

PENGELOMPOKAN TINGKAT RISIKO KESEHATAN LANSIA BERDASARKAN USIA, TEKANAN DARAH, DAN RIWAYAT PENYAKIT MENGUNAKAN ALGORITMA K-MEANS DI DESA NAGREG KENDAN

Novia Ramadhani¹, Nana Suarna², Agus Bahtiar³, Edi Tohidi⁴.

Program Studi Teknik Informatika¹²
Program Studi Sistem Informasi³
Program Studi Komputerisasi⁴

STMIK IKMI Cirebon
<https://ikmi.ac.id/page/18/?lang=de>
noviaarmd11@gmail.com

(*) Corresponding Author : noviaarmd11@gmail.com
Published : 17 Juli 2026

Abstract—The increasing number of elderly people in Indonesia has led to a rise in non-communicable diseases such as hypertension, diabetes, and cardiovascular disorders. This situation demands the use of information technology to more accurately identify health patterns and support data-driven medical decision-making. This study aims to cluster elderly individuals in Nagreg Kendan Village based on age, systolic–diastolic blood pressure, and disease history using the K-Means clustering algorithm. The data were obtained from direct examinations at the Elderly Posyandu and secondary medical records from the Nagreg Public Health Center. The research stages included data collection, data cleaning, categorical variable transformation, normalization using the Min–Max method, and the clustering process with K-Means++ initialization. To ensure cluster quality, the Davies-Bouldin Index (DBI) was used as the evaluation metric. The results indicate that the lowest DBI value was achieved for a specific number of clusters (based on analysis results), suggesting that the formed clusters have strong internal similarity and good separation between groups. Each cluster reflects different health characteristics among the elderly, ranging from low-risk groups to high-risk groups based on combinations of age, blood pressure, and chronic disease history. These findings provide a clear overview of the distribution of health risks among the elderly in Nagreg Kendan. This study demonstrates that the K-Means method is effective for area-based health segmentation and can serve as a foundation for developing medical decision-support systems at the village level. The clustering results can assist healthcare workers in conducting more targeted monitoring, designing preventive strategies, and determining priority interventions for high-risk elderly groups.

Keywords: K-Means Clustering Elderly, Blood Pressure, Disease History Davies-Bouldin Index Data Mining.

Abstrak—Peningkatan jumlah lansia di Indonesia berdampak pada meningkatnya kasus penyakit tidak menular seperti hipertensi, diabetes, dan gangguan kardiovaskular. Kondisi ini menuntut adanya pemanfaatan teknologi informasi untuk mengidentifikasi pola kesehatan secara lebih akurat dan mendukung pengambilan keputusan medis berbasis data. Penelitian ini bertujuan untuk melakukan pengelompokan lansia di Desa Nagreg Kendan berdasarkan variabel usia, tekanan darah *sistolik–diastolik*, dan riwayat penyakit menggunakan algoritma *K-Means clustering*. Data diperoleh dari hasil pemeriksaan langsung di Posyandu Lansia serta catatan kesehatan sekunder dari Puskesmas Nagreg. Tahapan penelitian meliputi pengumpulan data, pembersihan data, transformasi variabel kategorikal, normalisasi menggunakan metode *Min–Max*, serta proses klusterisasi dengan *inisialisasi K-Means++*. Untuk memastikan kualitas klaster, penelitian ini menggunakan *Davies-Bouldin Index* (DBI) sebagai metrik evaluasi. Hasil pengujian menunjukkan bahwa nilai DBI terendah diperoleh pada jumlah klaster tertentu, (berdasarkan hasil perhitungan pada analisis) yang mengindikasikan bahwa klaster yang terbentuk memiliki tingkat kemiripan internal yang kuat dan jarak antar klaster yang baik. Setiap klaster menggambarkan karakteristik kesehatan lansia yang berbeda, mulai dari kelompok dengan risiko rendah hingga kelompok risiko tinggi berdasarkan kombinasi usia, tekanan darah, serta riwayat penyakit kronis. Temuan ini memberikan gambaran yang jelas mengenai distribusi risiko kesehatan lansia di Nagreg Kendan. Penelitian ini membuktikan bahwa metode *K-Means* efektif digunakan untuk segmentasi kesehatan berbasis wilayah dan dapat menjadi dasar untuk pengembangan sistem pendukung keputusan medis di tingkat desa. Hasil

klasterisasi dapat dimanfaatkan oleh tenaga kesehatan untuk melakukan pemantauan lebih terarah, menyusun strategi pencegahan, serta menentukan prioritas intervensi bagi kelompok lansia dengan risiko lebih tinggi.

