% dari total populasi awal.

Karena satu individu pendonor dapat melakukan donasi lebih dari satu kali sepanjang periode Januari hingga Oktober 2026, dilakukan transformasi struktur data dari level transaksi menuju level profil individu menggunakan nomor identitas pendonor (NODONOR) yang telah dianonimisasi sebagai kunci utama (*primary key*). Agregasi tersebut menghasilkan tepat 78.074 entitas pendonor unik dengan menghitung total frekuensi kedatangan selama sepuluh bulan serta nilai rata-rata parameter fisiologis untuk setiap pendonor ke-i pada variabel ke-j sebagaimana dirumuskan pada Persamaan (1), di mana n_i menyatakan jumlah transaksi berhasil milik pendonor ke-i dan $x_{i,j,t}$ merupakan nilai pengukuran biomarker pada kunjungan ke-t.

$$\bar{x}_{i,j} = \frac{1}{n_i} \sum_{t=1}^{n_i} x_{i,j,t} \tag{1}$$

Sebelum memasuki tahap klastering, ketujuh fitur numerik terpilih, yakni rata-rata usia (UMUR_RATA), frekuensi donor (JUMLAH_DONOR), rata-rata hemoglobin (HB_RATA), rata-rata tekanan darah sistolik (TENSI1_RATA), rata-rata tekanan darah diastolik (TENSI2_RATA), rata-rata denyut nadi (NADI_RATA), dan rata-rata berat badan (BB_RATA), disetarakan skalanya menggunakan metode standardisasi Z-score (StandardScaler) pada lingkungan komputasi Python guna mencegah variabel bersatuan ratusan seperti tekanan darah mendominasi perhitungan jarak Euclidean terhadap variabel bersatuan belasan seperti kadar HB (Ezugwu et al., 2022; Ikotun et al., 2023). Transformasi standardisasi untuk setiap observasi ke-i pada variabel ke-j dihitung berdasarkan nilai rata-rata populasi μ_j dan simpangan baku σ_j mengikuti Persamaan (2).

$$z_{i,j} = \frac{x_{i,j} - \mu_j}{\sigma_j}$$

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Segmentasi populasi pendonor dilakukan menggunakan algoritma *K-Means Clustering* dengan inisialisasi pusat klaster *k-means++* untuk meminimalkan jumlah kuadrat jarak dalam klaster atau *Within-Cluster Sum of Squares* (WCSS) pada ruang fitur tujuh dimensi (Ezugwu et al., 2022; Ikotun et al., 2023; Ramadhan, 2023). Fungsi objektif optimasi partisi klaster K-Means dinyatakan pada Persamaan (3), dengan k melambangkan jumlah klaster, C_c menyatakan himpunan anggota klaster ke-c, $\mathbf{z}_i$ merupakan vektor fitur terstandardisasi pendonor ke-i, dan $\boldsymbol{\mu}_c$ adalah vektor titik pusat (*centroid*) klaster ke-c.

$$J = \sum_{c=1}^{k} \sum_{\mathbf{z}_i \in C_c} \|\mathbf{z}_i - \boldsymbol{\mu}_c\|^2 \tag{3}$$

Penentuan jumlah klaster optimal dievaluasi pada kandidat $k \in \{3,4,5,6\}$ menggunakan metrik *Silhouette Coefficient* (Shutaywi & Kachouie, 2021; Ningsih et al., 2025) pada 15.000 sampel acak terstratifikasi untuk efisiensi komputasi memori. Nilai koefisien silhouette $s(i)$ untuk setiap titik observasi ke-i dihitung dari selisih jarak rata-rata antar klaster terdekat $b(i)$ terhadap jarak rata-rata intra-klaster $a(i)$ sesuai Persamaan (4), di mana nilai rata-rata koefisien yang paling tinggi menunjukkan konfigurasi partisi terbaik.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{4}$$

Untuk membuktikan secara inferensial bahwa segmen-segmen yang terbentuk memiliki perbedaan distribusi biomarker dan keaktifan yang signifikan, diterapkan uji non-parametrik Kruskal-Wallis H-Test pada taraf signifikansi $\alpha = 0{,}05$ (Okoye & Hosseini, 2024; Arifin et al., 2026). Pemilihan uji berbasis peringkat ini didasarkan pada karakteristik variabel frekuensi donor dan denyut nadi yang memiliki kecondongan distribusi (*skewness*) tinggi sehingga tidak memenuhi asumsi normalitas parametrik. Statistik uji H dihitung berdasarkan total sampel N, jumlah anggota tiap klaster n_c, dan jumlah peringkat R_c pada klaster ke-c seperti ditunjukkan pada Persamaan (5).

