Penerapan Algoritma *K-Means Clustering* untuk Segmentasi Pasien Berdasarkan Kondisi Kesehatan dan Frekuensi Kunjungan Kesehatan

Muhammad Mei Ahsan Nur[1*], Irgi Fahreza San Bondhy[2], Yusuf Febrianto[3], Desti Listia Sari[4]

[1]Teknik Informatika/Fakultas Ilmu Komputer
Universitas Duta Bangsa Surakarta
[1*]230103200@mhs.udb.ac.id

[2]Teknik Informatika/Fakultas Ilmu Komputer
Universitas Duta Bangsa Surakarta
[2]230103196@mhs.udb.ac.id

[3]Teknik Informatika/Fakultas Ilmu Komputer
Universitas Duta Bangsa Surakarta
[3]230103211@mhs.udb.ac.id

[1]Teknik Informatika/Fakultas Ilmu Komputer
Universitas Duta Bangsa Surakarta
[4]230103188@mhs.udb.ac.id

***Abstrak*— Perkembangan teknologi informasi di bidang kesehatan menghasilkan data pasien dalam jumlah besar yang memerlukan teknik analisis untuk menemukan pola yang dapat mendukung pengambilan keputusan. Berdasarkan kajian pustaka, algoritma *K-Means Clustering* merupakan salah satu metode *data mining* yang efektif dalam mengelompokkan data berdasarkan tingkat kemiripan karakteristik sehingga banyak diterapkan pada segmentasi data kesehatan. Penelitian ini bertujuan menerapkan algoritma *K-Means Clustering* untuk melakukan segmentasi pasien berdasarkan kondisi kesehatan dan frekuensi kunjungan ke fasilitas layanan kesehatan. Metode penelitian yang digunakan adalah *Knowledge Discovery in Databases (KDD)* yang meliputi seleksi data, *preprocessing*, transformasi data, *data mining*, dan evaluasi. Data penelitian berupa 2.000 data pasien yang diproses melalui pembersihan data, *encoding*, normalisasi, dan penentuan jumlah *cluster* menggunakan *Elbow Method*, kemudian dievaluasi menggunakan *Silhouette Score*. Hasil penelitian menunjukkan bahwa jumlah *cluster* yang optimal adalah dua, yaitu kelompok pasien dengan kondisi kesehatan relatif baik dan frekuensi kunjungan rendah serta kelompok pasien dengan risiko kesehatan lebih tinggi dan frekuensi kunjungan lebih tinggi. Nilai *Silhouette Score* sebesar 0,1429 menunjukkan kualitas pemisahan *cluster* masih rendah, namun hasil segmentasi mampu memberikan informasi yang bermanfaat untuk mendukung pengambilan keputusan, penentuan prioritas pelayanan, dan pengelolaan sumber daya pada fasilitas kesehatan.**
Kata kunci— K-Means Clustering, Segmentasi Pasien, Data Mining, KDD, Layanan Kesehatan.

***Abstract*— The advancement of information technology in the healthcare sector has led to the rapid growth of patient data, requiring analytical techniques to identify patterns that support data-driven decision-making. Based on the literature review, the K-Means Clustering algorithm is one of the most widely used data mining methods for grouping data according to similarity in characteristics, making it suitable for healthcare data segmentation. This study aims to implement the K-Means Clustering algorithm to segment patients based on their health conditions and the frequency of healthcare visits. The research employed the Knowledge Discovery in Databases (KDD) methodology, which consists of data selection, *preprocessing*, transformation, data mining, and evaluation. The dataset comprised 2,000 patient records that underwent data cleaning, encoding, *normalization*, optimal cluster determination using the Elbow Method, and evaluation using the Silhouette Score. The results indicate that the optimal number of clusters is two, representing patients with relatively good health conditions and low visit frequency, and patients with higher health risks and more frequent healthcare visits. The Silhouette Score of 0.1429 indicates that the cluster separation quality is relatively low; however, the segmentation results still provide valuable insights for supporting decision-making, determining service priorities, and optimizing resource management in healthcare facilities.**
Keywords— K-Means Clustering, Patient Segmentation, Data Mining, Knowledge Discovery in Databases (KDD), Healthcare Services.

