

---

## ANALISIS ALGORITMA K-MEANS CLUSTERING DALAM IDENTIFIKASI TINGKAT RISIKO PENYAKIT BERDASARKAN DATA REKAM MEDIS PASIEN

Wina Aulia<sup>1</sup>, Andysah Putera Utama Siahaan<sup>2</sup>, Leni Marlina<sup>3</sup>,  
Khairul<sup>4</sup>, Muhammad Iqbal<sup>5</sup>

Universitas Pembangunan Panca Budi, Medan

e-mail: <sup>1</sup>winaaulia397@gmail.com, <sup>2</sup>andiesiahaan@gmail.com,

<sup>3</sup>khairul@dosen.pancabudi.ac.id, <sup>4</sup>lenimarlina@dosen.pancabudi.ac.id,

<sup>5</sup>muhammadiqbal@dosen.pancabudi.ac.id

**Abstract:** *This study aims to classify patients' health conditions based on six indicators: systolic blood pressure, diastolic blood pressure, fasting blood glucose, normal blood glucose, cholesterol level, and uric acid level using the K-Means Clustering method. The optimal number of clusters was determined using the Elbow Method and Silhouette Score, which resulted in five as the optimal number of clusters. The results show that the manual approach produces a more stable distribution that closely aligns with the clinical interpretation of cluster categories: Healthy (C1), Safe (C2), Alert (C3), Moderate (C4), and Severe (C5). Visualization was performed for each indicator through scatter plots and color mapping against normal value thresholds, aiding in the understanding of the distribution of patient conditions across clusters. The analysis reveals that even if a patient has one or more indicators within normal limits, they are not automatically classified into the Healthy or Safe clusters. Discrepancies in other indicators can place them in higher-risk clusters such as Alert, Moderate, or Severe. Therefore, this clustering approach provides a comprehensive view of health conditions based on a combination of features, rather than a single parameter. This research is useful in supporting early diagnosis and data-driven decision-making processes and can be integrated into health information systems for automatic risk classification of patient populations.*

**Keywords:** *K-Means, Clustering, Health, Blood Pressure, Blood Glucose, Cholesterol, Uric Acid, Data Visualization*

**Abstrak:** Penelitian ini bertujuan untuk mengelompokkan kondisi kesehatan pasien berdasarkan enam indikator, yaitu tekanan darah sistolik, tekanan darah diastolik, kadar gula puasa, kadar gula normal, kadar kolesterol, dan kadar asam urat menggunakan metode K-Means Clustering. Penentuan jumlah kluster optimal dilakukan dengan metode Elbow dan Silhouette Score, yang menghasilkan lima kluster sebagai jumlah optimal. Hasil menunjukkan bahwa pendekatan manual menghasilkan distribusi yang lebih stabil dan mendekati pemaknaan klinis dari kategori kluster, yaitu: Sehat (C1), Aman (C2), Waspada (C3), Sedang (C4), dan Berat (C5). Visualisasi dilakukan untuk setiap indikator melalui scatter plot dan pemetaan warna terhadap batas nilai normal, yang membantu dalam memahami sebaran kondisi pasien pada masing-masing kluster. Hasil analisis menunjukkan bahwa meskipun seorang pasien memiliki satu atau lebih indikator dalam batas normal, tidak secara otomatis tergolong dalam kluster Sehat atau Aman. Ketidaksiapan pada indikator lainnya dapat menempatkan pasien ke dalam kluster yang lebih tinggi risikonya, seperti Waspada, Sedang, atau Berat. Oleh karena itu, pendekatan clustering ini memberikan gambaran menyeluruh terhadap kondisi kesehatan berdasarkan kombinasi fitur, bukan hanya pada satu parameter. Penelitian ini bermanfaat untuk mendukung proses diagnosis awal dan pengambilan keputusan berbasis data, serta dapat diintegrasikan dalam sistem informasi kesehatan untuk klasifikasi risiko populasi pasien secara otomatis.