$$H = \frac{12}{N(N+1)} \sum_{c=1}^{k} \frac{R_c^2}{n_c} - 3(N+1) \tag{5}$$

Tahap akhir analisis adalah memproyeksikan struktur pengelompokan tujuh dimensi ke bidang visual dua dimensi menggunakan *Principal Component Analysis* (PCA) guna mengamati arah pemisahan klaster beserta kontribusi pembobotan (*loadings*) setiap indikator (Gewers et al., 2021; Greenacre et al., 2022). Matriks skor komponen utama $\mathbf{Y}$ diperoleh melalui perkalian matriks data terstandardisasi $\mathbf{Z}$ dengan matriks vektor eigen pembobot $\mathbf{W}$ sebagaimana dirumuskan pada Persamaan (6).

$$\mathbf{Y} = \mathbf{ZW} \tag{6}$$

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

3. HASIL DAN PEMBAHASAN

3.1 Karakteristik demografi dan pola retensi pendonor

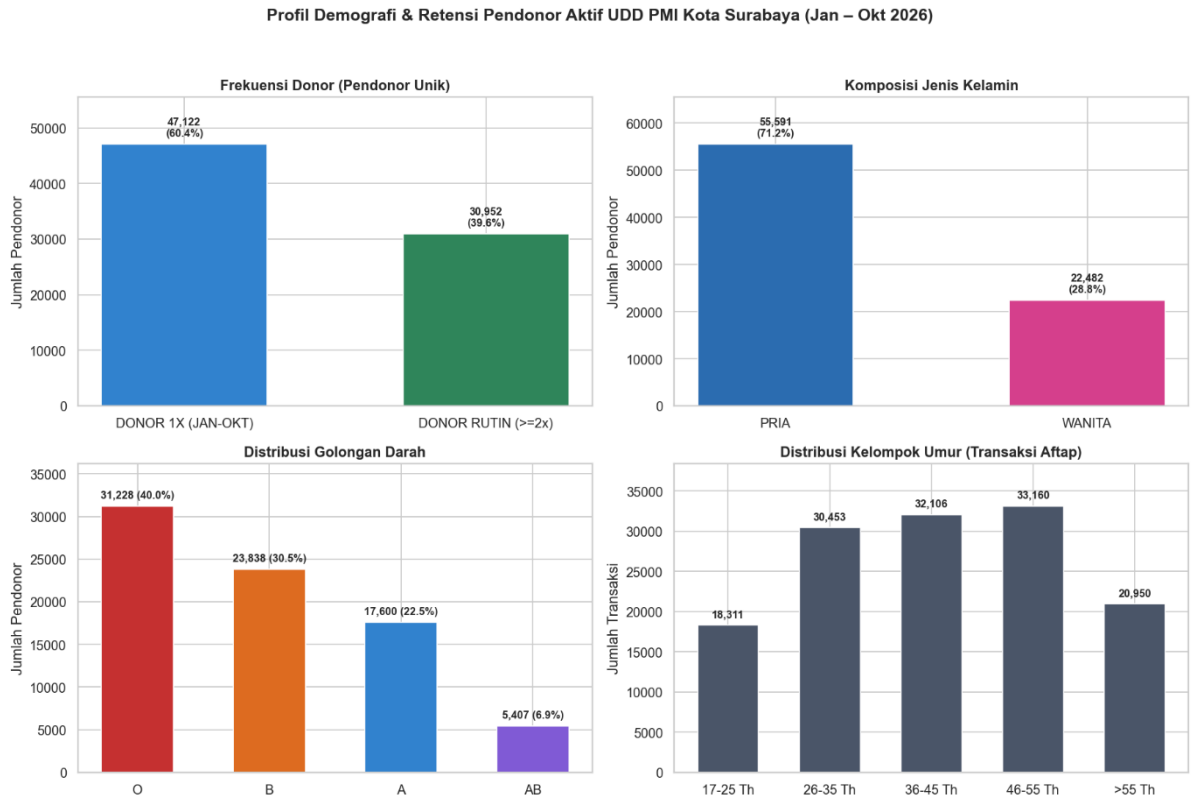
Hasil eksplorasi terhadap 134.980 transaksi aftap valid dan 78.074 pendonor unik di UDD PMI Kota Surabaya menunjukkan bahwa populasi pendonor aktif didominasi oleh laki-laki sebanyak 55.591 individu (71,20%), sedangkan perempuan berjumlah 22.482 individu (28,80%) dan 1 individu tidak tercatat keterangannya. Dominasi pendonor laki-laki ini sejalan dengan temuan studi penapisan donor di Indonesia yang menunjukkan bahwa calon pendonor perempuan memiliki tingkat penangguhan sementara lebih tinggi akibat kadar HB di bawah ambang batas kelayakan 12,5 g/dL serta faktor fisiologis siklus menstruasi (Armayanti & Astuti, 2023). Berdasarkan golongan darah, distribusi pendonor unik berturut-turut ditempati oleh golongan darah O sebanyak 31.228 orang (40,00%), golongan darah B sebanyak 23.838 orang (30,53%), golongan darah A sebanyak 17.600 orang (22,54%), dan golongan darah AB sebanyak 5.407 orang (6,93%).

