Jurnal Kolaborasi
Sains dan Ilmu Terapan
Volume 4 Nomor 2 Juni 2026 | Page -87-95
ISSN 2964- 5581 (media online)
https://utilityprojectsolution.org/ejournal/index.php/JuKSIT
Penerbit : Cv. Utility Project Solution

Implementasi *K-Means Clustering* Untuk Segmentasi Pasien Berdasarkan Riwayat Kunjungan Dan Diagnosis

Natanael Maruli Tua Limbong[1], N P Dharshinni [2*]

[1-2] Fakultas Sains Dan Teknologi, Universitas Prima Indonesia, Medan, Indonesia
Email: [1]natanaellimbong870@gmail.com, [2,*]priyadharshinninaidu@gmail.com*
Email Penulis Korespondensi: [2]priyadharshinninaidu@gmail.com*

Info Artikel

Kata Kunci:
Segmentasi Pasien
Riwayat Kunjungan Pasien
Diagnosis Medis
Pengelompokan K-Means
KNIME

Abstrak

Penelitian ini bertujuan untuk mengimplementasikan algoritma K-Means Clustering untuk mengelompokkan pasien berdasarkan riwayat kunjungan dan diagnosis medis. Dataset yang digunakan diperoleh dari Kaggle, berisi data pasien seperti usia, kondisi medis, tipe admisi, jumlah tagihan, dan lama rawat inap. Proses penelitian meliputi pra-pemrosesan data, transformasi, normalisasi, dan implementasi algoritma KMeans menggunakan aplikasi KNIME. Jumlah klaster yang digunakan dalam penelitian ini adalah tiga. Hasil menunjukkan bahwa data pasien dapat dikelompokkan menjadi tiga kategori: tingkat kondisi rendah, sedang, dan tinggi. Klaster dengan tingkat kondisi tinggi memiliki jumlah anggota terbanyak, mengindikasikan pasien yang membutuhkan perawatan lebih intensif. Hasil pengelompokan ini memberikan wawasan tentang pola segmentasi pasien dan dapat digunakan untuk mendukung pengambilan keputusan serta meningkatkan kualitas layanan kesehatan.

Keywords:
Patient Segmentation
Patient Visit History
Medical diagnosis
K-Means Clustering
KNIME

Abstract

This study aims to implement the K-Means Clustering algorithm to segment patients based on visit history and medical diagnosis. The dataset used was obtained from Kaggle, containing patient data such as age, medical condition, admission type, billing amount, and length of stay. The research process includes data preprocessing, transformation, normalization, and the implementation of the KMeans algorithm using the KNIME application. The number of clusters used in this study is three. The results show that patient data can be grouped into three categories: low, medium, and high condition levels. The cluster with high condition levels has the largest number of members, indicating patients who require more intensive care. These clustering results provide insights into patient segmentation patterns and can be used to support decision-making and improve healthcare service quality.

1. PENDAHULUAN

Perkembangan teknologi informasi di era digital telah membawa perubahan signifikan dalam berbagai bidang, termasuk sektor kesehatan. Sistem informasi kesehatan saat ini mampu menyimpan data dalam jumlah besar seperti data rekam medis pasien, riwayat kunjungan, hasil pemeriksaan, serta diagnosis penyakit secara terstruktur dan terintegrasi dalam database rumah sakit [1]. Keberadaan data dalam jumlah besar tersebut menjadi aset yang sangat berharga apabila dapat diolah dan dimanfaatkan dengan baik untuk mendukung pengambilan keputusan.

Namun, pada kenyataannya banyak institusi kesehatan yang belum mampu memanfaatkan data tersebut secara optimal. Data yang tersimpan sering sekali hanya digunakan sebagai arsip administratif tanpa dilakukan analisis lebih lanjut untuk menggali informasi penting yang terkandung di dalamnya [2]. Hal ini menyebabkan potensi besar dari data

Submitted: **29/05/2026**; Accepted: **12/06/2026**; Published:**22/06/2026**

Jurnal Kolaborasi
Sains dan Ilmu Terapan
Volume 4 Nomor 2 Juni 2026 | Page -87-95
ISSN 2964- 5581 (media online)
https://utilityprojectsolution.org/ejournal/index.php/JuKSIT Penerbit : Cv. Utility Project Solution

medis tidak dimanfaatkan secara maksimal, terutama dalam memahami pola kunjungan pasien dan karakteristik penyakit.