**Kata kunci:** K-Means, Clustering, Kesehatan, Tekanan Darah, Gula Darah, Kolesterol, Asam Urat, Visualisasi Data

## PENDAHULUAN

Perkembangan teknologi informasi telah membawa dampak signifikan terhadap transformasi digital di berbagai sektor (Siahaan & Afandi Syahputra, 2024)(Marlina et al., 2016), termasuk bidang kesehatan. Digitalisasi sistem rekam medis memungkinkan penyimpanan dan pengolahan data pasien dalam jumlah besar secara sistematis. Hal ini tidak hanya meningkatkan efisiensi layanan medis, tetapi juga membuka peluang baru dalam pemanfaatan data untuk mendukung pengambilan keputusan klinis yang lebih tepat sasaran (data-driven decision making). Informasi yang tersimpan dalam rekam medis digital mencerminkan kondisi fisiologis pasien dan dapat dianalisis lebih lanjut untuk menilai risiko kesehatan secara lebih komprehensif.

Penyakit tidak menular seperti hipertensi, diabetes, dan gangguan metabolik lainnya masih menjadi permasalahan utama karena prevalensinya yang terus meningkat. Kondisi ini menuntut adanya upaya deteksi dini yang lebih efektif, agar komplikasi serius dapat dicegah sedini mungkin. Salah satu pendekatan analitis yang relevan untuk membantu proses ini adalah teknik data clustering, khususnya algoritma K-Means Clustering. Algoritma ini mampu mengelompokkan pasien ke dalam kategori-kategori risiko berdasarkan kesamaan karakteristik pada indikator seperti tekanan darah sistolik dan diastolik, kadar gula darah (puasa dan sesudah makan), kolesterol, dan asam urat.

Pemanfaatan K-Means dalam pengelolaan data kesehatan memiliki potensi besar dalam mendukung pengembangan sistem pendukung keputusan medis (Iqbal et al., 2016). Dengan mengelompokkan pasien berdasarkan tingkat risiko penyakit,

institusi kesehatan dapat melakukan langkah preventif yang lebih tepat sasaran. Hal ini berdampak pada efisiensi penggunaan sumber daya medis, serta peningkatan kualitas pelayanan terhadap pasien yang masuk dalam kelompok rentan.

Sejumlah studi sebelumnya telah menunjukkan efektivitas algoritma ini dalam dunia medis. (Laksono et al., 2024) mengelompokkan data pasien rawat inap berdasarkan usia dan diagnosis penyakit, yang menghasilkan kluster distribusi penyakit pada kelompok usia tertentu. (Wala et al., 2024) menerapkan K-Means pada data pasien penyakit jantung, menghasilkan strategi perawatan yang lebih personal. (Danny & Muhidin, 2023) menunjukkan bagaimana clustering dapat digunakan untuk mengidentifikasi indikator awal risiko diabetes. Selain itu, (Setiawati et al., 2023) dalam prosidingnya berhasil mengklasifikasikan penyakit diabetes berdasarkan diagnosis tingkat keparahan penyakit menggunakan algoritma logistik regresi. Studi lain oleh (Hartini et al., 2025) menegaskan bahwa teknik K-Means clustering bermanfaat dalam klasifikasi diagnosis penyakit lambung.

Meskipun demikian, efektivitas K-Means sangat bergantung pada penentuan pusat kluster awal (initial centroids). Pemilihan secara acak dapat menyebabkan hasil yang tidak konsisten atau kurang akurat. Oleh karena itu, pendekatan manual berbasis pengetahuan domain mulai dipertimbangkan sebagai alternatif untuk meningkatkan kualitas pengelompokan. Sayangnya,

perbandingan menyeluruh antara pendekatan acak dan manual dalam pengolahan data rekam medis belum banyak dieksplorasi secara mendalam. Melalui penelitian ini, dilakukan analisis dan evaluasi terhadap hasil pengelompokan pasien menggunakan pendekatan K-Means baik dengan pusat

klaster acak maupun manual. Visualisasi distribusi data juga dilakukan untuk mengamati sejauh mana pola pengelompokan mampu merepresentasikan tingkat risiko kesehatan. Diharapkan, hasil penelitian ini dapat memberikan kontribusi nyata dalam pengembangan sistem analisis kesehatan berbasis data science, serta mendorong implementasi machine learning yang lebih luas dalam sistem rekam medis digital di Indonesia.