Ditinjau dari sebaran usia pada level transaksi pengambilan darah, kelompok usia produktif dewasa mendominasi kontribusi kantong darah, dipimpin oleh rentang usia 46–55 tahun sebanyak 33.160 transaksi (24,57%), diikuti rentang 36–45 tahun sebanyak 32.106 transaksi (23,79%), dan rentang 26–35 tahun sebanyak 30.453 transaksi (22,56%). Sementara itu, kelompok pendonor pemula usia 17–25 tahun menyumbang 18.311 transaksi (13,57%) dan kelompok usia di atas 55 tahun menyumbang 20.950 transaksi (15,52%). Ringkasan lengkap komposisi demografi dan tingkat keaktifan disajikan pada Tabel 1 serta diilustrasikan secara visual pada Gambar 1.

Tabel 1. Karakteristik demografi dan tingkat keaktifan pendonor

Dimensi Analisis	Kategori	Jumlah	Persentase (%)
Tipe Keaktifan (n = 78.074)	Donor 1 Kali	47.122	60,36
	Donor Rutin (>= 2 Kali)	30.952	39,64
Jenis Kelamin (n = 78.074)	Laki-Laki (Pria)	55.591	71,20
	Perempuan (Wanita)	22.482	28,80
	Tidak Terketerangan	1	0,00
Golongan Darah (n = 78.074)	Golongan O	31.228	40,00
	Golongan B	23.838	30,53
	Golongan A	17.600	22,54
	Golongan AB	5.407	6,93
Kelompok Usia Transaksi (n = 134.980)	17–25 Tahun	18.311	13,57
	26–35 Tahun	30.453	22,56
	36–45 Tahun	32.106	23,79
	46–55 Tahun	33.160	24,57
	> 55 Tahun	20.950	15,52

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)



Gambar 1. Distribusi demografi dan tingkat keaktifan pendonor

Temuan penting pada analisis retensi adalah keterkaitan antara proporsi pendonor satu kali (60,36%) dengan aturan jeda medis minimal 61 hari setelah donor berhasil. Dalam rentang pengamatan selama 304 hari dari 1 Januari hingga 31 Oktober 2026, seorang pendonor yang menyumbangkan darah pada awal Januari dan selalu kembali tepat setelah 61 hari hanya dapat mendonor maksimal sebanyak lima kali sebelum periode pencatatan berakhir. Lebih lanjut, individu yang baru melakukan donor pertama kali setelah tanggal 31 Agustus 2026 (sepanjang September dan Oktober 2026) secara medis memang belum memasuki batas kelayakan 61 hari untuk melakukan donor kedua hingga akhir Oktober 2026. Pemahaman atas efek batas waktu pengamatan ini mencegah kesalahan interpretasi manajerial dengan membedakan pendonor yang memang belum jatuh tempo jadwal donor ulang dari pendonor yang telah melewati masa tunggu 61 hari namun belum kembali.