I. PENDAHULUAN

Perkembangan teknologi informasi di bidang kesehatan menghasilkan lonjakan data pasien, seperti riwayat penyakit dan frekuensi kunjungan[1]. Besarnya volume data membuat analisis manual kurang efektif dalam menemukan pola tersembunyi[2]. Oleh karena itu, pendekatan *data mining* diperlukan untuk mengelompokkan data pasien secara terstruktur[1]. Dalam analisis data kesehatan, teknik *clustering*, khususnya algoritma *K-Means*, menjadi metode yang banyak digunakan[2].

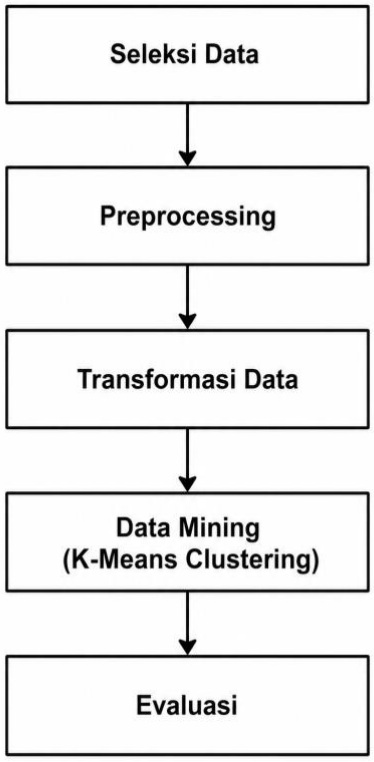
Penelitian menunjukkan bahwa metode *clustering* efektif dalam mengidentifikasi pola diagnosis dan penggunaan layanan kesehatan [3]. Selain itu, pendekatan *data-driven* ini mampu mengelompokkan pasien berdasarkan tingkat utilisasi layanan dan kondisi komorbiditas, menjadikannya relevan untuk mendukung analisis sistem kesehatan modern [4].

K-Means banyak digunakan karena sederhana, efisien, dan mampu menangani data besar [5]. Metode ini terbukti membantu mengidentifikasi pola tersembunyi [6], mengelompokkan penyakit [8], serta meningkatkan kualitas layanan rekam medis [9]. Meskipun penelitian mengenai penerapan *K-Means* di Indonesia telah memberikan informasi terstruktur bagi pengambil keputusan [7], model yang ada umumnya masih menggunakan variabel secara terpisah, seperti hanya penyakit atau pola kunjungan saja. Oleh karena itu, pendekatan yang mengintegrasikan kondisi kesehatan dan frekuensi kunjungan dalam satu model segmentasi masih belum banyak dikembangkan [10].

Perkembangan teknologi informasi di bidang kesehatan menghasilkan lonjakan data pasien yang kompleks, sehingga diperlukan pendekatan *data mining* seperti algoritma *K-Means* untuk mengidentifikasi pola tersembunyi yang sulit ditemukan melalui analisis konvensional. Meskipun penelitian sebelumnya menunjukkan bahwa *K-Means* efektif dalam mengelompokkan data medis, penyakit, dan pola layanan guna mendukung pengambilan keputusan, model yang ada umumnya masih menggunakan variabel secara terpisah. Oleh karena itu, penelitian ini berfokus pada penerapan *K-Means Clustering* yang mengintegrasikan kondisi kesehatan dan frekuensi kunjungan pasien dalam satu model segmentasi untuk menghasilkan kelompok yang lebih representatif, sehingga membantu fasilitas kesehatan dalam memahami karakteristik pasien, menentukan prioritas layanan, serta meningkatkan efisiensi pengelolaan sumber daya.

