

Pengelompokan Pasien Penyakit Kronis Diabetes, Hipertensi, dan Gagal Ginjal Menggunakan *K-Means* Berdasarkan Variabel Klinis

Jurusan Rekam Medis, Akkes Sapta Bakti Bengkulu, Indonesia¹

Fakultas Teknik, Teknik Informatika, Universitas Muhammadiyah Bengkulu, Indonesia²

E-mail: fitratridamayanti93@gmail.com

Abstract

The increasing number of patients with chronic diseases poses challenges in delivering effective and efficient healthcare services. The heterogeneity of patients' clinical characteristics makes risk identification and clinical decision-making processes more complex. This study aimed to classify patients based on similarities in clinical characteristics using the K-Means algorithm to identify specific health patterns. The study employed a quantitative approach with a cross-sectional observational design and data mining techniques. Research data were obtained from a public Kaggle dataset which, after the data cleaning process, resulted in 51 patient records with analytical variables including age, blood pressure, blood glucose level, hemoglobin, diabetes mellitus, hypertension, anemia, and red blood cell (RBC) condition. Cluster analysis was performed using the K-Means algorithm, while cluster validity was evaluated using the Silhouette and Dunn indices. The results showed that the optimal number of clusters was three, with a Silhouette score of 0.537 and a Dunn index of 0.612. Cluster 1 (n=38) represented a low-risk profile characterized by relatively normal blood pressure and blood glucose levels without cases of diabetes mellitus or hypertension. Cluster 2 (n=8) was characterized by a high prevalence of anemia (87.5%) and low hemoglobin levels. Meanwhile, Cluster 3 (n=5) exhibited the most severe metabolic profile, with hypertension prevalence reaching 100%, diabetes mellitus 80%, and the highest average blood glucose level 348.8 mg/dL. These findings indicate that the K-Means method is effective in identifying patient segmentation based on distinct clinical characteristics. However, due to the relatively limited sample size and imbalanced cluster distribution, the findings should be interpreted as an exploratory analysis that requires further validation using larger and more diverse datasets.

Keywords: clinical pattern, data mining, health stratification, patient segmentation, unsupervised learning

Abstrak

Peningkatan jumlah pasien dengan penyakit kronis menimbulkan tantangan dalam pengelolaan layanan kesehatan yang efektif dan efisien. Heterogenitas karakteristik klinis pasien menyebabkan proses identifikasi risiko dan pengambilan keputusan klinis menjadi lebih kompleks. Penelitian ini bertujuan mengelompokkan pasien berdasarkan kesamaan karakteristik klinis menggunakan algoritma *K-Means* untuk mengidentifikasi pola kesehatan tertentu. Penelitian menggunakan pendekatan kuantitatif dengan desain observasional *cross-sectional* dan teknik data mining. Data penelitian diperoleh dari dataset publik Kaggle yang setelah proses pembersihan data menghasilkan 51 pasien dengan variabel analisis berupa usia, tekanan darah, kadar glukosa darah, hemoglobin, DM, hipertensi, anemia, dan kondisi sel darah merah. Analisis kluster dilakukan menggunakan algoritma *K-Means* dengan evaluasi validitas kluster menggunakan indeks *Silhouette* and *Dunn*. Hasil penelitian menunjukkan bahwa jumlah kluster optimal adalah tiga kluster dengan nilai *Silhouette* sebesar 0,537 dan indeks *Dunn* sebesar 0,612. Kluster 1 (n=38) menunjukkan profil risiko rendah dengan tekanan darah dan kadar glukosa darah relatif normal serta tanpa kasus DM maupun hipertensi. Kluster 2 (n=8) ditandai prevalensi anemia tinggi (87,5%) dan kadar hemoglobin rendah. Sementara itu, Kluster 3 (n=5) menunjukkan profil metabolik paling berat dengan prevalensi hipertensi sebesar 100%, DM 80%, serta rata-rata glukosa darah tertinggi 348,8 mg/dL. Temuan ini menunjukkan bahwa metode *K-Means* mampu mengidentifikasi segmentasi pasien berdasarkan karakteristik klinis yang berbeda. Namun, karena ukuran sampel relatif terbatas dan distribusi kluster tidak seimbang, hasil penelitian lebih tepat diposisikan sebagai analisis eksploratif awal yang memerlukan validasi lebih lanjut menggunakan dataset yang lebih besar dan beragam.