Seiring dengan meningkatnya jumlah pasien setiap tahun, volume data yang dihasilkan juga semakin besar dan kompleks. Kondisi ini memerlukan pendekatan khusus dalam pengolahan data agar dapat menghasilkan informasi yang akurat dan relevan. Salah satu pendekatan yang dapat digunakan adalah data mining, yaitu proses pengolahan data dalam jumlah besar untuk menemukan pola atau informasi tersembunyi yang tidak dapat diperoleh secara manual [3]. Data mining memungkinkan pengolahan data kesehatan menjadi informasi yang lebih bermakna untuk mendukung pengambilan keputusan berbasis data.

Dalam bidang kesehatan, salah satu pemanfaatan data mining adalah untuk melakukan segmentasi pasien. Segmentasi pasien merupakan proses pengelompokan pasien berdasarkan karakteristik tertentu seperti frekuensi kunjungan, jenis penyakit, usia, dan faktor lainnya. Segmentasi ini sangat penting karena dapat membantu pihak rumah sakit dalam memahami perilaku pasien serta menentukan strategi pelayanan yang lebih efektif [4].

Teknik yang umum digunakan dalam segmentasi adalah *clustering*. *Clustering* merupakan metode pengelompokan data berdasarkan tingkat kemiripan antar objek sehingga objek dalam satu kelompok memiliki karakteristik yang serupa dan berbeda dengan kelompok lainnya [5]. Metode ini sangat cocok digunakan dalam analisis data kesehatan karena mampu mengelompokkan pasien tanpa memerlukan label atau kategori sebelumnya.

Salah satu algoritma *clustering* yang paling banyak digunakan adalah *K-Means*. Algoritma ini bekerja dengan membagi data ke dalam sejumlah *cluster* berdasarkan jarak terdekat terhadap titik pusat (*centroid*) yang telah ditentukan. *K-Means* memiliki keunggulan dalam hal kemudahan implementasi, kecepatan proses, serta kemampuan dalam mengolah data dalam jumlah besar secara efisien. Beberapa penelitian sebelumnya menunjukkan bahwa algoritma *K-Means* dapat diterapkan secara efektif dalam bidang kesehatan. Penelitian yang dilakukan oleh peneliti sebelumnya menunjukkan bahwa *K-Means* mampu mengelompokkan data rekam medis pasien berdasarkan pola penyakit sehingga dapat membantu dalam analisis distribusi penyakit. Selain itu, penelitian lain juga menunjukkan bahwa metode ini dapat digunakan untuk mengidentifikasi kelompok pasien dengan karakteristik tertentu yang dapat digunakan sebagai dasar dalam perencanaan layanan kesehatan [6].

Selain itu, penerapan *K-Means* dalam analisis data kesehatan juga dapat membantu dalam meningkatkan efisiensi pelayanan rumah sakit. Dengan mengetahui kelompok pasien berdasarkan pola kunjungan dan diagnosis, pihak rumah sakit dapat mengalokasikan sumber daya secara lebih tepat, seperti tenaga medis, fasilitas, dan jadwal pelayanan [7]. Hal ini tentunya akan berdampak pada peningkatan kualitas pelayanan dan kepuasan pasien.

Dalam penelitian ini, data yang digunakan tidak diambil langsung dari rumah sakit, melainkan menggunakan dataset publik dari *Kaggle* sebagai simulasi data pasien. Penggunaan dataset *Kaggle* dalam penelitian data mining telah banyak dilakukan karena menyediakan data yang beragam, terstruktur, dan dapat digunakan untuk eksperimen analisis data. Dataset tersebut biasanya mencakup informasi seperti identitas pasien, riwayat kunjungan, serta diagnosis penyakit yang dapat digunakan untuk proses *clustering*.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk mengimplementasikan algoritma *K-Means Clustering* dalam melakukan segmentasi pasien berdasarkan riwayat kunjungan dan diagnosis penyakit. Dengan adanya segmentasi ini, diharapkan dapat memberikan gambaran mengenai pola pasien yang dapat digunakan sebagai dasar dalam pengambilan keputusan di bidang kesehatan.

2. METODOLOGI PENELITIAN

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan data mining. Penelitian kuantitatif digunakan karena data yang diolah berupa data numerik yang dapat dianalisis menggunakan metode statistik dan algoritma tertentu.