II. METODOLOGI PENELITIAN

Penelitian ini menggunakan metode *Knowledge Discovery in Databases (KDD)*, yaitu suatu proses untuk memperoleh pengetahuan yang bermanfaat dari sekumpulan data melalui beberapa tahapan yang sistematis, mulai dari seleksi data hingga evaluasi hasil *clustering*[10]. Penelitian dilakukan dengan menggunakan dataset sebanyak 2000 data pasien dari rumah sakit yang disimulasikan. Tahapan dalam proses *KDD* meliputi beberapa langkah, berikut:



Gambar 1. Tahapan Penelitian

Tahap pertama adalah seleksi data (data selection), yaitu proses memilih data dan atribut yang sesuai dengan tujuan penelitian sehingga informasi yang digunakan benar-benar relevan

untuk proses segmentasi pasien. Berdasarkan dataset yang digunakan, atribut yang dipilih meliputi Age, Gender, BMI, Insurance_Type, Primary_Condition,Num_Chronic_Conditions, Annual_Visits,Avg_Billing_Amount,Days_Since_Last_Visit,danPreventive_Care_Flag.

Pemilihan atribut tersebut didasarkan pada kemampuannya dalam merepresentasikan karakteristik pasien yang berkaitan dengan kondisi kesehatan serta intensitas kunjungan ke layanan kesehatan. Penggunaan atribut yang tepat diharapkan mampu menghasilkan cluster pasien yang lebih representatif [11].

Tahap berikutnya adalah *preprocessing* data, yang bertujuan untuk meningkatkan kualitas data sebelum dilakukan proses clustering. Pada tahap ini dilakukan pemeriksaan terhadap data yang hilang *(missing value)*, data duplikat, serta data yang tidak konsisten. Selain itu, atribut yang tidak memiliki kontribusi terhadap proses segmentasi dihapus, yaitu PatientID, State, City, dan Last_Visit_Date, karena atribut tersebut tidak menggambarkan karakteristik kesehatan pasien maupun frekuensi kunjungan. Melalui tahap ini dihasilkan dataset yang lebih bersih, konsisten, dan siap digunakan pada proses analisis berikutnya [12]

.Selanjutnya dilakukan transformasi data (data transformation) untuk mengubah data ke dalam format yang sesuai dengan algoritma clustering. Variabel kategorikal terlebih dahulu dikonversi menjadi bentuk numerik melalui proses *encoding*. Setelah itu dilakukan normalisasi menggunakan metode Min-Max Scaling agar seluruh atribut berada pada rentang nilai yang sama sehingga tidak terjadi dominasi atribut tertentu dalam proses pengelompokan.

Selain proses normalisasi, penelitian ini juga menerapkan Principal Component Analysis (PCA) sebagai teknik reduksi dimensi. Penggunaan PCA bertujuan untuk menyederhanakan data berdimensi tinggi menjadi dimensi yang lebih rendah sehingga proses visualisasi hasil clustering menjadi lebih mudah dilakukan. Pada penelitian ini digunakan dua principal component (PC1 dan PC2) sebagai representasi data dalam ruang dua dimensi. Pemilihan dua komponen utama tersebut dilakukan karena sudah cukup untuk menampilkan pola penyebaran data dan hubungan antar cluster secara lebih jelas. Dalam penelitian ini, PCA tidak digunakan sebagai metode evaluasi maupun analisis variasi data (explained variance), melainkan hanya sebagai teknik visualisasi hasil clustering sehingga proses interpretasi terhadap hasil pengelompokan pasien menjadi lebih mudah dipahami.

Tahap inti penelitian adalah data mining, yaitu proses pengelompokan pasien menggunakan algoritma K-Means Clustering berdasarkan karakteristik kondisi kesehatan dan frekuensi kunjungan layanan kesehatan. Implementasi algoritma dilakukan menggunakan bahasa pemrograman Python dengan bantuan library Scikit-learn (sklearn), sedangkan proses pengolahan data memanfaatkan Pandas dan NumPy.