Kata Kunci: analisis kluster, karakteristik klinis, rekam medis, segmentasi pasien, stratifikasi kesehatan

Naskah masuk: 20 Desember 2025, Naskah direvisi: 2 Januari 2026, Naskah diterima: 3 Maret 2026

Naskah diterbitkan secara online: 30 April 2026

©2022/ Penulis. Artikel ini merupakan artikel dengan akses terbuka di bawah lisensi CC BY-SA

(<https://creativecommons.org/licenses/by-sa/4.0>)

1. Pendahuluan

Penyakit kronis seperti Diabetes Melitus (DM), hipertensi, dan gagal ginjal kronik merupakan beban kesehatan masyarakat global. Penyakit-penyakit ini menyumbang proporsi besar kematian dunia sekitar 60% dari total kematian global disebabkan oleh penyakit tidak menular (Aldhyani, et al., 2020). DM menjadi tantangan kesehatan tersendiri dengan jumlah pengidap dewasa mencapai 537 juta orang pada tahun 2021 (Taurbekova, et al., 2025). Fenomena tersebut juga terjadi di Indonesia, yang ditandai dengan meningkatnya prevalensi DM dan hipertensi. Survei Kesehatan Indonesia (SKI) tahun 2023 menunjukkan bahwa prevalensi hipertensi pada penduduk usia ≥ 18 tahun di Indonesia mencapai 30,8% berdasarkan hasil pengukuran tekanan darah, sedangkan prevalensi DM pada penduduk usia ≥ 15 tahun mencapai 11,7% berdasarkan hasil pemeriksaan kadar glukosa darah (Kementerian Kesehatan Republik Indonesia, 2023). Penyakit ginjal kronik juga mengalami peningkatan dan seringkali terjadi sebagai komplikasi DM dan hipertensi. Meskipun demikian, hanya sekitar 0,38% penduduk Indonesia yang dilaporkan mengidap penyakit ginjal kronik (*Chronic Kidney Disease/CKD*) pada tahun 2023 (Agatha, et al., 2025). Sebagian besar penderita belum menyadari penyakitnya sampai pada tahap lanjut. Pengelolaan CKD menimbulkan beban biaya tinggi. Pendanaan perawatan penyakit ginjal menempati urutan kedua tertinggi di BPJS Kesehatan setelah penyakit jantung (Shabrina, et al., 2022). Dengan latar belakang tersebut, penanganan profil kesehatan pasien kronis menjadi penting untuk mengidentifikasi kelompok risiko sehingga intervensi dapat ditargetkan secara efektif.

Beban penyakit kronis pada tingkat lokal juga mengalami peningkatan. Kombinasi DM dan hipertensi menjadi faktor risiko yang berhubungan signifikan dengan kejadian CKD (Shabrina, et al., 2022). Perubahan gaya hidup modern serta proses penuaan penduduk juga berkontribusi terhadap peningkatan prevalensi hipertensi dan DM yang pada akhirnya meningkatkan

risiko CKD. Data Survei Kesehatan Indonesia (SKI) tahun 2023 menunjukkan bahwa prevalensi DM, hipertensi, dan penyakit ginjal kronik terus meningkat, terutama pada kelompok usia dewasa dan lanjut usia (Kementerian Kesehatan Republik Indonesia, 2023). Kondisi tersebut menunjukkan perlunya analisis data secara sistematis karena pola kunjungan pasien dengan penyakit kardiometabolik dan gangguan ginjal memiliki karakteristik klinis yang beragam. Oleh karena itu, pendekatan segmentasi seperti klasterisasi diperlukan untuk mengidentifikasi kelompok pasien dengan karakteristik klinis serupa sehingga mendukung perencanaan layanan kesehatan yang lebih tepat sasaran.

Berbagai studi sebelumnya telah menerapkan teknik klasterisasi *unsupervised* untuk segmentasi pasien. *Algoritma K-means* adalah metode klasterisasi tak berlabel yang sering digunakan untuk membagi data ke dalam klaster berdasarkan kesamaan antar data (Phuong, et al., 2025). Hasil literatur menunjukkan efektivitasnya dalam ranah klinis menggunakan *K-means* untuk mengelompokkan pasien DM menjadi lima subtipen fenotipik berdasarkan variabel klinis utama (Taurbekova, et al., 2025). Robertson et al. (2022) menemukan klaster multimorbiditas pada pasien rawat inap, termasuk klaster gabungan CKD dan DM yang menunjukkan pola penyerta penyakit ginjal dan DM. Metode *K-means* telah digunakan dalam berbagai penelitian di Indonesia untuk melakukan segmentasi pasien, misalnya klasterisasi pasien lansia berdasarkan tanda-tanda vital dan demografi menggunakan algoritma *K-means* (Afriany, et al., 2025). Secara keseluruhan, berbagai analisis klaster menekankan kemampuan *K-means* mengungkap struktur laten dalam data klinis pasien dan mendukung pendekatan perawatan yang lebih terpersonalisasi.