Kata Kunci: *K-Means Clustering*, Lansia, Tekanan Darah, Riwayat Penyakit *Davies-Bouldin Index* Data Mining.

INTRODUCTION

Perkembangan teknologi informasi telah memberikan dampak besar terhadap berbagai sektor, termasuk bidang kesehatan. Di negara berkembang seperti Indonesia, tantangan dalam pelayanan kesehatan bagi populasi lanjut usia (lansia) semakin kompleks seiring meningkatnya angka harapan hidup dan prevalensi penyakit tidak menular (PTM). Penyakit seperti hipertensi, diabetes mellitus, dan gangguan kardiovaskular kini menjadi penyebab utama morbiditas dan mortalitas pada kelompok usia lanjut. Kondisi ini diperparah oleh keterbatasan sumber daya kesehatan, rendahnya tingkat deteksi dini, serta pengendalian penyakit kronis yang belum optimal [1], [2]

Multimorbiditas, yaitu keberadaan dua atau lebih penyakit kronis dalam satu individu, semakin meluas pada kelompok lansia di negara berkembang dan memberikan tekanan besar terhadap sistem pelayanan kesehatan ([3], [4]). Dalam konteks tersebut, analisis data kesehatan berbasis teknologi informasi menjadi sangat penting untuk membantu tenaga medis dalam mengenali pola kesehatan populasi lansia serta menentukan intervensi yang tepat sasaran [5]. Pendekatan berbasis data memungkinkan pengambilan keputusan klinis yang lebih cepat, efisien, dan akurat.

Sebagai bagian dari bidang teknik informatika, *data mining* dan *machine learning* berperan penting dalam mendukung analisis kesehatan modern. Salah satu teknik yang banyak digunakan adalah *clustering* atau pengelompokan data, yang dapat membantu dalam mengidentifikasi kelompok pasien dengan karakteristik serupa. Metode *K-Means clustering* menjadi salah satu algoritma populer karena kesederhanaannya, efisiensi komputasi, serta kemampuannya dalam menangani data berskala besar dengan hasil yang mudah diinterpretasikan [6].

Dalam konteks data kesehatan lansia, *K-Means* dapat dimanfaatkan untuk mengelompokkan pasien berdasarkan variabel klinis seperti usia, tekanan darah, kadar gula, dan riwayat penyakit. Hasil pengelompokan tersebut dapat membantu tenaga medis dan

peneliti dalam mengidentifikasi kelompok risiko tinggi dan merancang strategi pencegahan serta perawatan yang lebih terarah [7]. Pendekatan ini juga berpotensi mendukung *clinical decision support systems (CDSS)* yang berbasis kecerdasan buatan untuk meningkatkan efisiensi pelayanan kesehatan.

Lebih lanjut, metode *K-Means* dapat diintegrasikan dengan teknik reduksi dimensi seperti *Principal Component Analysis (PCA)* atau *t-distributed Stochastic Neighbor Embedding (t-SNE)* untuk mempermudah visualisasi pola data kesehatan yang kompleks [8]. Namun, metode ini juga memiliki kelemahan seperti sensitivitas terhadap jumlah kluster (*K*) dan inisialisasi awal, sehingga validasi menggunakan metrik seperti *Silhouette Score* atau *Davies-Bouldin Index (DBI)* sangat diperlukan untuk memastikan kualitas hasil pengelompokan [6].

Oleh karena itu, penelitian dalam bidang teknik informatika yang berfokus pada penerapan algoritma *K-Means* untuk analisis data kesehatan lansia memiliki nilai strategis. Penelitian semacam ini tidak hanya memberikan kontribusi terhadap pengembangan ilmu pengetahuan di bidang *machine learning* dan *data analytics*, tetapi juga mendukung upaya peningkatan kualitas pelayanan kesehatan berbasis data di Indonesia. Dengan memanfaatkan pendekatan komputasional yang efektif, diharapkan sistem kesehatan dapat beradaptasi lebih baik terhadap kebutuhan populasi lansia yang terus bertambah dan semakin kompleks.