## METODE

Pengelompokan jenis penyakit berdasarkan data Rekam Medis pasien memiliki peran krusial dalam meningkatkan keakuratan dan kecepatan diagnosis. *Algoritma K-Means Clustering* telah terbukti efektif sebagai metode untuk menganalisis Data Rekam Medis. Bab ini mengulas dasar teori di balik penggunaan algoritma *K-Means Clustering* untuk tujuan pengelompokan data rekam medis pasien.

*K-Means Clustering* adalah sebuah metode yang digunakan untuk mengelompokkan data menjadi beberapa bagian berdasarkan jarak, yang hanya dapat bekerja pada atribut numerik. Metode ini termasuk dalam kategori clustering partitioning, yang memisahkan data ke dalam wilayah yang berbeda. *K-Means* merupakan metode non-hierarkis di mana pusat kluster dipilih secara acak dari kumpulan data dan setiap data ditempatkan ke kluster terdekat berdasarkan jarak minimum. Posisi pusat kluster diperbarui iteratif sampai semua data diklasifikasikan ke dalam kluster-klusternya, menciptakan posisi pusat kluster yang baru (Febrianto et al., 2021).

Menurut (Rahmawati et al., 2016), Algoritma *K-Means Clustering* adalah metode analisis data atau data mining yang bersifat tanpa supervisi (*unsupervised*), di mana tujuannya adalah untuk meminimalkan variasi di dalam kluster tertentu dan memaksimalkan variasi antar kluster lainnya.

## Penelitian Terkait

Penelitian yang relevan berisi penelitian sebelumnya yang berkaitan dengan tema yang diteliti. Adapun sumber penelitian relevan berupa jurnal tentang Analisa Algoritma K-Means Clustering untuk pengelompokan Data Rekam Medis pasien berdasarkan jenis penyakit. Berikut hasil penelitian yang relevan:

**Tabel 1 Penelitian Terkait**

No.	Nama	Judul	Metode
1	(Ordila et al., 2020)	Penerapan Data Mining untuk Pengelompokan Data Rekam Medis Pasien Berdasarkan Jenis Penyakit dengan Algoritma Clustering (Studi Kasus: Poli Klinik PT. Inecda)	Metode <i>K-Means Clustering</i>
2	(Agustina & Danar Dana, 2024)	Implementasi K-Means Clustering untuk Pengelompokan Tingkat Kemiskinan di Pulau Jawa Menggunakan Bahasa Pemrograman Python	Algoritma <i>K-Means Clustering</i>
3	(Bahri et al., 2017)	Analisis Pemetaan Tingkat Pengangguran di Pulau Jawa dan Bali dengan	Metode <i>K-Means</i> .

		Metode K-Means	
4	(Manalu & Gunadi, 2022)	Implementasi Metode Data Mining K-Means Clustering terhadap Data Pembayaran Transaksi Menggunakan Bahasa Pemrograman Python pada CV Digital Dimensi	Data Mining K-Means Clustering
5	(Yolanda & Suhardi, 2023)	Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Data Pasien Rehabilitasi Narkoba	Algoritma K-Means Clustering
6	(Maesroh et al., 2024)	Penerapan Algoritma K-Means Clustering pada Pengelompokan Daerah Penyebaran Diare di Provinsi Jawa Barat	Algoritma K-Means Clustering