Berdasarkan kajian literatur, sebagian besar penelitian fokus pada satu penyakit kronis tunggal atau pola multimorbiditas umum. Oleh karena itu, penelitian ini bertujuan menerapkan algoritma *K-means* untuk mengelompokkan pasien dengan komorbiditas DM, hipertensi, dan gagal ginjal

kronik berdasarkan variabel klinis. Variabel seperti usia, tekanan darah, kadar glukosa darah dan hemoglobin dipilih karena merepresentasikan kondisi fisiologis utama yang berkaitan dengan perkembangan dan tingkat keparahan penyakit kronis tersebut. Usia digunakan sebagai indikator faktor risiko degeneratif, tekanan darah mencerminkan kondisi kardiovaskular, kadar glukosa darah menunjukkan kontrol metabolik pasien DM, sedangkan kadar hemoglobin berkaitan dengan gangguan fungsi ginjal dan kondisi anemia yang sering menyertai penyakit ginjal kronik. Pendekatan ini diharapkan dapat mengidentifikasi profil kelompok pasien dengan tingkat risiko komplikasi tertentu sehingga memberikan kontribusi dalam pengembangan strategi manajemen penyakit kronis yang lebih terarah dan berbasis data.

2. Metode

Penelitian ini menggunakan pendekatan kuantitatif dengan desain observasional *cross-sectional*, dimana analisis data dilakukan pada satu titik waktu tertentu. Teknik analisis yang digunakan adalah *K-means clustering*, yaitu metode pembelajaran tak-terawasi (*unsupervised learning*) yang populer karena implementasinya sederhana dan kompleksitas komputasi rendah. Metode *K-means* dipilih karena efektif dalam mengelompokkan data tak berlabel ke dalam beberapa klaster berdasarkan kemiripan atribut (Ikotun, et al., 2023). Dengan demikian, objek data dengan karakteristik serupa terkumpul dalam satu klaster, sedangkan antar klaster berbeda secara signifikan.

2.1 Metode Pengumpulan Data

Data yang digunakan dalam penelitian ini bersumber dari dataset publik yang diunduh dari *Kaggle* yaitu platform penyedia koleksi dataset untuk keperluan riset dan analisis data (Nafiyah & Hidayatullah, 2026). Dataset tersebut diunggah oleh akun *Co-learning Lounge* dan diakses pada bulan Mei tahun 2026 melalui tautan <https://www.kaggle.com/datasets/colearninglounge/chronic-kidney-disease>.

Variabel yang tersedia meliputi tekanan darah, kadar glukosa darah, albumin, hemoglobin, kreatinin serum, ureum darah, jumlah sel darah merah, serta berbagai parameter klinis lain yang relevan dalam identifikasi penyakit DM, hipertensi, dan gagal ginjal. Dataset tersebut kemudian dipersiapkan dengan proses pembersihan (*data cleansing*) untuk menghilangkan nilai duplikat, data hilang, atau kesalahan input sehingga siap dianalisis (Putra, 2023).

Setelah proses *data cleaning* dan seleksi variabel yang sesuai dengan tujuan penelitian, jumlah data yang memenuhi kriteria analisis menjadi 51 pasien. Jumlah sampel ini relatif terbatas untuk penelitian berbasis *machine learning* dan *data mining*, sehingga hasil klasterisasi pada penelitian ini lebih tepat diposisikan sebagai analisis eksploratif awal (*exploratory analysis*) dibandingkan model prediktif berskala besar. Meskipun demikian, penelitian klasterisasi berbasis *K-Means* pada data klinis tetap dapat dilakukan pada ukuran sampel terbatas apabila pola antar variabel menunjukkan pemisahan yang cukup jelas dan didukung oleh evaluasi validitas klaster, seperti indeks *Silhouette* dan *Dunn*. Namun, ukuran sampel yang kecil berpotensi memengaruhi stabilitas klaster dan membatasi generalisasi hasil terhadap populasi yang lebih luas. Oleh karena itu, interpretasi hasil penelitian dilakukan secara hati-hati dan memerlukan validasi lebih lanjut menggunakan dataset dengan ukuran yang lebih besar dan lebih beragam. Karena data bersifat sekunder dan anonim, penelitian ini menggunakan seluruh data yang memenuhi kriteria inklusi penelitian tanpa proses sampling tambahan dan tidak memerlukan persetujuan etik khusus.