Pendekatan data mining digunakan untuk menggali informasi dan menemukan pola tersembunyi dari data pasien, khususnya dalam melakukan segmentasi berdasarkan riwayat kunjungan dan diagnosis penyakit. Metode yang digunakan dalam penelitian ini adalah algoritma *K-Means Clustering*, yaitu metode pengelompokan data yang membagi data ke dalam beberapa *cluster* berdasarkan tingkat kemiripan antar data.

Penelitian ini dilaksanakan dalam rentang waktu kurang lebih 6 bulan, yang dimulai dari tahap pengumpulan data hingga penyusunan laporan penelitian.

Submitted: **29/05/2026**; Accepted: **12/06/2026**; Published:**22/06/2026**

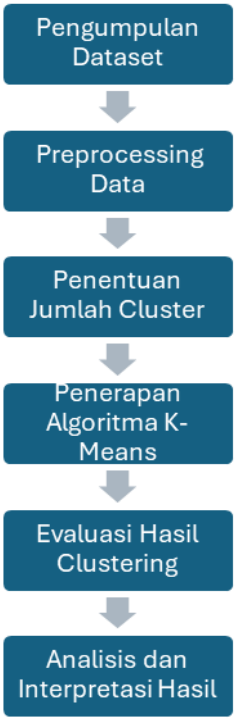
Jurnal Kolaborasi
Sains dan Ilmu Terapan
Volume 4 Nomor 2 Juni 2026 | Page -87-95
ISSN 2964- 5581 (media online)
https://utilityprojectsolution.org/ejournal/index.php/JuKSIT Penerbit : Cv. Utility Project Solution

Adapun tahapan waktu penelitian meliputi:
1. Pengumpulan *dataset* dari *Kaggle*
2. *Preprocessing* data
3. Implementasi algoritma *K-Means*
4. Analisis hasil *clustering*
5. Penyusunan laporan skripsi

Tempat pelaksanaan penelitian dilakukan secara mandiri menggunakan perangkat komputer/laptop pribadi dengan bantuan software pengolahan data.

1.1 Prosedur Kerja

Langkah-langkah prosedur penelitian ditunjukkan melalui alur dibawah ini :



Gambar 1. Prosedur Kerja

Prosedur kerja dalam penelitian ini diawali dengan tahap pengumpulan dataset yang diperoleh dari platform *Kaggle*. Dataset yang digunakan merupakan data pasien yang mencakup beberapa atribut penting seperti riwayat kunjungan, diagnosis penyakit, usia, dan jenis kelamin. Data tersebut digunakan sebagai bahan utama dalam proses analisis untuk melakukan segmentasi pasien.

Selanjutnya dilakukan tahap *preprocessing* data yang bertujuan untuk mempersiapkan data agar siap diolah. Pada tahap ini dilakukan proses *cleaning* data untuk menghapus data yang tidak lengkap, duplikat, atau tidak relevan. Kemudian dilakukan transformasi data dengan mengubah data kategorikal menjadi data numerik agar dapat diproses oleh algoritma *K-Means*. Selain itu, dilakukan juga proses normalisasi data untuk menyamakan skala nilai antar atribut sehingga tidak terjadi bias dalam perhitungan jarak antar data.

Setelah data siap digunakan, langkah berikutnya adalah menentukan jumlah *cluster* yang optimal. Penentuan jumlah *cluster* dilakukan menggunakan metode tertentu seperti *Elbow Method*, yang bertujuan untuk menemukan jumlah *cluster* terbaik berdasarkan nilai *error* terkecil.

Tahap selanjutnya adalah penerapan algoritma *K-Means Clustering*. Proses ini dimulai dengan menentukan jumlah *cluster* yang telah diperoleh sebelumnya, kemudian menetapkan *centroid* awal secara acak. Selanjutnya dilakukan perhitungan jarak antara setiap data dengan *centroid* menggunakan metode perhitungan jarak tertentu. Data

Submitted: **29/05/2026**; Accepted: **12/06/2026**; Published:**22/06/2026**

Jurnal Kolaborasi
Sains dan Ilmu Terapan
Volume 4 Nomor 2 Juni 2026 | Page -87-95
ISSN 2964- 5581 (media online)
https://utilityprojectsolution.org/ejournal/index.php/JuKSIT Penerbit : Cv. Utility Project Solution

kemudian dikelompokkan ke dalam *cluster* yang memiliki jarak terdekat dengan *centroid*. Proses ini dilakukan secara berulang dengan memperbarui posisi *centroid* hingga tidak terjadi perubahan signifikan atau mencapai kondisi konvergen.