### Metode Analisis Data

Metode analisis data dalam penelitian ini dirancang untuk mengelompokkan pasien berdasarkan data rekam medis menggunakan algoritma K-Means Clustering. Proses analisis dilakukan secara bertahap, dimulai dari pra-pemrosesan data yang mencakup pembersihan data, seleksi atribut, dan transformasi data agar siap dianalisis. Data seperti jenis kelamin, umur, tekanan

darah, kadar gula, kolesterol, dan asam urat dipilih sebagai fitur, sementara data non-numerik seperti nama dan diagnosa diabaikan. Normalisasi data juga dilakukan agar setiap fitur memiliki skala yang setara. Selanjutnya, jumlah kluster optimal ditentukan menggunakan metode Elbow dan Silhouette Score, yang membantu mengevaluasi seberapa baik data terkelompok. Setelah nilai K ditentukan, algoritma K-Means diterapkan dengan langkah-langkah seperti inisialisasi centroid, penghitungan jarak Euclidean, pengelompokan data, dan pembaruan centroid secara iteratif. Hasil clustering dievaluasi menggunakan Silhouette Score dan divisualisasikan dengan PCA untuk melihat distribusi kluster secara dua dimensi. Akhirnya, tiap kluster dianalisis berdasarkan ciri medis dominan, lalu diberi label seperti Sehat, Aman, Waspada, Sedang, dan Berat. Pemberian label ini bukan berdasarkan diagnosis medis, melainkan berdasarkan kesamaan nilai-nilai fitur yang terdeteksi oleh algoritma.

### Klasterisasi

Klasterisasi merupakan inti dari analisis dalam penelitian ini, yang bertujuan mengelompokkan pasien berdasarkan kesamaan indikator kesehatan seperti tekanan darah, kadar gula, kolesterol, dan asam urat, menggunakan algoritma K-Means Clustering. Tujuan utamanya bukan untuk mendiagnosis penyakit secara langsung, melainkan untuk mengidentifikasi tingkat risiko kesehatan pasien. Proses klasterisasi diawali dengan menentukan jumlah kluster optimal menggunakan metode Elbow, yang dalam penelitian ini menghasilkan lima kluster ( $K = 5$ ), masing-masing mewakili tingkat risiko dari sangat ringan hingga sangat berat. Setelah jumlah kluster ditetapkan, centroid awal dipilih secara acak dari dataset, lalu proses iteratif dilakukan dengan menghitung jarak tiap data ke centroid menggunakan Euclidean Distance, mengelompokkan data berdasarkan jarak terdekat, dan

memperbarui posisi centroid hingga hasil stabil. Hasil klasterisasi ini diharapkan dapat memberikan wawasan kolektif terhadap pola kesehatan pasien

dan mendukung strategi preventif dan promotif rumah sakit. Berikut adalah tabel representasi awal dari nilai centroid pada masing-masing klaster:

**Tabel 2 Penentuan nilai centroid awal**

Klaster	Sistolik	Diastolik	Kadar Gula Puasa	Kadar Gula Normal	Kolesterol	Asam Urat
C1 (Sehat)	110	70	80	90	150	4.0
C2 (Aman)	120	80	100	110	180	5.0
C3 (Waspada)	135	90	125	140	200	6.5
C4 (Sedang)	150	100	150	160	230	7.5
C5 (Berat)	170	110	180	200	260	9.0

Centroid menjadi acuan awal dalam proses pembentukan klaster yang kemudian diperbarui dalam setiap iterasi berdasarkan rata-rata nilai anggota klaster.

## HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil implementasi metode K-Means Clustering terhadap data rekam medis pasien yang telah melalui tahap prapemrosesan sebelumnya. Analisis dilakukan dengan menggunakan perangkat lunak Jupyter Notebook dan Google Colab sebagai media komputasi. Fokus utama dalam bab ini adalah menampilkan hasil pengelompokan yang terbentuk berdasarkan fitur-fitur medis yang telah ditentukan, seperti tekanan darah, kadar gula, kolesterol, dan asam urat. Selain itu, disajikan pula evaluasi terhadap kualitas klasterisasi serta interpretasi karakteristik dari masing-masing klaster. Penjabaran

hasil ini bertujuan untuk mengidentifikasi pola-pola kesehatan tertentu dalam kelompok pasien, sehingga dapat digunakan sebagai dasar pertimbangan dalam pengambilan keputusan medis dan penyusunan strategi tindakan preventif terhadap pasien dengan kondisi serupa.