2.2 Metode Analisis Data

Analisis data dilakukan dengan menerapkan algoritma *K-means* pada data yang telah dibersihkan. Langkah awal adalah menormalisasi atau mentransformasi data agar skala variabel menjadi seragam. Selanjutnya, ditentukan jumlah klaster (*k*) yang akan dibentuk sesuai dengan tujuan penelitian, kemudian dipilih secara acak titik pusat awal (*centroid*). Setiap literasi analisis

menghitung jarak *Euclidean* antara setiap titik data dan semua *centroid*, lalu setiap data di-*assign* ke klaster dengan jarak terpendek. Langkah ini diulang hingga pembagian klaster stabil (konvergen) (Fikria & Sriani, 2024). Setelah pengelompokan selesai, kualitas klaster dievaluasi menggunakan indeks validasi internal. Indeks seperti *Silhouette* dan *Dunn* digunakan sebagai tolok ukur, karena nilai indeks yang tinggi menunjukkan kohesi klaster yang baik (Novaldi & Wijayanto, 2023). Evaluasi ini memastikan jumlah klaster yang dipilih menghasilkan pengelompokan data yang tepat dan bermakna.

3. Hasil dan Pembahasan

3.1 Karakteristik Responden

Setelah proses pembersihan data, diperoleh 51 pasien dengan riwayat DM, hipertensi, dan/atau penyakit ginjal kronis. Rata-rata usia responden adalah 49,3 tahun dengan tekanan darah rata-rata 76,1 mmHg, kadar glukosa darah sewaktu rata-rata 133,9 mg/dL dan kadar hemoglobin rata-rata 13,9 g/dL. Dari sisi komorbiditas, 7 pasien (13,7%) memiliki DM, 11 pasien (21,6%) hipertensi, 5 pasien (9,8%) penyakit jantung koroner dan 8 pasien (15,7%) anemia. Data tersebut dirangkum pada Tabel 1 yang memberikan gambaran distribusi karakteristik klinis responden secara keseluruhan. Hasil ini menunjukkan heterogenitas profil klinis sampel, misalnya kadar glukosa dan hemoglobin yang beragam, yang menjadi alasan penggunaan metode klasterisasi untuk mengidentifikasi pola tersembunyi.

Tabel 1. Karakteristik responden (n=51).

Variabel	Nilai
Usia (tahun), rata-rata (SD)	49,3 ± 13,7
Tekanan darah sistolik (mmHg), rata-rata (SD)	76,1 ± 11,8
Glukosa darah sewaktu (mg/dL), rata-rata (SD)	133,9 ± 82,8
Hemoglobin (g/dL), rata-rata (SD)	13,9 ± 2,9
Diabetes Melitus (%)	13,7%

Hipertensi (%)	21,6%
Penyakit jantung koroner (%)	9,8%
Anemia (%)	15,7%

Sumber : Data primer (rekam medis pasien).

Berdasarkan Tabel 1, terlihat bahwa sebagian besar pasien relatif berusia menengah dengan komplikasi komorbid yang bervariasi. Kehadiran DM dan hipertensi relatif rendah (<25%), sedangkan kadar glukosa darah dan anemia menunjukkan keragaman yang cukup besar. Hasil ini menegaskan adanya heterogenitas profil klinis pasien, sesuai dengan latar belakang bahwa kombinasi DM, hipertensi, dan CKD menciptakan beban penyakit yang kompleks.

3.2 Hasil Klasterisasi dan Diskusi

Analisis klaster dilakukan dengan algoritma *K-Means* setelah variabel dinormalisasi. Pengujian untuk menentukan jumlah klaster optimal (k) meliputi perhitungan *Silhouette* dan indeks *Dunn*. Nilai *Silhouette* tertinggi diperoleh pada k=3 (0,537), sedangkan indeks *Dunn* juga relatif tinggi (0,612) pada k=3, yang menunjukkan pemisahan klaster yang baik. Dengan demikian, dipilih tiga klaster pasien. Hasil klasterisasi menghasilkan tiga kelompok dengan karakteristik klinis yang berbeda. Tabel 2 menyajikan ringkasan statistik utama tiap klaster.