Proses clustering dilakukan menggunakan dataset hasil preprocessing dan transformasi. Pada tahap preprocessing dilakukan penghapusan data duplikat, penanganan *missing value*, *encoding* pada atribut kategorikal, serta normalisasi menggunakan metode Min-Max Scaling agar seluruh data berada pada skala yang sama. Selanjutnya, jumlah cluster optimal ditentukan menggunakan Elbow Method, yang menghasilkan 3 cluster (K=3). Proses pengelompokan dilakukan menggunakan algoritma K-Means Clustering dengan parameter init = 'k-means++', n_init = 10, max_iter = 300, dan random_state = 42. Jarak antar data dan centroid dihitung menggunakan Euclidean Distance, kemudian centroid diperbarui secara berulang hingga proses clustering mencapai kondisi konvergen atau batas iterasi maksimum. Hasil implementasi menunjukkan jumlah cluster optimal sebanyak 3 cluster (K = 3) dengan nilai Silhouette Score sebesar 0,1364.

Tahap terakhir adalah evaluasi (evaluation) yang bertujuan untuk menilai kualitas hasil clustering yang diperoleh. Evaluasi dilakukan menggunakan Elbow Method untuk memastikan jumlah cluster yang paling optimal, sedangkan Silhouette Score digunakan untuk mengukur tingkat kekompakan anggota dalam suatu cluster dan tingkat pemisahan antar cluster. Semakin

tinggi nilai Silhouette Score, maka semakin baik kualitas cluster yang dihasilkan. Selain itu, hasil clustering divisualisasikan menggunakan Principal Component Analysis (PCA) sehingga distribusi setiap cluster dapat diamati secara lebih jelas. Hasil evaluasi tersebut selanjutnya digunakan sebagai dasar dalam menarik kesimpulan mengenai segmentasi pasien berdasarkan karakteristik kondisi kesehatan dan intensitas kunjungan layanan kesehatan.

III. HASIL DAN PEMBAHASAN

A. Data Selection

Dataset yang digunakan dalam penelitian ini berasal dari Kaggle dengan judul *Patient Segmentation Data*. Dataset tersebut terdiri atas 2.000 data pasien yang berisi informasi mengenai kondisi kesehatan serta riwayat kunjungan pasien ke layanan kesehatan. Sebelum digunakan dalam proses analisis, data terlebih dahulu melalui tahap seleksi atribut dan *preprocessing* untuk memastikan kualitas serta kesesuaiannya dengan metode yang digunakan. Selanjutnya, dataset diproses menggunakan algoritma *K-Means Clustering* untuk mengelompokkan pasien berdasarkan kemiripan karakteristik yang dimiliki. Atribut yang dipilih sebagai variable dalam proses *clustering* meliputi:

No	Atribut	Keterangan
1	Age	Usia pasien dalam satuan tahun.
2	BMI	Indeks massa tubuh pasien berdasarkan tinggi dan berat badan.
3	Num_Chronic_Conditions	Jumlah penyakit kronis yang dimiliki pasien.
4	Annual_Visits	Total kunjungan pasien ke fasilitas kesehatan dalam satu tahun.
5	Avg_Billing_Amount	Rata-rata biaya tagihan layanan kesehatan yang dikeluarkan pasien.
6	Days_Since_Last_Visit	Jumlah hari sejak kunjungan terakhir pasien ke fasilitas kesehatan.
7	Preventive_Care_Flag	Menunjukkan apakah pasien mengikuti program perawatan preventif (0 = Tidak, 1 = Ya).