MATERIALS AND METHODS

Penelitian ini termasuk dalam kategori penelitian kuantitatif deskriptif dengan pendekatan data mining, yang bertujuan untuk mengidentifikasi pola dan kelompok risiko pada data kesehatan lansia menggunakan metode *unsupervised learning*. Pendekatan ini digunakan karena data yang dianalisis tidak memiliki label atau kategori awal, sehingga diperlukan metode analisis berbasis pola untuk menemukan struktur tersembunyi dalam data.

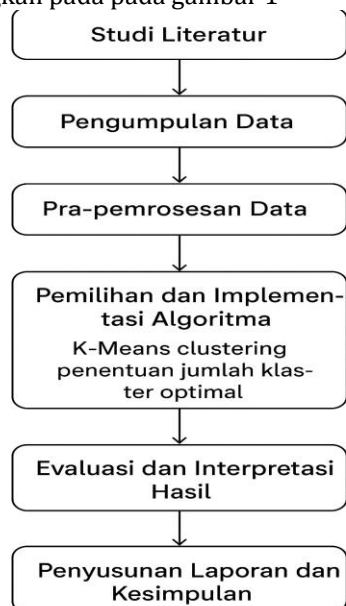
Objek dalam penelitian ini berupa data kesehatan lansia yang berisi sejumlah variabel klinis, seperti usia, tekanan darah, kadar gula darah, indeks massa tubuh (IMT), serta riwayat penyakit

tidak menular seperti hipertensi dan diabetes mellitus. Data tersebut digunakan untuk membangun model pengelompokan yang mampu mengidentifikasi kelompok lansia berdasarkan kesamaan karakteristik klinis.

Pemilihan metode *unsupervised learning* didasarkan pada kebutuhan untuk melakukan eksplorasi pola tanpa adanya variabel target. Algoritma *K-Means clustering* dipilih karena memiliki keunggulan dalam hal efisiensi komputasi, kemudahan implementasi, dan interpretabilitas hasil. Metode ini memungkinkan analisis struktur data yang kompleks dengan membagi populasi lansia ke dalam beberapa kelompok homogen yang dapat digunakan untuk mendukung perencanaan intervensi kesehatan secara terarah. Selain itu, hasil klasterisasi dapat menjadi dasar untuk penelitian lanjutan menggunakan pendekatan *supervised learning* guna membangun model prediksi risiko kesehatan.

Desain penelitian ini menggunakan eksperimen komputasional yang dilakukan melalui beberapa tahap sistematis, dimulai dari pengumpulan dan pra-pemrosesan data hingga analisis hasil pengelompokan. Eksperimen ini dilaksanakan menggunakan perangkat lunak pengolahan data seperti Python dengan pustaka pendukung *machine learning* (misalnya *scikit-learn*, *pandas*, dan *matplotlib*) untuk memastikan reproduktibilitas dan transparansi hasil penelitian.

Tahapan utama penelitian ini meliputi langkah-langkah pada gambar 1



Gambar 1. Alur Penelitian

Alur penelitian pada diagram tersebut dimulai dari tahap studi literatur, yaitu peneliti

mengkaji berbagai sumber ilmiah seperti jurnal, buku, dan penelitian terdahulu untuk memahami konsep serta metode yang relevan. Selanjutnya, dilakukan pengumpulan data yang bertujuan memperoleh dataset yang akan digunakan dalam penelitian. Setelah data terkumpul, tahap berikutnya adalah pra-pemrosesan data, yang meliputi pembersihan data (cleaning), transformasi, dan normalisasi agar data siap diolah dan menghasilkan analisis yang lebih akurat.

Tahap inti penelitian terdapat pada pemilihan dan implementasi algoritma, di mana metode *K-Means clustering* digunakan untuk mengelompokkan data. Pada tahap ini juga dilakukan penentuan jumlah kluster optimal agar hasil pengelompokan lebih representatif dan bermakna. Setelah model diterapkan, dilakukan evaluasi dan interpretasi hasil untuk menilai kualitas kluster yang terbentuk serta memahami pola yang dihasilkan dari data. Terakhir, seluruh proses dan temuan penelitian dirangkum dalam tahap penyusunan laporan dan kesimpulan, yang berisi hasil analisis, pembahasan, serta rekomendasi berdasarkan penelitian yang telah dilakukan.