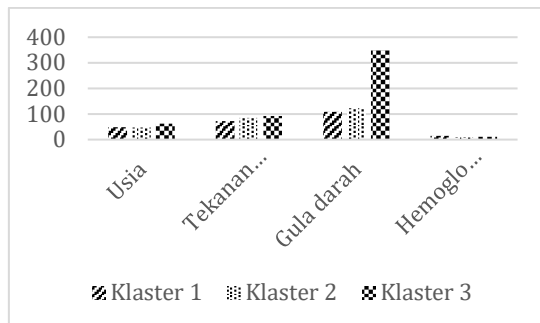
Tabel 2. Karakteristik klinis dan komorbiditas tiap klaster pasien.

Variabel	Klaster 1	Klaster 2	Klaster 3
Usia rata-rata (tahun)	47,9 ± 14,0	47,5 ± 11,5	62,8 ± 5,9
Tekanan darah rata-rata (mmHg)	72,1 ± 8,4	85,0 ± 13,1	92,0 ± 13,0
Glukosa darah sewaktu rata-rata (mg/dL)	108,1 ± 19,8	122,1 ± 64,3	348,8 ± 102,5

Hemoglobin rata-rata (g/dL)	15,3 ± 1,5	9,14 ± 1,31	10,72 ± 2,42
Persentase DM	0,0%	37,5%	80,0%
Persentase Hipertensi	0,0%	75,0%	100,0%
Persentase Anemia	0,0%	87,5%	20,0%
Persentase sel darah merah normal	100,0%	50,0%	40,0%

Sumber : Analisis *K-Means* (data rekam medis).

Untuk memperjelas perbedaan karakteristik klinis antar klaster, visualisasi perbandingan nilai rata-rata variabel utama ditampilkan pada Gambar 1. Grafik ini menunjukkan distribusi usia, tekanan darah, kadar glukosa darah, dan hemoglobin pada masing-masing klaster pasien.



Gambar 1. Perbandingan karakteristik klinis antar klaster pasien

Pada Gambar 1 terlihat perbedaan yang cukup jelas antar klaster. Klaster 3 memiliki nilai glukosa darah yang jauh lebih tinggi dibandingkan klaster lainnya, yang menunjukkan dominasi gangguan metabolik berat. Selain itu, tekanan darah pada klaster 3 juga paling tinggi, memperkuat indikasi risiko hipertensi yang signifikan. Sebaliknya, klaster 1 menunjukkan nilai yang relatif stabil dan mendekati normal pada seluruh variabel, mencerminkan kelompok dengan risiko rendah. Klaster 2 berada di antara keduanya,

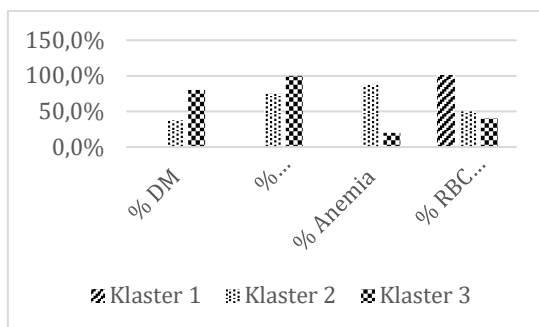
namun memiliki nilai hemoglobin paling rendah, yang mengindikasikan prevalensi anemia yang tinggi pada kelompok ini. Pola ini konsisten dengan hasil deskriptif pada Tabel 2 dan memperkuat interpretasi bahwa masing-masing klaster memiliki karakteristik klinis yang berbeda secara signifikan.