3.2 Evaluasi klastering K-Means dan profil tipologi biomarker pendonor

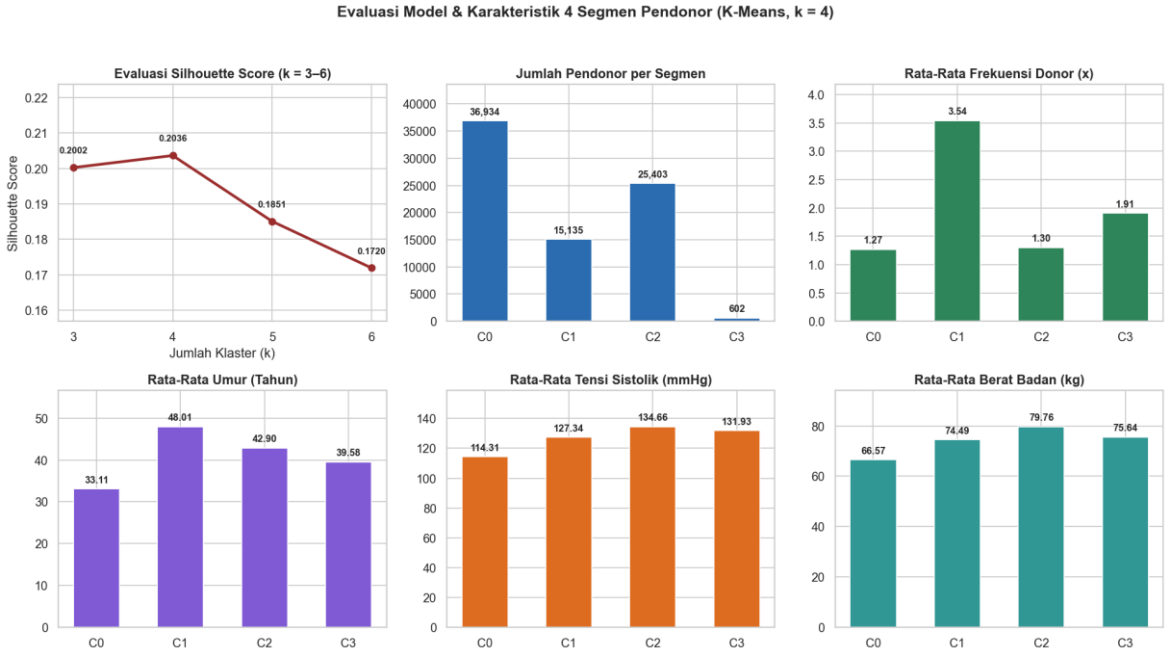
Evaluasi pemodelan K-Means pada ruang tujuh variabel terstandardisasi menghasilkan nilai *Silhouette Score* sebesar 0,2002 untuk $k = 3$, meningkat mencapai puncak tertinggi sebesar 0,2036 pada $k = 4$, lalu menurun menjadi 0,1851 pada $k = 5$ dan 0,1720 pada $k = 6$. Dengan demikian, konfigurasi empat klaster ($k = 4$) ditetapkan sebagai struktur segmentasi paling optimal. Nilai *Silhouette Score* pada rentang moderat tersebut merupakan karakteristik wajar pada data skrining kesehatan populasi umum karena seluruh subjek yang tercatat merupakan pendonor lolos penapisan medis sehingga sebaran nilai biomarker klinisnya berada dalam koridor fisiologis normal yang saling berdekatan.

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Keempat klaster yang terbentuk memperlihatkan pemisahan tipologi fisiologis dan perilaku donasi yang kontras sebagaimana dirangkum pada Tabel 2 dan ditampilkan pada Gambar 2. Klaster 0 (C0: Dewasa Muda & Berat Badan Ringan) mencakup proporsi populasi terbesar sebanyak 36.934 pendonor (47,31%) dengan rata-rata usia termuda 33,11 tahun, proporsi pendonor perempuan tertinggi sebesar 40,88%, rata-rata BB paling ringan 66,57 kg, tekanan darah sistolik-diastolik rendah-normal sebesar 114,31/75,70 mmHg, dan rata-rata frekuensi donasi masih rendah yakni 1,27 kali. Sebaliknya, Klaster 1 (C1: Rutin & Matang) beranggotakan 15.135 pendonor (19,39%) yang didominasi laki-laki (83,85%) dengan usia paling matang rata-rata 48,00 tahun, profil hemodinamik stabil (127,34/81,88 mmHg), serta rata-rata frekuensi donasi tertinggi mencapai 3,54 kali dalam sepuluh bulan, yang menunjukkan kedisiplinan kembali mendonor rata-rata setiap 75 hingga 85 hari setelah masa tunggu 61 hari selesai.