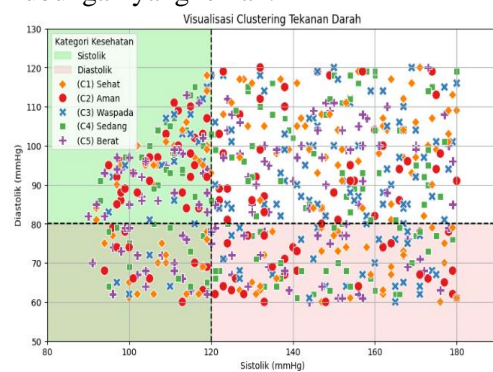
### Hasil Clustering

Pada bagian ini, dibahas hasil clustering menggunakan algoritma K-Means pada data rekam medis pasien untuk mengelompokkan pasien berdasarkan fitur medis seperti tekanan darah, kadar gula, kolesterol, dan asam urat. Hasil clustering ini menghasilkan klaster yang mencerminkan tingkat risiko penyakit pasien. Untuk menganalisis hasilnya, dilakukan visualisasi menggunakan scatter plot yang menunjukkan hubungan antara dua fitur utama, seperti sistolik dan diastolik, atau kombinasi fitur lain. Setiap titik pada scatter plot mewakili data pasien, dengan koordinat yang menggambarkan nilai

sistolik dan diastolik, memberikan gambaran distribusi data tekanan darah pasien dalam dataset.

### *Distribusi Tekanan Darah Sistolik dan Diastolik*

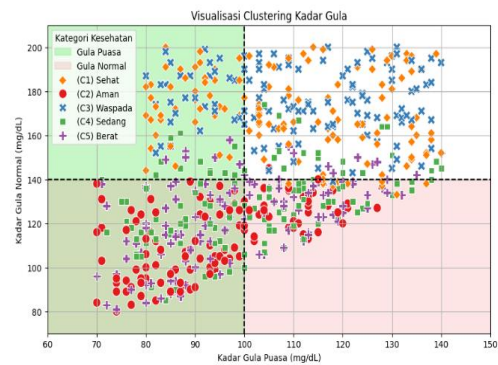
Penelitian ini menganalisis tekanan darah sistolik dan diastolik pasien sebagai indikator utama untuk menentukan kondisi kesehatan, terutama terkait risiko hipertensi atau masalah kardiovaskular. Scatter plot digunakan untuk menunjukkan korelasi antara kedua variabel kontinu ini. Grafik ini membantu mengidentifikasi pola atau kecenderungan antara tekanan darah sistolik dan diastolik, di mana titik-titik yang membentuk pola garis atau kelompok menunjukkan hubungan yang kuat, sementara sebaran acak menunjukkan hubungan yang lemah.



**Gambar 1 Visualisasi clustering tekanan darah**

### *Distribusi Kadar Gula Puasa dan Kadar Gula Normal*

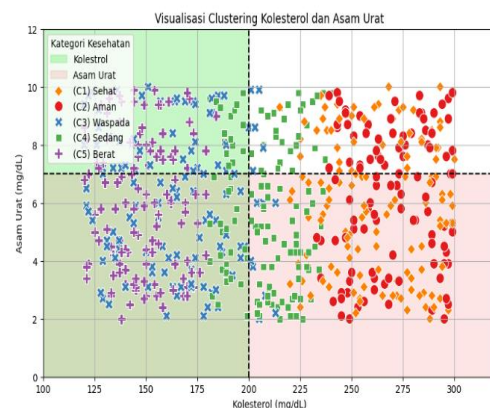
Komponen berikut yang dianalisis adalah kadar gula puasa dan kadar gula normal pasien, yang penting untuk menentukan risiko penyakit seperti diabetes. Pemahaman tentang hubungan kedua kadar gula ini sangat penting untuk mengetahui pola kesehatan pasien. Visualisasi pada Gambar 4.4 menunjukkan sebaran data dan apakah klaster-klaster yang terbentuk memiliki pola konsisten dalam kadar gula puasa dan normal, memberikan wawasan lebih dalam mengenai keterkaitan antara kondisi gula darah pasien dan tingkat risiko penyakit yang mereka miliki.