Setelah proses *clustering* selesai, dilakukan tahap evaluasi hasil untuk mengetahui kualitas *cluster* yang terbentuk. Evaluasi dilakukan menggunakan metode seperti *Silhouette Score* dan *Inertia* untuk mengukur tingkat kesesuaian dan pemisahan antar *cluster*.

Tahap terakhir adalah analisis dan interpretasi hasil *clustering*. Pada tahap ini dilakukan analisis terhadap hasil pengelompokan pasien untuk mengetahui pola segmentasi berdasarkan riwayat kunjungan dan diagnosis penyakit. Hasil analisis tersebut kemudian digunakan sebagai dasar dalam penarikan kesimpulan serta memberikan rekomendasi yang dapat mendukung pengambilan keputusan di bidang pelayanan kesehatan.

3. HASIL DAN PEMBAHASAN

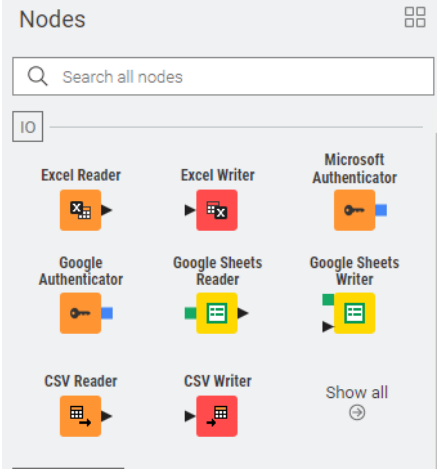
Penelitian ini menggunakan *dataset* yang diperoleh dari platform *Kaggle* berupa data pasien pada layanan kesehatan. *Dataset* terdiri dari beberapa atribut yaitu *Name, Age, Gender, Blood Type, Medical Condition, Date of Admission, Doctor, Hospital, Insurance Provider, Billing amount, Room Number, Admission type, Discharge Date, Medication,* dan *Test Results*.

Berdasarkan hasil analisis awal, jumlah data yang digunakan sebanyak 40 data pasien (menyesuaikan dataset yang digunakan). Dalam proses pengolahan data, tidak semua atribut digunakan secara langsung dalam bentuk asli, melainkan dilakukan transformasi agar sesuai dengan kebutuhan algoritma *cluster*ing. Tahap normalisasi data dilakukan menggunakan Microsoft Excel, sedangkan proses *cluster*ing dilakukan menggunakan algoritma *K-Means* yang diimplementasikan melalui *KNIME Analytics Platform*.

Dalam penelitian ini, proses analisis data pada *KNIME* dilakukan melalui beberapa tahapan dengan menggunakan operator (*node*) sebagai berikut:

1. Excel Reader

   Node ini digunakan untuk mengimpor *dataset* pasien ke dalam *KNIME*. *Dataset* yang telah diunduh dari *Kaggle* dimasukkan dalam format *.xlsx*, kemudian dilakukan pengecekan struktur data seperti jumlah atribut dan tipe data.



Gambar 2. *Node Excel Reader (I/O)*

2. Column Filter

   Node ini digunakan untuk memilih atribut yang relevan dengan penelitian, yaitu *Age, Medical Condition, Admission type, Billing amount,* dan *Length of stay*. Atribut yang tidak relevan seperti *Name, Doctor*, dan *Room Number* dihapus agar tidak mempengaruhi proses *clustering*.

Submitted: **29/05/2026**; Accepted: **12/06/2026**; Published:**22/06/2026**

Jurnal Kolaborasi
Sains dan Ilmu Terapan
Volume 4 Nomor 2 Juni 2026 | Page -87-95
ISSN 2964- 5581 (media online)
https://utilityprojectsolution.org/ejournal/index.php/JuKSIT Penerbit : Cv. Utility Project Solution

Gambar 3. *Column Filter*

3. *K-Means Cluster*ing

   Node ini merupakan inti dari proses *clustering*, dimana data pasien dikelompokkan ke dalam 4 *cluster* berdasarkan tingkat kemiripan karakteristik data. Parameter jumlah *cluster* (K) diatur sebanyak 3 sesuai dengan kebutuhan penelitian.