RESULTS AND DISCUSSION

Hasil akhir proses *K-Means* diekspor ke file DATA CLUSTER.xlsx (sheet "Cluster Model"), yang berisi daftar anggota tiap kluster beserta centroid-nya.

Setiap baris mewakili satu responden dengan atribut numerik (usia, tekanan darah sistolik, tekanan darah diastolik, dan variabel hasil *one-hot encoding* penyakit).

Tabel 1 memperlihatkan tampilan tab *Cluster Model* di RapidMiner yang menunjukkan hasil klasterisasi dengan dua kelompok berbeda.

Tabel 1 Hasil Klasterisasi

Klaster	Item	Usia Rata-rata	TD Sistolik	TD Diastolik	Ciri Dominan
1	249				
2	1				

Data ini menunjukkan bahwa terdapat dua klaster dengan karakteristik risiko tinggi (1) klaster risiko rendah (2).

Proses pengelompokan data kesehatan lansia di Desa Nagreg Kendan menggunakan algoritma *K-Means* menghasilkan dua cluster dengan karakteristik yang sangat berbeda. Pembahasan ini menguraikan profil masing-masing cluster beserta jumlah anggotanya dan implikasi kesehatan berdasarkan variabel usia, tekanan darah, dan riwayat penyakit.

Profil Cluster 0 (249 item) Kelompok Lansia Risiko Sedang-Tinggi Cluster 0 merupakan cluster dengan jumlah anggota terbesar, yaitu 249 lansia atau sekitar 99,6% dari total data. Profil ini menunjukkan bahwa sebagian besar lansia di desa tersebut memiliki karakteristik yang relatif serupa. Cluster ini memiliki rata-rata usia 73,50 tahun, yang menempatkan kelompok ini pada kategori lansia lanjut (*elderly-old group*). Usia lanjut merupakan faktor penentu penting terhadap meningkatnya risiko penyakit degeneratif dan penurunan fungsi fisiologis.

Nilai riwayat penyakit dalam cluster ini menunjukkan adanya beberapa kondisi seperti asam urat, diabetes, dan hipertensi, meskipun tidak dimiliki oleh semua anggota. Temuan ini menunjukkan bahwa sebagian anggota cluster pernah atau sedang mengalami gangguan metabolik dan kardiovaskular. Selain itu, nilai tekanan darah rata-rata 130,21 mengindikasikan kondisi *prehypertension* menuju hipertensi tahap awal, yang dapat memperparah risiko komplikasi kesehatan pada lansia.

Dengan kombinasi usia lanjut, tekanan darah tinggi, dan sejumlah riwayat penyakit, Cluster 0 dapat diidentifikasi sebagai kelompok lansia berisiko sedang hingga tinggi. Kelompok ini menjadi prioritas utama untuk pemantauan kesehatan berkala, intervensi dini, serta edukasi mengenai pola hidup sehat dan manajemen penyakit kronis.

Profil Cluster 1 (1 item) Kelompok Lansia Risiko Rendah Cluster 1 hanya beranggotakan 1 lansia dari total 250 data, atau sekitar 0,4%. Meskipun jumlahnya sangat kecil, cluster ini tetap memiliki karakteristik berbeda yang memisahkannya dari cluster utama. Lansia dalam cluster ini memiliki usia rata-rata 68 tahun, yang lebih muda dibandingkan dengan cluster 0. Kondisi usia yang lebih rendah umumnya diikuti oleh fungsi fisik dan metabolisme tubuh yang masih relatif stabil.

Riwayat penyakit pada cluster ini tercatat nihil, kecuali satu indikator hipertensi. Tidak adanya riwayat penyakit lain seperti jantung, diabetes, atau asam urat menunjukkan bahwa lansia dalam cluster ini berada dalam kondisi kesehatan yang lebih baik. Meskipun demikian, nilai tekanan darah sebesar 110 menunjukkan potensi risiko awal yang tetap perlu dipantau, karena tekanan darah merupakan salah satu indikator kesehatan penting pada lansia.

Dengan demikian, Cluster 1 dapat dikategorikan sebagai lansia berisiko rendah hingga sedang, terutama karena minimnya

riwayat penyakit dan usia yang lebih muda dibanding mayoritas populasi.