**Gambar 2 Visualisasi clustering kadar gula**

### *Distribusi Kadar Kolesterol dan Asam Urat*

Kolesterol dan asam urat adalah dua komponen penting yang menggambarkan kesehatan metabolik pasien, di mana kadar kolesterol tinggi dan asam urat berlebih dapat menjadi indikator risiko penyakit jantung, hipertensi, dan gout. Visualisasi dengan box plot bertujuan untuk menggambarkan variasi dan sebaran kadar kolesterol dan asam urat dalam data pasien serta memeriksa distribusinya di setiap klaster. Box plot ini memungkinkan analisis perbedaan kadar kedua variabel antara klaster-klaster dan menunjukkan apakah kadar kolesterol dan asam urat yang lebih tinggi sering ditemukan pada klaster dengan risiko penyakit yang lebih tinggi, memberikan wawasan lebih lanjut tentang faktor-faktor yang mempengaruhi klasifikasi risiko penyakit.



**Gambar 3 Visualisasi clustering kolesterol dan asam urat**

### Perbandingan Klaster Manual dan Random Clustering

Dalam clustering menggunakan algoritma K-Means, pemilihan pusat klaster (centroid) sangat mempengaruhi hasil pengelompokan data. Penelitian ini menggunakan dua pendekatan untuk menentukan centroid, yaitu secara manual (berdasarkan nilai representatif dari masing-masing klaster) dan secara acak (random initialization). Hasil distribusi klaster dari kedua metode akan dianalisis untuk menilai ketepatan pengelompokan pasien berdasarkan indikator kesehatan dan logika klinis, guna menentukan pendekatan mana yang memberikan hasil lebih representatif.

#### *Distribusi Pasien Berdasarkan Kategori Kesehatan (Manual)*

Pada pendekatan manual, pusat klaster ditentukan berdasarkan nilai yang mewakili kondisi kesehatan pasien dalam setiap kategori, yang menghasilkan distribusi lebih logis dan konsisten dengan indikator kesehatan. Subbab ini menyajikan hasil distribusi pasien berdasarkan kategori tekanan darah, kadar gula, kolesterol, dan asam urat dengan metode centroid manual, yang kemudian dibandingkan untuk melihat kesesuaian klasifikasi pasien dengan kondisi kesehatannya. Hasil distribusi pasien berdasarkan clustering dengan pendekatan centroid manual ditampilkan pada Tabel 3.

**Tabel 3 Distribusi Pasien Berdasarkan Kategori Kesehatan (Manual)**

Kategori Kesehatan	Tekanan Darah Normal	Kadar Gula Normal	Kolesterol dan Asam Urat Normal
(C1) Sehat	19	48	99
(C2) Aman	13	63	33
(C3) Waspada	7	-	39
(C4) Sedang	14	75	-
(C5)	15	1	-

Berat			
<b>Total</b>	<b>68</b>	<b>187</b>	<b>171</b>

#### *Analisis Distribusi Pasien (Manual)*

Pemilihan pusat klaster (centroid) dalam algoritma K-Means sangat mempengaruhi hasil akhir clustering. Perbandingan antara dua pendekatan penentuan centroid, acak (random) dan manual (berdasarkan pengetahuan domain), menunjukkan perbedaan signifikan dalam hasil clustering meskipun menggunakan data dan jumlah klaster yang sama. Tabel 2 dan Tabel 3 mengilustrasikan perbedaan distribusi jumlah pasien berdasarkan tiga komponen utama: tekanan darah, kadar gula, serta kolesterol dan asam urat. Pendekatan manual cenderung menghasilkan distribusi yang lebih konsisten dan proporsional, terutama pada kategori C1 (Sehat) dan C2 (Aman), di mana pasien lebih terklasifikasi sesuai dengan status kesehatannya. Sebaliknya, pendekatan acak menghasilkan distribusi yang kurang konsisten, dengan kategori seperti C5 (Berat) yang tidak menunjukkan kelainan pada beberapa fitur, meskipun secara klinis seharusnya ada anomali pada banyak komponen. Perbedaan ini lebih jelas dapat dilihat dalam Tabel 4.