Gambar 4 *K-Means Properties*

4. Scatter Plot / *Cluster* View

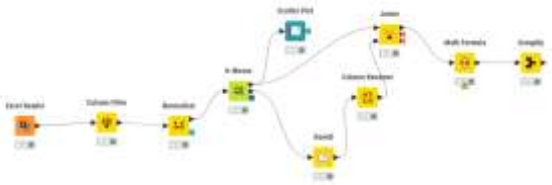
   Node ini digunakan untuk menampilkan visualisasi hasil *clustering* dalam bentuk grafik, sehingga dapat dilihat persebaran data dan perbedaan antar *cluster*. Berikut ini adalah hasil dari analisis dengan *view scatter plot*

Gambar 5. *Scatter Plots*

Setelah seluruh proses dijalankan, diperoleh hasil berupa pengelompokan data pasien ke dalam beberapa *cluster*. Setiap data pasien akan memiliki label *cluster* yang menunjukkan kelompoknya masing-masing. Hasil ini kemudian digunakan untuk menganalisis karakteristik masing-masing *cluster* berdasarkan riwayat kunjungan dan diagnosis penyakit. Berikut ini adalah ilustrasi node yang digunakan

Submitted: **29/05/2026**; Accepted: **12/06/2026**; Published:**22/06/2026**

Jurnal Kolaborasi
Sains dan Ilmu Terapan
Volume 4 Nomor 2 Juni 2026 | Page -87-95
ISSN 2964- 5581 (media online)
https://utilityprojectsolution.org/ejournal/index.php/JuKSIT Penerbit : Cv. Utility Project Solution



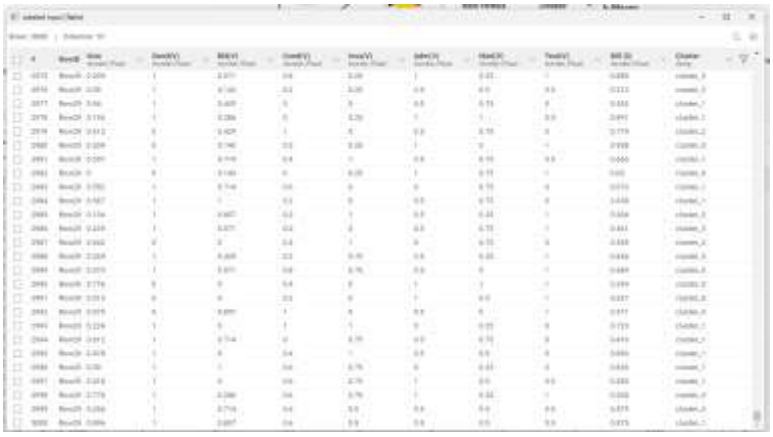
Gambar 6 Node *KNIME*

Proses pengolahan data dalam penelitian ini dilakukan menggunakan perangkat lunak *KNIME* dengan pendekatan *clustering* menggunakan algoritma *K-Means*. Tahapan dimulai dari node Excel Reader yang berfungsi untuk mengimpor data dari file Excel ke dalam lingkungan kerja *KNIME*. Selanjutnya, node Column Filter digunakan untuk memilih atribut-atribut yang relevan dan bertipe numerik sebagai variabel input dalam proses *cluster*ing. Data yang telah diseleksi kemudian diproses menggunakan node Normalizer untuk menyamakan skala antar variabel, sehingga setiap atribut memiliki kontribusi yang seimbang dalam perhitungan jarak pada algoritma *K-Means*.

3.1 Evaluasi Hasil *Clustering*

Evaluasi hasil *clustering* dilakukan untuk mengetahui kualitas pengelompokan data pasien yang dihasilkan menggunakan algoritma *K-Means*. Evaluasi ini bertujuan untuk memastikan bahwa data dalam satu *cluster* memiliki tingkat kemiripan yang tinggi (homogen) serta memiliki perbedaan yang jelas dengan *cluster* lainnya (heterogen).

Gambar 7 Evaluasi Statistik Hasil *KNIME*



Gambar 8 Evaluasi Hasil *KNIME*

3.2 Pembahasan

Hasil pengolahan data yang telah dilakukan, dihasilkan **4 *cluster***, yaitu 841pasien (*cluster*_0), 725 pasien (*cluster*_1), 661pasien (*cluster*_2), 773 pasien (*cluster*_3).

(*cluster*_0) menunjukkan pasien kondisi kritis dan kebutuhan penanganan tinggi, kelompok ini terdiri dari pasien dengan tingkat keparahan penyakit tinggi, sering masuk melalui jalur emergency, memiliki biaya tagihan besar, serta lama rawat inap yang cenderung lebih panjang.