Perbandingan Antarcluster dan Implikasi Kesehatan

Analisis perbandingan kedua cluster menunjukkan perbedaan profil kesehatan yang sangat jelas. Cluster 0, yang beranggotakan hampir seluruh lansia dalam dataset, memperlihatkan bahwa sebagian besar lansia memiliki kondisi kesehatan yang memerlukan perhatian lebih, terutama dari sisi tekanan darah dan usia lanjut. Di sisi lain, Cluster 1, meskipun jumlah anggotanya hanya satu, menunjukkan kondisi kesehatan yang relatif lebih baik dan lebih stabil.

Distribusi yang sangat tidak seimbang ini menegaskan bahwa populasi lansia di Desa Nagreg Kendan secara dominan berada pada kelompok risiko kesehatan yang lebih tinggi. Hal ini menjadi temuan penting bagi pemerintah desa dan tenaga kesehatan, karena menunjukkan perlunya program intervensi kesehatan yang terfokus pada pencegahan penyakit degeneratif, pemeriksaan tekanan darah rutin, serta peningkatan literasi kesehatan pada lansia.

Selain itu, hasil clustering ini membuktikan bahwa metode K-Means efektif dalam mengidentifikasi kelompok risiko berdasarkan variabel numerik dan kategorikal sederhana. Temuan ini dapat menjadi dasar bagi penelitian lanjutan untuk mengembangkan sistem pendukung keputusan (Decision Support System) atau dashboard pemantauan kesehatan lansia secara real-time di tingkat desa.

CONCLUSION

Penelitian ini bertujuan untuk mengelompokkan tingkat risiko kesehatan lansia di Desa Nagreg Kendan berdasarkan usia, tekanan darah, dan riwayat penyakit menggunakan algoritma K-Means. Berdasarkan hasil pengolahan data dan analisis pada bab sebelumnya, diperoleh kesimpulan sebagai berikut:

1. Algoritma K-Means berhasil membentuk dua cluster yang menunjukkan perbedaan karakteristik kesehatan lansia. Cluster 0 beranggotakan 249 lansia (99,6%) sedangkan Cluster 1 berisi 1 lansia (0,4%). Ketimpangan jumlah anggota cluster menunjukkan bahwa sebagian besar lansia memiliki kondisi kesehatan yang hampir serupa.
2. Cluster 0 merupakan kelompok dengan tingkat risiko kesehatan sedang hingga tinggi. Cluster ini didominasi oleh lansia berusia rata-rata 73,50 tahun dengan tekanan darah 130,21 serta memiliki riwayat beberapa penyakit seperti asam

- urat, diabetes, dan hipertensi. Kondisi ini menunjukkan bahwa kelompok ini membutuhkan perhatian khusus dan pemantauan kesehatan yang intensif.
3. Cluster 1 merupakan kelompok lansia dengan tingkat risiko rendah. Kelompok ini hanya berisi 1 lansia berusia 68 tahun, tanpa riwayat penyakit signifikan, dan memiliki tekanan darah relatif lebih rendah. Meskipun demikian, keberadaan hanya satu anggota dalam cluster ini menunjukkan bahwa kelompok lansia dengan profil kesehatan baik sangat sedikit.
 4. Variabel usia, tekanan darah, dan riwayat penyakit terbukti berpengaruh signifikan terhadap pembentukan cluster. Ketiga variabel tersebut mampu membedakan tingkat risiko kesehatan lansia melalui pola kemiripan data yang diidentifikasi oleh K-Means.
 5. Hasil penelitian ini dapat menjadi dasar pengambilan keputusan dalam program kesehatan lansia di desa, terutama dalam menentukan prioritas layanan kesehatan bagi lansia berisiko tinggi.

REFERENCE

- [1] A. E. Schutte, "Global burden of hypertension in low- and middle-income countries," 2021.
- [2] O. A. Uthman, "Epidemiology of non-communicable diseases in aging populations," 2022.
- [3] M. Khan, "Multimorbidity and health outcomes among older adults in developing countries," 2022.
- [4] C. Hategeka, "Chronic disease management challenges in low-resource settings," 2022.
- [5] B. Zhou, "Hypertension and diabetes prevalence in aging populations of developing countries," 2021.
- [6] T. J. Loftus, "K-Means clustering in clinical decision support systems," 2022.
- [7] A. Habebh and S. Gohel, "Comparative study of clustering algorithms in healthcare data," 2021.
- [8] Y. Yang, "Visualization-enhanced clustering for health data analytics," 2024.