**Tabel 4 Perbedaan Utama Manual dan Random Clustering**

Aspek	Manual Clustering	Random Clustering
Kesesuaian Klinis	Tinggi, sesuai dengan indikator kesehatan	Rendah, beberapa klaster tidak representatif
Konsistensi Diagnostik	Terlihat dari distribusi merata dan logis	Kurang konsisten, kategori berat tidak muncul penuh
Penempatan Klaster Sehat	Logis dan proporsional	Tidak konsisten, hanya muncul di

		dua fitur
Kategori Aman dan Sedang	Tersebar merata	Hanya fokus di beberapa fitur

Berdasarkan hasil analisis perbandingan, dapat disimpulkan bahwa penentuan pusat kluster secara manual memberikan hasil clustering yang lebih akurat dan representatif secara klinis dibandingkan pendekatan acak. Penentuan centroid berdasarkan pengetahuan domain dan karakteristik data memungkinkan terbentuknya kluster yang lebih stabil, logis, dan relevan untuk diagnosis kesehatan. Beberapa faktor yang memengaruhi perbedaan ini antara lain adalah inisialisasi acak yang tidak terarah, sehingga pusat kluster dapat berada di titik yang tidak mencerminkan pola data secara tepat; kurangnya pertimbangan terhadap batas-batas klinis dalam pendekatan acak; serta ketidakstabilan hasil clustering acak yang bisa berubah setiap kali algoritma dijalankan, berbeda dengan pendekatan manual yang lebih konsisten.

## SIMPULAN

Berdasarkan hasil penelitian dan analisis yang dilakukan, dapat disimpulkan bahwa algoritma K-Means Clustering merupakan metode yang efektif dalam mengelompokkan data pasien ke dalam lima kluster berdasarkan enam indikator kesehatan, yaitu tekanan darah sistolik, diastolik, gula darah puasa, gula darah setelah makan, kolesterol, dan asam urat. Penggunaan pusat kluster yang ditentukan secara manual menghasilkan pengelompokan yang lebih stabil dan relevan secara klinis dibandingkan dengan pendekatan acak, yang cenderung menghasilkan variasi hasil antar eksekusi. Hasil visualisasi dan distribusi pasien pada masing-masing kluster menunjukkan bahwa status kesehatan pasien tidak dapat ditentukan hanya berdasarkan satu atau dua indikator, melainkan dari gabungan

keseluruhan fitur kesehatan. Oleh karena itu, pendekatan ini dapat mendukung proses klasifikasi risiko kesehatan secara lebih komprehensif dan membantu dalam pengambilan keputusan medis berbasis data.

## DAFTAR PUSTAKA

- Agustina, I., & Danar Dana, R. (2024). Implementasi K-Means Clustering untuk Pengelompokan Tingkat Kemiskinan di Pulau Jawa Menggunakan Bahasa Pemrograman Python. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 945–951. <https://doi.org/10.36040/jati.v8i1.8858>
- Bahri, S., Sembiring, F., Aziz, M. A., & Firmansyah, D. (2017). Analisis Pemetaan Tingkat Pengangguran di Pulau Jawa dan Bali dengan Metode K-Means. *Jurnal Rekayasa Teknologi Nusa Putra*, 4(1), 13–18. <https://doi.org/10.52005/rekayasa.v4i1.142>
- Danny, M., & Muhidin, A. (2023). Analisis Prediksi Resiko Diabetes Tahap Awal Menggunakan Algoritma Naive Bayes. *Jurnal Teknologi Informatika Dan Komputer*, 9(2), 1443–1459. <https://doi.org/10.37012/jtik.v9i2.2017>
- Febrianto, A., Achmadi, S., Sasmito, A. P., Industri, F. T., Baik, S., & Mining, D. (2021). PENERAPAN METODE K-MEANS UNTUK CLUSTERING PENGUNJUNG PERPUSTAKAAN ITN MALANG. 5(1), 61–70.
- Hartini, T., Purnamasari, A. I., Bahtiar, A., & Kaslani. (2025). Implementasi Algoritma K-Means Clustering untuk Klasifikasi Diagnosis Penyakit Lambung. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(1), 76–83.
- Iqbal, M., Mayasari, N., & Muslim. (2016). Analisis Performansi Software WEKA dan Rapidminer dalam Pengolahan Data Mining. *Jurnal Teknik Dan Informatika*, 3,