Tabel 2. Profil rata-rata biomarker dan frekuensi donor empat klaster

Segmen Klaster	Pendonor (%)	Usia (Th)	Frek. (x)	HB (g/dL)	Tensi (mmHg)	Nadi (bpm)	BB (kg)	Pria (%)
C0: Dewasa Muda & BB Ringan	36.934 (47,31%)	33,11	1,27	14,14	114,31 / 75,70	79,98	66,57	59,12
C1: Rutin & Matang	15.135 (19,39%)	48,00	3,54	14,25	127,34 / 81,88	80,08	74,49	83,85
C2: Tensi & BB Menengah-Atas	25.403 (32,54%)	42,90	1,30	14,70	134,66 / 86,29	79,96	79,77	81,30
C3: Nadi & Tensi Tinggi	602 (0,77%)	39,58	1,91	14,30	131,93 / 86,23	91,35	75,64	68,44



Gambar 2. Evaluasi Silhouette Score dan profil empat segmen pendonor

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Sementara itu, Klaster 2 (C2: Tensi & Berat Badan Menengah-Atas) menghimpun 25.403 pendonor (32,54%) pada rentang usia produktif menengah rata-rata 42,90 tahun yang memiliki kadar HB rata-rata tertinggi sebesar 14,70 g/dL dan BB rata-rata terbesar sebesar 79,77 kg, disertai tekanan darah rata-rata 134,66/86,29 mmHg yang berada pada batas atas rentang normal, namun rata-rata frekuensi donornya masih tergolong rendah yakni 1,30 kali. Adapun Klaster 3 (C3: Nadi & Tensi Tinggi) merupakan segmen minoritas beranggotakan 602 pendonor (0,77%) dengan rata-rata denyut nadi menonjol sebesar 91,35 bpm dan tekanan darah 131,93/86,23 mmHg. Terpisahnya C3 sebagai klaster tersendiri disebabkan oleh tingginya konsentrasi pencatatan denyut nadi populasi umum pada angka basal 80 bpm sehingga kelompok pendonor dengan aktivitas simpatis atau denyut nadi di atas 90 bpm memiliki jarak Euclidean yang jauh dari pusat massa populasi.

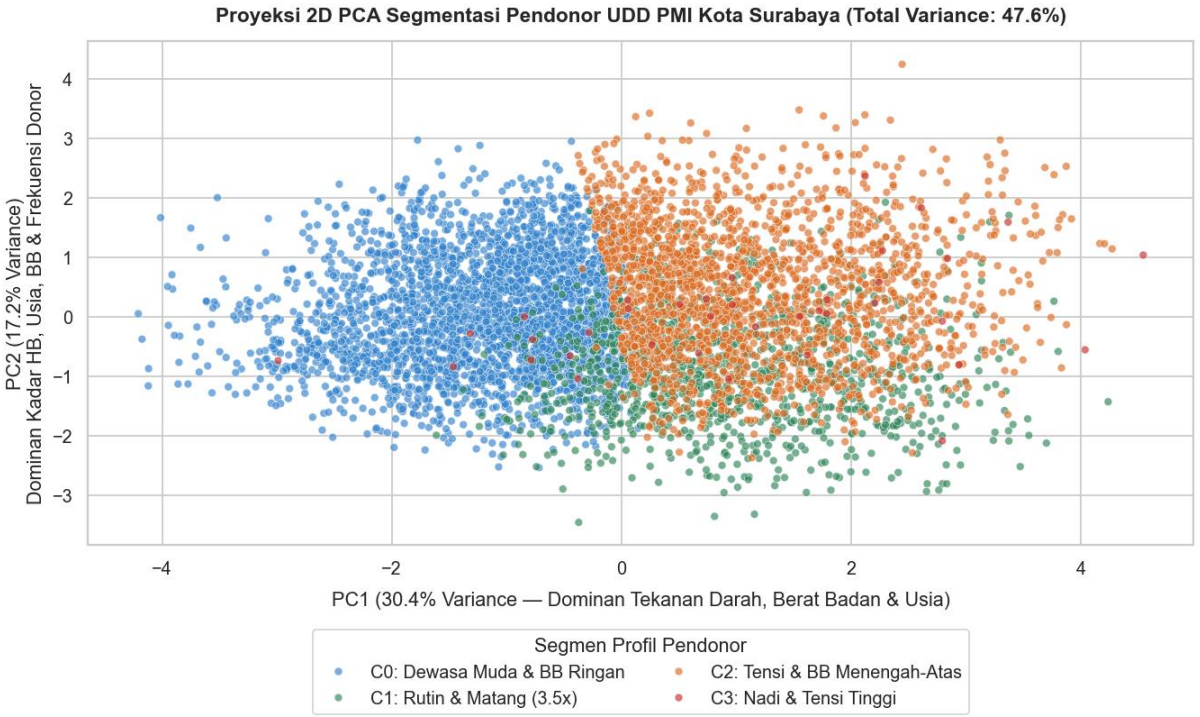
3.3 Validasi statistik Kruskal-Wallis dan proyeksi reduksi dimensi PCA