Karakterisasi klaster. Klaster 1 (n=38) terdiri dari pasien relatif muda (rata-rata 47,9 tahun) dengan tekanan darah terendah (72 mmHg) dan glukosa sewaktu normal (108 mg/dL). Tidak ada kasus DM atau hipertensi di klaster ini, dan kadar hemoglobin tertinggi (~15,3 g/dL). Profil ini mencerminkan kelompok pasien dengan kondisi kesehatan relatif baik (low-risk). Sebaliknya, Klaster 3 (n=5) berisi pasien tertua (rata-rata 62,8 tahun) dengan hipertensi 100% dan DM 80%. Klaster ini memiliki glukosa darah sewaktu sangat tinggi (rata-rata 348,8 mg/dL), serta hemoglobin menurun (10,7 g/dL), menunjukkan profil komorbiditas metabolik yang berat. Klaster 2 (n=8) berada di antara: pasien usia menengah (47,5 tahun) dengan tekanan darah moderat (85 mmHg), proporsi hipertensi 75%, DM 37,5%, namun kadar hemoglobin rendah (9,14 g/dL) dan anemia mencapai 87,5%. Karakteristik tersebut mengindikasikan adanya kondisi klinis yang lebih berat dibandingkan klaster 1, terutama ditinjau dari tingginya prevalensi hipertensi dan anemia serta rendahnya kadar hemoglobin pada kelompok pasien ini. Namun, karena penelitian ini tidak menggunakan indikator fungsi ginjal lanjutan seperti laju filtrasi glomerulus (eGFR), kreatinin serum, maupun albumin urin sebagai variabel analisis utama, interpretasi terkait tingkat keparahan kerusakan ginjal tidak dapat disimpulkan secara definitif.

Meskipun pemilihan k=3 menghasilkan nilai *Silhouette* dan indeks *Dunn* terbaik, distribusi anggota antar klaster masih menunjukkan ketidakseimbangan, khususnya pada Klaster 2 (n=8) dan Klaster 3 (n=5). Kondisi ini perlu diperhatikan karena ukuran klaster yang kecil dapat meningkatkan kemungkinan bahwa kelompok yang terbentuk merepresentasikan pola ekstrem atau *outlier*, bukan subpopulasi klinis yang sepenuhnya stabil dan representatif. Namun

demikian, klaster kecil pada analisis *K-Means* tidak selalu menunjukkan kesalahan pemodelan. Dalam penelitian ini, klaster 3 memperlihatkan pola klinis yang konsisten, yaitu dominasi hipertensi, DM, serta kadar glukosa darah yang sangat tinggi dibandingkan klaster lain. Pemisahan karakteristik yang cukup jelas tersebut didukung oleh nilai *Silhouette* dan indeks *Dunn* yang relatif baik, sehingga mengindikasikan bahwa klaster yang terbentuk memiliki tingkat separasi yang memadai dan bukan semata-mata hasil pengelompokan acak. Meskipun demikian, karena jumlah anggota klaster 3 relatif kecil, interpretasi hasil lebih tepat diposisikan sebagai temuan eksploratif awal (*exploratory finding*) dan belum dapat digeneralisasikan secara luas terhadap populasi pasien penyakit ginjal kronis secara keseluruhan. Oleh karena itu, penelitian lanjutan dengan ukuran sampel lebih besar dan distribusi data yang lebih seimbang diperlukan untuk mengevaluasi stabilitas klaster yang terbentuk.

Untuk memperjelas distribusi komorbiditas pada masing-masing klaster, Gambar 2 menyajikan perbandingan persentase DM, hipertensi, anemia, dan kondisi sel darah merah normal pada setiap kelompok pasien.



Gambar 2. Perbandingan persentase komorbiditas antar klaster pasien

Berdasarkan Gambar 2, terlihat bahwa Klaster 3 memiliki prevalensi DM dan hipertensi tertinggi dibandingkan klaster lainnya, masing-masing mencapai 80% dan 100%. Hal ini menunjukkan bahwa klaster 3 merupakan kelompok dengan beban komorbiditas metabolik yang paling berat.

Sebaliknya, klaster 1 tidak menunjukkan adanya kasus DM maupun hipertensi, serta memiliki persentase sel darah merah normal yang mencapai 100%, yang mengindikasikan kondisi klinis relatif stabil. Klaster 2 menunjukkan karakteristik yang berbeda, yaitu prevalensi anemia yang paling tinggi (87,5%) dengan persentase sel darah merah normal yang lebih rendah. Pola ini mengindikasikan adanya gangguan hematologis yang dominan pada klaster 2. Secara keseluruhan, visualisasi ini memperkuat hasil analisis pada Tabel 2 bahwa setiap klaster memiliki profil komorbiditas yang berbeda secara signifikan, sehingga mendukung interpretasi segmentasi risiko klinis pasien.