- 35–38.
- Laksono, B., Syahidin, Y., & Yunengsih, Y. (2024). Implementasi Data Mining Klasterisasi Data Pasien Rawat Inap dengan Algoritma K-Means Clustering. *Jurnal Teknologi Sistem Informasi Dan Aplikasi*, 7(2), 621–627.  
<https://doi.org/10.32493/jtsi.v7i2.39354>
- Maesaroh, M., Nur Padilah, T., & Haerul Jaman, J. (2024). Penerapan Algoritma K-Means Clustering pada Pengelompokan Daerah Penyebaran Diare di Provinsi Jawa Barat. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(4), 2783–2787.  
<https://doi.org/10.36040/jati.v7i4.7208>
- Manalu, D. A., & Gunadi, G. (2022). Implementasi Metode Data Mining K-Means Clustering terhadap Data Pembayaran Transaksi Menggunakan Bahasa Pemrograman Python pada CV Digital Dimensi. *Infotech: Journal of Technology Information*, 8(1), 43–54.  
<https://doi.org/10.37365/jti.v8i1.131>
- Marlina, L., Muslim, & Siahaan, A. P. U. (2016). Data Mining Classification Comparison (Naïve Bayes and C4.5 Algorithms). *International Journal of Engineering Trends and Technology*, 38(7), 380–383.
- Ordila, R., Wahyuni, R., Irawan, Y., & Yulia Sari, M. (2020). Penerapan Data Mining untuk Pengelompokan Data Rekam Medis Pasien Berdasarkan Jenis Penyakit dengan Algoritma Clustering (Studi Kasus: Poli Klinik PT. Inecda). *Jurnal Ilmu Komputer*, 9(2), 148–153.  
<https://doi.org/10.33060/JIK/2020/Vo19.Iss2.181>
- Rahmawati, L., Widya Sihwi, S., & Suryani, E. (2016). Analisa Clustering Menggunakan Metode K-Means Dan Hierarchical Clustering (Studi Kasus: Dokumen Skripsi Jurusan Kimia, Fmipa, Universitas Sebelas Maret). *Jurnal Teknologi & Informasi ITSmart*, 3(2), 66.  
<https://doi.org/10.20961/its.v3i2.654>
- Setiawati, N. P., Nugroho, B. A., Setiawan, A. T., & Hartanti, D. (2023). Data Mining Untuk Klasifikasi Diagnosis Tingkat Keparahan Penyakit Diabetes Dengan Algoritma Logistik Regresi. *Prosiding Seminar Nasional Teknologi Informasi Dan Bisnis*.
- Siahaan, A. P. U., & Afandi Syahputra. (2024). Pengenalan Sistem Antrian Berbasis Web di Kantor Dinas Kependudukan dan Pencatatan Sipil Kota Medan. *Jurnal Hasil Pengabdian Masyarakat (JURIBMAS)*, 2(3), 251–263.  
<https://doi.org/10.62712/juribmas.v2i3.154>
- Wala, J., Herman, & Umar, R. (2024). Implementasi K-Means Clustering pada Pengelompokan Pasien Penyakit Jantung. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 9(3), 205–216.
- Yolanda, E., & Suhardi. (2023). Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Data Pasien Rehabilitasi Narkoba. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, 4(1), 182–191.  
<https://doi.org/10.30865/klik.v4i1.1107>