Pengujian signifikansi perbedaan antar keempat segmen menggunakan uji non-parametrik Kruskal-Wallis H-Test membuktikan bahwa seluruh tujuh variabel biomarker dan keaktifan memiliki perbedaan distribusi yang sangat signifikan secara statistik dengan nilai p < 0,0001 pada seluruh indikator (Tabel 3). Nilai statistik H terbesar tercatat pada variabel frekuensi donor (H = 43.259,50) yang menegaskan pemisahan tajam antara segmen loyal C1 (median 3,00 kali) terhadap C0 dan C2 (median 1,00 kali), diikuti oleh tekanan darah sistolik (H = 39.249,80) dan tekanan darah diastolik (H = 31.798,99) yang membedakan segmen C0 (median 115,00/77,50 mmHg) dari segmen C2 (median 132,50/85,50 mmHg).

Tabel 3. Hasil uji Kruskal-Wallis dan pembobotan PCA antar klaster

Variabel Analisis	Med. C0	Med. C1	Med. C2	Med. C3	Statistik H	p-value	PC1 Load.	PC2 Load.
UMUR_RATA (Th)	31,00	49,00	44,00	39,00	17.480,46	< 0,0001	0,3743	-0,4660
JUMLAH_DONOR (x)	1,00	3,00	1,00	2,00	43.259,50	< 0,0001	0,2424	-0,3539
HB_RATA (g/dL)	14,00	14,20	14,70	14,20	3.808,36	< 0,0001	0,1230	0,6723
TENSI1_RATA (mmHg)	115,00	126,25	132,50	130,67	39.249,80	< 0,0001	0,5972	0,0052
TENSI2_RATA (mmHg)	77,50	81,50	85,50	86,00	31.798,99	< 0,0001	0,5657	0,0505
NADI_RATA (bpm)	80,00	80,00	80,00	90,00	15.602,60	< 0,0001	0,0686	0,0023
BB_RATA (kg)	65,00	73,18	78,00	73,88	13.974,06	< 0,0001	0,3234	0,4506

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)



Gambar 3. Proyeksi dua dimensi PCA pada empat segmen profil pendonor

Pada tahap reduksi dimensi, dua komponen utama pertama dari PCA mampu merangkum 47,62% varians total dari ruang tujuh dimensi (Tabel 3 dan Gambar 3). Komponen Utama Pertama (PC1) yang menangkap 30,39% varians memiliki bobot positif tertinggi pada tekanan darah sistolik (0,5972), tekanan darah diastolik (0,5657), usia (0,3743), dan BB (0,3234), sejalan dengan prinsip ortogonalitas komponen utama dalam memaksimalkan ragam multivariat (Gewers et al., 2021; Greenacre et al., 2022). Secara visual pada Gambar 3, sumbu horizontal PC1 memisahkan secara tegas distribusi segmen C0 (biru) di kuadran kiri yang berprofil tekanan darah rendah dan postur lebih ringan terhadap segmen C2 (oranye) di kuadran kanan yang berprofil tekanan darah serta BB lebih tinggi. Titik sebaran segmen C3 tampak tumpang tindih pada bidang proyeksi dua dimensi PC1 dan PC2 karena bobot pemuatan (*loading*) variabel denyut nadi pada kedua komponen utama tersebut relatif kecil (0,0686 dan 0,0023), sehingga variasi denyut nadi lebih tercermin pada komponen utama berdimensi lebih tinggi. Sementara itu, Komponen Utama Kedua (PC2) yang menangkap 17,23% varians dipengaruhi secara berlawanan oleh kadar HB (0,6723) dan BB (0,4506) di arah positif terhadap usia (-0,4660) dan frekuensi donor (-0,3539) di arah negatif, sehingga menempatkan konsentrasi segmen pendonor rutin matang C1 (hijau) pada bagian bawah bidang proyeksi.