Hasil ini konsisten dengan penelitian sebelumnya bahwa klasterisasi *K-Means* mampu mengungkap subkelompok pasien dengan tingkat risiko berbeda. Penelitian yang dilakukan Wang et al. (2023) menemukan empat subklaster pasien DM dengan label *low* hingga *high-risk*, dimana klaster risiko tinggi memiliki komplikasi signifikan. Demikian pula, klaster bergaya “metabolik berat” (klaster 3) pada hasil penelitian ini setara dengan profil risiko tinggi tersebut. Studi Gong et al. (2026) juga menunjukkan *K-Means* berhasil membagi pasien gagal jantung kronik menjadi dua kelompok (CKD vs non-CKD) dengan indikator klinis berbeda. Temuan ini sejalan dengan literatur bahwa klaster *K-Means* dapat mengidentifikasi pola klinis laten yang relevan. Taurbekova et al (2025) melaporkan bahwa subtype *Severe Insulin-Resistant Diabetes* (SIRD) dalam pengelompokan DM ditandai oleh komplikasi ginjal yang luas dan risiko CKD lebih tinggi, sama seperti pengelompokan pasien pada penelitian ini yang menunjukkan klaster berprofil ginjal parah.

Segmentasi pasien berperan penting dalam mendukung stratifikasi risiko dan perencanaan intervensi kesehatan yang lebih terarah. Pengelompokan pasien berdasarkan kesamaan karakteristik klinis memungkinkan tenaga kesehatan mengidentifikasi kelompok dengan risiko tinggi serta menyusun strategi perawatan yang sesuai dengan kondisi masing-masing kelompok. Literatur menunjukkan bahwa klasterisasi MLTC

(*Multiple Long-Term Conditions*) dapat mendukung proses stratifikasi risiko, pengembangan program pencegahan yang lebih tepat sasaran, serta perancangan layanan kesehatan yang terintegrasi. Menurut studi Beaney *et al.*, (2026), pengelompokan ini dapat membimbing strategi perawatan DM dan CKD yang spesifik: kelompok dengan profil metabolik berat (klaster 3) mungkin memerlukan penanganan glukosa darah dan hipertensi yang agresif, sedangkan kelompok sehat (klaster 1) bisa fokus pada pencegahan primer. Keseluruhan hasil ini memperkuat pandangan bahwa analisis klaster tak-terawasi seperti *K-Means* efektif mengungkap subgrup pasien dengan karakteristik klinis khas yang dapat dihubungkan dengan tingkat risiko komplikasi tertentu (Beaney, et al., 2026). Dengan demikian, hasil penelitian ini menunjukkan bahwa pendekatan klusterisasi berbasis data klinis dapat digunakan untuk mendukung identifikasi kelompok risiko pasien sebagai dasar perencanaan layanan kesehatan yang lebih terarah.

4. Kesimpulan dan Saran

Pengelompokan pasien menggunakan metode *K-Means* menghasilkan tiga klaster dengan karakteristik klinis yang berbeda, yaitu 1) klaster risiko rendah, 2) klaster dengan dominasi anemia dan hipertensi, serta 3) klaster dengan profil metabolik berat yang ditandai prevalensi DM dan hipertensi tinggi. Hasil penelitian menunjukkan bahwa metode *K-Means* mampu mengidentifikasi segmentasi pasien berdasarkan kesamaan karakteristik klinis. Namun, ukuran sampel yang terbatas dan distribusi klaster yang tidak seimbang menyebabkan hasil penelitian ini masih bersifat eksploratif dan memerlukan validasi lebih lanjut menggunakan dataset yang lebih besar dan beragam.

Penelitian selanjutnya disarankan menggunakan dataset dengan jumlah sampel yang lebih besar dan distribusi data yang lebih seimbang agar stabilitas klaster dan generalisasi hasil dapat meningkat. Selain itu, penambahan variabel klinis yang lebih komprehensif, seperti parameter fungsi ginjal lanjutan dan faktor gaya hidup, diperlukan untuk menghasilkan segmentasi pasien yang

lebih akurat. Penggunaan metode klusterisasi lain atau pengembangan algoritma *K-Means* juga dapat dipertimbangkan untuk meningkatkan kualitas pengelompokan data.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada pihak-pihak yang telah memberikan kontribusi nyata dalam pelaksanaan penelitian ini. Terima kasih disampaikan kepada penyedia dataset melalui platform Kaggle yang telah menyediakan data terbuka sehingga penelitian ini dapat dilakukan dengan baik. Penulis juga mengapresiasi kontribusi para peneliti dan pengembang di bidang data mining yang telah mengembangkan metode dan referensi ilmiah yang menjadi dasar dalam penelitian ini. Selain itu, terima kasih kepada rekan-rekan yang telah membantu dalam proses pengolahan data, analisis, serta penyusunan naskah, sehingga penelitian ini dapat terselesaikan dengan optimal.