Submitted: **29/05/2026**; Accepted: **12/06/2026**; Published:**22/06/2026**

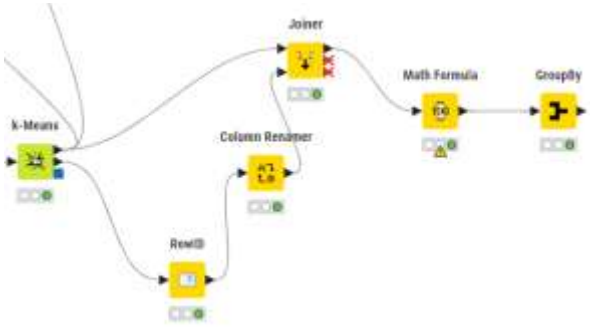
Jurnal Kolaborasi
Sains dan Ilmu Terapan
Volume 4 Nomor 2 Juni 2026 | Page -87-95
ISSN 2964- 5581 (media online)
https://utilityprojectsolution.org/ejournal/index.php/JuKSIT Penerbit : Cv. Utility Project Solution

(*cluster*_1) menunjukkan pasien dengan kondisi sedang dan perawatan intensif menengah, Pasien dalam *cluster* ini memiliki kondisi yang tidak terlalu kritis namun tetap membutuhkan perhatian medis yang cukup, dengan biaya dan lama rawat inap pada tingkat menengah.

(*cluster*_2) menunjukkan pasien dengan kondisi ringan dan kunjungan terjadwal, Kelompok ini umumnya terdiri dari pasien dengan kondisi yang relatif ringan, masuk melalui jalur *elective*, biaya lebih rendah, dan lama rawat inap yang singkat.

Pasien dengan kondisi khusus atau variatif (*cluster*_3) menunjukkan pasien dengan kondisi khusus, *cluster* ini mencakup pasien dengan karakteristik yang beragam *(outlier)*, seperti kombinasi kondisi medis tertentu, hasil tes tidak pasti *(inconclusive)*, atau pola biaya dan lama rawat inap yang tidak konsisten dengan *cluster* lainnya.

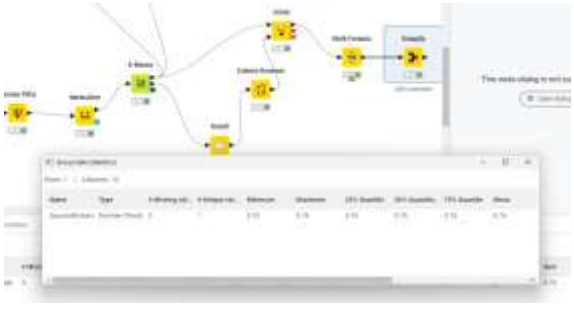
Hasil perhitungan *Mean Squared Error* (*MSE*) yang diperoleh melalui rangkaian proses pada perangkat lunak *KNIME* dengan memanfaatkan node *Joiner*, *Math Formula*, dan *GroupBy*. Proses diawali dengan penggabungan data hasil *cluster*ing dengan *centroid* menggunakan node Joiner berdasarkan kesesuaian label *cluster*. Hasil penggabungan ini menghasilkan dataset yang memuat nilai atribut asli beserta nilai *centroid* dari masing-masing *cluster*.



Gambar 9. *Node* Untuk Menghitung *MSE*

Selanjutnya, node *Math Formula* digunakan untuk menghitung jarak kuadrat antara setiap data dengan *centroid cluster*-nya menggunakan pendekatan jarak Euclidean. Perhitungan ini menghasilkan nilai *squared distance* untuk setiap data. Tahap akhir dilakukan menggunakan node *GroupBy* dengan metode agregasi mean untuk memperoleh nilai rata-rata dari seluruh squared distance. Nilai rata-rata tersebut merupakan nilai *Mean Squared Error* (*MSE*) yang digunakan sebagai indikator kualitas hasil *cluster*ing.

Berdasarkan hasil perhitungan, diperoleh nilai *MSE* sebesar **0,76**. Nilai ini menunjukkan tingkat kedekatan data terhadap *centroid cluster*, di mana semakin kecil nilai *MSE* yang dihasilkan, maka semakin baik kualitas pengelompokan data yang diperoleh dari algoritma *K-Means*.