3.4 Implikasi operasional dan integrasi pengingat donor ulang berbasis jeda 61 hari

Perbandingan kondisi sebelum dan sesudah penerapan pemodelan segmentasi menunjukkan peningkatan signifikan dalam kapabilitas analitik pelayanan donor di UDD PMI Kota Surabaya. Sebelumnya, seluruh 78.074 pendonor diperlakukan secara seragam dalam komunikasi pasca-donasi tanpa membedakan profil kesehatan maupun potensi retensinya. Melalui hasil segmentasi yang diintegrasikan ke dalam prototipe dashboard analitik interaktif bersama tim UDD PMI Kota Surabaya, pengelola layanan darah kini dapat mengidentifikasi

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

dua kelompok prioritas utama untuk konversi menjadi pendonor rutin, yaitu segmen C0 (47,31%) dan segmen C2 (32,54%), yang secara akumulatif mencakup 79,85% populasi pendonor aktif namun masih memiliki rata-rata frekuensi kedatangan di bawah 1,35 kali.

Adopsi inovasi praktis yang direkomendasikan dari hasil studi ini adalah penerapan sistem pengingat donor ulang otomatis berbasis tanggal donasi terakhir masing-masing individu (TGL_TERAKHIR + 61 hari) yang disesuaikan dengan karakteristik klinis tiap klaster. Untuk segmen C0 yang didominasi dewasa muda dengan proporsi perempuan mencapai 40,88%, pesan pengingat kelayakan donor ulang di hari ke-61 perlu disertai edukasi persiapan pra-donor tiga hari sebelumnya (H-3) berupa anjuran kecukupan istirahat dan asupan nutrisi kaya zat besi guna mencegah penolakan akibat HB rendah. Sebaliknya, bagi segmen C2 dan C3 yang memiliki kecenderungan tekanan darah sistolik-diastolik di ambang atas normal serta denyut nadi lebih tinggi, pesan pengingat jadwal donor ulang difokuskan pada imbauan menjaga kualitas tidur, mengurangi konsumsi kafein menjelang hari pemeriksaan, dan beristirahat tenang sebelum penapisan medis agar tingkat keberhasilan pengambilan darah tetap terjaga.

4. KESIMPULAN

Pemodelan segmentasi profil biomarker medis dan tingkat keaktifan pendonor darah di Unit Donor Darah Palang Merah Indonesia Kota Surabaya berhasil memetakan 134.980 transaksi valid dari 78.074 pendonor unik periode Januari hingga Oktober 2026 ke dalam empat klaster yang tervalidasi secara empiris. Penerapan algoritma K-Means pada tujuh indikator klinis dan frekuensi donasi mencapai struktur partisi optimal pada empat klaster dengan skor silhouette sebesar 0,2036, didukung uji Kruskal-Wallis yang membuktikan perbedaan signifikan pada seluruh variabel serta proyeksi Principal Component Analysis yang merangkum 47,62% varians total. Hasil pemetaan memisahkan kelompok dewasa muda berbobot ringan sebesar 47,31%, kelompok pendonor rutin matang sebesar 19,39% dengan frekuensi tertinggi 3,54 kali, kelompok tekanan darah dan berat badan menengah atas sebesar 32,54%, serta kelompok minoritas nadi tinggi sebesar 0,77%. Temuan ini memberikan manfaat operasional nyata bagi pengelola layanan transfusi darah dalam merancang pengingat donor ulang otomatis tepat setelah masa tunggu medis minimal 61 hari berakhir disertai edukasi persiapan pra-donor sesuai profil kesehatan masing-masing segmen. Penelitian selanjutnya disarankan memperluas cakupan data penuh satu tahun kalender dan mengintegrasikan pemodelan prediktif peluang kembali mendonor.

Ucapan Terima Kasih

Penulis menyampaikan terima kasih yang sebesar-besarnya kepada pimpinan, pembimbing lapangan, serta seluruh jajaran staf Unit Donor Darah Palang Merah Indonesia (UDD PMI) Kota Surabaya atas izin akses data teranonimisasi, pendampingan teknis, dan masukan klinis selama pelaksanaan program magang sains data. Ucapan terima kasih juga disampaikan kepada Program Studi Sains Data, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jawa Timur atas dukungan akademik dalam penyusunan artikel ini. Penulis menyatakan tidak terdapat konflik kepentingan finansial, komersial, maupun profesional yang memengaruhi hasil penelitian ini. Studi ini tidak menerima hibah khusus dari lembaga pendanaan di sektor publik, komersial, atau nirlaba.