Daftar Pustaka

- Afriany, R., Sinaga, R. & Samsinar, 2025. Segmentasi pasien berbasis K-Means dari tanda-tanda vital dan demografi: pendekatan unsupervised learning untuk profil risiko klinis. *Jurnal Ilmiah Sistem Informasi dan Ilmu Komputer*, 5(3), pp. 638-650.
- Agatha, C. D., Mukti, A. I. & Girsang, E., 2025. Prevalensi penyakit gagal ginjal kronik di kalangan usia produktif akibat gaya hidup di Rumah Sakit Royal Prima Medan. *Jurnal Ilmiah Indonesia*, 10(8), pp. 6377-6386.
- Aldhyani, T. H., Alshebami, A. S. & Alzahrani, M. Y., 2020. Soft clustering for enhancing the diagnosis of chronic diseases over machine learning algorithms. *Journal of Healthcare Engineering*, Volume 2020, pp. 1-16.
- Beaney, T. et al., 2026. Clustering in multiple long-term conditions: methodological

Jurnal Kesehatan

Author(s) : Fitra Tri Damayanti¹, Muhammad Iffran Ceria Rizal¹, Sakinah Derajad¹, Ihksan Kurnia Afnis¹,
Joko Purnomo¹, Fitriah², Khairunnisa¹

- and translational challenges and solutions. *Journal of the Royal Society of Medicine*, 119(1), pp. 5-9.
- Fikria, R. & Sriani, 2024. Analisis metode K-Means clustering dalam pengelompokan penjualan sembako. *Journal of Information System Research (JOSH)*, 5(4), pp. 1464-1471.
- Gong, L., Hu, Z., Lin, X. & Chen, W., 2026. Unsupervised K-Means clustering identified routine clinical biomarkers for classifying chronic kidney disease in patients with chronic heart failure. *Laboratory Medicine*, 57(1), pp. 82-91.
- Ikotun, A. M. et al., 2023. K-Means clustering algorithms: a comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, Volume 622, pp. 178-210.
- Kementerian Kesehatan Republik Indonesia, 2023. *Survei Kesehatan Indonesia (SKI) 2023*, Jakarta: Kementerian Kesehatan RI.
- Nafiiyah, N. & Hidayatullah, A. T., 2026. Analisis penyakit diabetes dengan support vector machine (data Kaggle). *Jurnal Sistem Informasi dan Informatika (Simika)*, 9(1), pp. 25-33.
- Novaldi, J. & Wijayanto, A. W., 2023. Analisis cluster kualitas pemuda di Indonesia pada tahun 2022 dengan agglomerative hierarchical dan K-Means. *Komputika: Jurnal Sistem Komputer*, 12(2), pp. 211-219.
- Phuong, N. D. et al., 2025. Machine learning techniques in chronic kidney diseases: a comparative study of classification model performance. *Bioinformatics and Biology Insights*, Volume 19, pp. 1-18.
- Putra, I. K. N., 2023. Implementasi metode K-Means clustering pada ekspor kopi berdasarkan negara tujuan. *Jurnal Ilmu Komputer dan Bisnis (JIKB)*, 14(1), pp. 45-55.
- Robertson, L. et al., 2022. Identifying multimorbidity clusters in an unselected population of hospitalised patients. *Scientific Reports*, Volume 12, pp. 1-10.
- Shabrina, S. A., Saftarina, F. & Pramesona, B. A., 2022. Faktor risiko penyakit ginjal kronik pada pasien diabetes. *JK Unila*, 6(2), pp. 58-62.
- Taurbekova, B. et al., 2025. Cluster analysis in diabetes research: a systematic review enhanced by a cross-sectional study. *Journal of Clinical Medicine*, 14(10), pp. 1-22.
- Wang, C. et al., 2023. Unsupervised cluster analysis of clinical and metabolite characteristics in patients with chronic complications of T2DM: an observational study of real data. *Frontiers in Endocrinology*, Volume 14, pp. 1-12.