Gambar 10 Hasil *MSE*

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan mengenai penerapan algoritma *K-Means Clustering* untuk segmentasi pasien berdasarkan riwayat kunjungan dan diagnosis, maka dapat diambil beberapa kesimpulan sebagai berikut Algoritma *K-Means* berhasil diimplementasikan untuk mengelompokkan data pasien berdasarkan variabel usia, kondisi medis (diagnosis), jenis pelayanan, biaya tagihan, dan lama rawat inap. Proses *clustering* dilakukan melalui tahapan *preprocessing*, normalisasi data, serta pengujian menggunakan aplikasi *KNIME* sehingga menghasilkan

Submitted: **29/05/2026**; Accepted: **12/06/2026**; Published:**22/06/2026**

Jurnal Kolaborasi
Sains dan Ilmu Terapan
Volume 4 Nomor 2 Juni 2026 | Page -87-95
ISSN 2964- 5581 (media online)
https://utilityprojectsolution.org/ejournal/index.php/JuKSIT Penerbit : Cv. Utility Project Solution

pengelompokan data yang sistematis. Hasil *clustering* menunjukkan bahwa data pasien dapat dikelompokkan menjadi 4 *cluster*, yaitu (*cluster*_0) : pasien kondisi kritis dan kebutuhan penanganan medis yang tinggi, (*cluster*_1) : pasien dengan kondisi tidak terlalu kritis tetapi tetap membutuhkan penanganan medis yang cukup, *cluster*_2) : pasien dengan kondisi tidak kritis dan membutuhkan penanganan medis yang ringan, (*cluster*_3) : pasien dengan kondisi khusus. Dari hasil evaluasi *clustering*, diperoleh bahwa (*cluster*_0) memiliki jumlah anggota terbanyak, yang menunjukkan bahwa sebagian besar pasien memiliki kondisi yang membutuhkan perhatian lebih. Selain itu, setiap *cluster* memiliki karakteristik yang berbeda sehingga dapat menggambarkan pola segmentasi pasien secara jelas. Hasil segmentasi dalam penelitian ini menunjukkan bahwa algoritma *K-Means* mampu memberikan informasi yang berguna dalam memahami pola pasien berdasarkan riwayat kunjungan dan diagnosis, sehingga dapat digunakan sebagai dasar dalam mendukung pengambilan keputusan di bidang pelayanan kesehatan.

UCAPAN TERIMAKASIH

Terima kasih disampaikan kepada pihak-pihak yang telah mendukung terlaksananya penelitian ini.

REFERENSI

[1] A. Sapitri and Y. Afrilia, “Implementation of *Clustering* Method Using *K-Means* Algorithm for Grouping BPJS Health Patient Medical Record Data,” vol. 9, no. 5, 2025.

[2] L. Mayola, M. Hafizh, and H. Syahputra, “Klasterisasi Rumah Sakit berdasarkan Kunjungan Pasien menggunakan Algoritma *K-Means* : Data 2019-2023,” vol. 7, no. 1, pp. 15–21, 2025.

[3] G. Mafela, M. Sujak, H. N. Rofiq, and F. I. Tawakal, “Implementation of *K-Means Clustering* for Optimizing Non-Communicable Disease Budgets Implementasi *K-Means Clustering* untuk Optimalisasi Anggaran Penyakit Tidak Menular,” vol. 5, no. January, pp. 67–74, 2025.

[4] W. Aulia, A. Putera, U. Siahaan, L. Marlina, and M. Iqbal, “*K-Means Clustering* Algorithm Analysis For Grouping Patient Medical Record Data Based On Disease Type,” vol. 14, no. 04, pp. 832–843, 2024, doi: 10.54209/infosains.v14i04.

[5] T. A. Munandar, A. Yunizar, and Y. Pratama, “Regional *Clustering* Based on Types of Non-Communicable Diseases Using *K-Means* Algorithm,” vol. 23, no. 1, pp. 285–296, 2024, doi: 10.30812/matrik.v23i2.3352.

[6] M. Solihin, S. C. Shani, and T. Sari, “Segmentation of Clinical Patients Based on Visit Patterns and Diagnoses Using *Clustering* Algorithms at Klinik Pratama UIN Sunan Kalijaga,” vol. 6, no. 2, pp. 30–40, 2025.

[7] I. Lestari, “Formulasi Baru *K-Means* dalam Pengelompokkan Desa Berdasarkan Ketersediaan Fasilitas Kesehatan,” pp. 181–190, 2025.

[8] R. Anggraini, E. Haerani, and I. Afrianty, “Pengelompokkan Penyakit Pasien Menggunakan Algoritma *K-Means*,” vol. 9, no. 6, pp. 1840–1849, 2022, doi: 10.30865/jurikom.v9i6.5145.