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

DAFTAR PUSTAKA

Andini, T. D., & Farokhah, L. (2022). Peningkatan ketersediaan darah sesuai segmentasi umur menggunakan K-Means clustering. *Jurnal Manajemen Informatika (JAMIKA)*, 12(2), 126–136. https://doi.org/10.34010/jamika.v12i2.7897

Arifin, S., Manurung, M. M., Jonathan, S., & Alifah, M. (2026). A review of Kruskal-Wallis test applications in scientific research. *Journal of Applied Finance & Risk Management*, 4(2), 42–55. https://e-journal.president.ac.id/index.php/JAFRM/article/download/6301/2413

Armayanti, D., & Astuti, Y. (2023). Gambaran penangguhan pendonor di Unit Donor Darah PMI Kota Yogyakarta tahun 2022. *Jurnal Sehat Mandiri*, 18(2), 11–23. https://doi.org/10.33761/jsm.v18i2.1015

Ben Elmir, W., Hemmak, A., & Senouci, B. (2023). Smart platform for data blood bank management: Forecasting demand in blood supply chain using machine learning. *Information*, 14(1), Article 31. https://doi.org/10.3390/info14010031

Ezugwu, A. E., Ikotun, A. M., Oyelade, O. O., Abualigah, L., Agushaka, J. O., Eke, C. I., & Akinyelu, A. A. (2022). A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. *Engineering Applications of Artificial Intelligence*, 110, Article 104743. https://doi.org/10.1016/j.engappai.2022.104743

Gewers, F. L., Ferreira, G. R., Arruda, H. F. D., Silva, F. N., Comin, C. H., Amancio, D. R., & Costa, L. D. F. (2021). Principal component analysis: A natural approach to data exploration. *ACM Computing Surveys*, 54(4), Article 70. https://doi.org/10.1145/3447755

Greenacre, M., Groenen, P. J. F., Hastie, T., D'Enza, A. I., Markos, A., & Tuzhilina, E. (2022). Principal component analysis. *Nature Reviews Methods Primers*, 2, Article 100. https://doi.org/10.1038/s43586-022-00184-w

Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 622, 178–210. https://doi.org/10.1016/j.ins.2022.11.139

Ningsih, A. W., Novriyenni, & Khair, H. (2025). Pengelompokkan pendonor darah berdasarkan golongan darah di PMI Kabupaten Langkat menggunakan metode clustering K-Means. *Fusion: Journal of Research in Engineering, Technology and Applied Sciences*, 2(1), 40–48. https://doi.org/10.66341/fusion.v2i1.261

Okoye, K., & Hosseini, S. (2024). Mann–Whitney U test and Kruskal–Wallis H test statistics in R. In *R programming: Statistical data analysis in research* (pp. 225–246). Springer Nature Singapore. https://doi.org/10.1007/978-981-97-3385-9_11

Pratiwi, D., & Prasetyo, A. B. (2026). Klasterisasi donor darah reaktif IMLTD di PMI DKI Jakarta dengan K-Means. *Indonesian Journal of Multidisciplinary on Social and Technology*, 4(3), 2085–2098. https://doi.org/10.69693/ijmst.v4i3.12496

Ramadhan, M. R. (2023). Clustering pasien yang layak donor darah dengan algoritma K-Means studi kasus PMI Kota Medan. *Jurnal Sains dan Teknologi Informasi*, 2(3), 74–79. https://doi.org/10.47065/jussi.v2i3.4596

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Shutaywi, M., & Kachouie, N. N. (2021). Silhouette analysis for performance evaluation in machine learning with applications to clustering. *Entropy*, 23(6), Article 759. https://doi.org/10.3390/e23060759

Twumasi, C., & Twumasi, J. (2022). Machine learning algorithms for forecasting and backcasting blood demand data with missing values and outliers: A study of Tema General Hospital of Ghana. *International Journal of Forecasting*, 38(3), 1258–1277. https://doi.org/10.1016/j.ijforecast.2021.10.008

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)