[9] R. D. Putri and R. Muliono, “Penerapan Algoritma *K-Means* Untuk Klasterisasi Pasien Berdasarkan Riwayat Kesehatan dan Jenis Layanan Kesehatan,” vol. 5, no. 2, pp. 250–268, 2025, doi: 10.34007/incoding.v5i2.973.

[10] O. J. Harmaja, H. Halawa, W. S. Hulu, and S. Loi, “Implementasi Algoritma *K-Means Clustering* Untuk Pengelompokkan Penyakit Pasien Pada Puskesmas Pulo Brayan,” vol. 5, no. 1, pp. 150–157, 2023.

[11] R. B. Prasetyo, Y. A. Pranoto, and R. P. Prasetya, “IMPLEMENTASI DATA MINING MENGGUNAKAN ALGORITMA *K-MEANS CLUSTERING* PENYAKIT PASIEN RAWAT JALAN PADA KLINIK DR . ATIRAH,” vol. 7, no. 4, pp. 2144–2151, 2023.

[12] M. Andriani, A. Manaor, H. Pardede, and M. Simanjuntak, “Pengelompokan Penyakit pada Pasien Berdasarkan Usia dengan Metode *K-Means Clustering* ,” no. 4, 2024.

[13] H. Dilawati, “Klasterisasi Data Rekam Medis Pasien Menggunakan Metode,” vol. 5, no. 2, pp. 5–8, 2024.

[14] E. A. Herdiaman, A. Sudiarjo, M. Hikmatyar, T. Informatika, U. Perjuangan, and J. Barat, “RSUD MENGGUNAKAN METODE *K-MEANS*,” vol. 12, no. 3, 2024.

[15] I. Kanedi and E. Suryana, “Penerapan Metode *K-Means Clustering* Dalam Pengelompokan Data Pasien Rawat Inap Peserta BPJS Di Rumah Sakit Umum Daerah Kabupaten Kaur,” vol. 20, no. 2, pp. 493–500, 2024.

Submitted: **29/05/2026**; Accepted: **12/06/2026**; Published:**22/06/2026**

Jurnal Kolaborasi
Sains dan Ilmu Terapan
Volume 4 Nomor 2 Juni 2026 | Page -87-95
ISSN 2964- 5581 (media online)
https://utilityprojectsolution.org/ejournal/index.php/JuKSIT Penerbit : Cv. Utility Project Solution

[16] P. Supratman, I. Teknologi, B. Dian, and C. Cendikia, "PENERAPAN METODE *K-MEANS* UNTUK MENGELOMPOKKAN REKAM MEDIS PASIEN BERDASARKAN DIAGNOSA P," no. November, 2024.

[17] M. S. Sasmita, "IMPLEMENTASI ALGORITMA *K-MEANS* UNTUK *CLUSTERING* DATA PENYAKIT DI PUSKESMAS KOTAGEDE 2 YOGYAKARTA," vol. 4, no. 1, pp. 1–9, 2024.

[18] D. Natalia, B. Purba, M. Sihombing, and I. Ambarita, "Pengelompokan Data Rekam Medis pada Pasien Penyakit dalam Untuk Meningkatkan Manajemen Informasi Kesehatan Berdasarkan Wilayah Kota Binjai .," vol. 2, no. 3, 2024.

[19] G. Alasi and R. A. Putri, "Penerapan *K-Means Clustering* untuk Pengelompokan Pasien Rumah Sakit berdasarkan Tingkat Keparahan Penyakit," vol. 07, no. 03, pp. 475–486, 2025.

[20] I. N. Sjamsuddin, D. A. Firmansyah, and Y. Laferani, "Klasterisasi Pasien Rawat Inap BPJS pada RS Islam Assyifa Sukabumi menggunakan Metode *K-Means*," vol. 7, no. 01, pp. 355–368, 2025.

[21] S. H. Putra, "Data Mining untuk Pola Perbaikan Pelayanan Pasien dengan Metode Klustering *K-Means* pada Rumah Sakit Mitra Sejati," vol. 9, no. 3, pp. 1000–1012, 2025.

[22] A. A. Ardiansyah, E. A. Khaeroshi, and R. R. Wicaksono, "Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K- Means *Clustering* Di RSUD Muhammadiyah Bantul," vol. 13, no. 2, pp. 141–150, 2025.

Submitted: **29/05/2026**; Accepted: **12/06/2026**; Published:**22/06/2026**