Gambar 2. Hasil Atribut Data Selection

B. *Processing*

Tabel 1. Data Pasien Setelah Seleksi

ID	Age	BMI	Chronic Conditions	Annual Visits	Billing Amount	Days Since Last Visit
0	64	50.4	3	7	2.995,0	186
1	59	19.0	1	8	1.209,0	39
2	58	37.4	1	4	999,0	126
3	43	39.8	1	6	5.638,5	286
4	53	18.3	1	4	5.796,0	319

Tabel 2. Hasil Normalisasi Menggunakan Min-Max Scaling

ID	Age	BMI	Chronic Conditions	Annual Visits	Billing Amount	Days Since Last Visit
0	0,667	0,847	1,000	0,545	0,227	0,508
1	0,594	0,128	0,333	0,636	0,082	0,104
2	0,580	0,549	0,333	0,273	0,065	0,343
3	0,362	0,604	0,333	0,455	0,443	0,783
4	0,507	0,112	0,333	0,273	0,456	0,874

Tahap *preprocessing* dilakukan untuk mempersiapkan data sebelum proses clustering menggunakan algoritma K-Means. Proses ini meliputi penghapusan data duplikat, penanganan *missing value*, *encoding* pada atribut kategorikal, serta normalisasi menggunakan metode Min-Max Scaling. Tabel 1 menampilkan contoh hasil seleksi atribut yang digunakan dalam proses clustering, sedangkan Tabel 2 menunjukkan hasil normalisasi menggunakan *Min-Max Scaling*. Terlihat bahwa seluruh atribut numerik telah berada pada rentang nilai 0–1, sehingga tidak terdapat atribut yang mendominasi proses perhitungan jarak pada algoritma *K-Means*. Setelah tahap *preprocessing* selesai dilakukan, sebanyak 2.000 data pasien siap digunakan pada proses penentuan jumlah cluster dan pembentukan *cluster*. Setelah seluruh tahapan *preprocessing* selesai dilakukan, sebanyak 2.000 data pasien digunakan sebagai masukan pada proses penentuan jumlah *cluster* dan pembentukan cluster menggunakan algoritma *K-Means*.

C. Penentuan Jumlah Cluster Menggunakan Elbow Method

Pada tahap ini dilakukan penentuan jumlah cluster optimal menggunakan *Elbow Method* dengan menguji nilai K dari 1 sampai 10. Nilai *Within Cluster Sum of Squares* (WCSS) dihitung pada setiap nilai K untuk mengetahui jumlah *cluster* yang paling optimal. Berdasarkan Hasil pengujian menggunakan *Elbow Method* menunjukkan bahwa jumlah *cluster* yang optimal adalah K = 3. Oleh karena itu, proses pengelompokan data pasien pada penelitian ini dilakukan menggunakan 3 *cluster*. Pemilihan tiga *cluster* dinilai mampu memberikan keseimbangan yang baik dalam mengelompokkan pasien berdasarkan kemiripan karakteristik kondisi kesehatan dan frekuensi kunjungan ke fasilitas layanan kesehatan.

D. Hasil *Clustering* Menggunakan *K-Means*

Pada tahap ini dilakukan proses *clustering* menggunakan algoritma *K-Means* dengan jumlah *cluster* sebanyak 3, sesuai dengan hasil penentuan menggunakan *Elbow Method*. Proses *clustering* dilakukan menggunakan tujuh atribut, yaitu Age, BMI, Num_Chronic_Conditions, Annual_Visits, Avg_Billing_Amount, Days_Since_Last_Visit, dan Preventive_Care_Flag. Hasil proses ini mengelompokkan data pasien ke dalam tiga *cluster* berdasarkan tingkat kemiripan karakteristik yang dimiliki[13]. Hasil pengelompokan data pasien ke dalam tiga *cluster* berdasarkan tingkat kemiripan karakteristik disajikan pada Tabel 3 berikut.

Tabel 3. Hasil *Clustering* Pasien

Cluster	**Jumlah Pasien**	**Presentase**	**Karakteristik**
Cluster 0	585	29,3%	Pasien dengan kondisi kesehatan relatif baik dan frekuensi kunjungan rendah
Cluster 1	593	29,7%	Pasien dengan kondisi kesehatan sedang dan kunjungan rutin
Cluster 2	822	41,1%	Pasien dengan risiko kesehatan tinggi dan frekuensi kunjungan tinggi

Berdasarkan hasil *clustering*, *Cluster* 0 terdiri atas 585 pasien (29,3%) yang memiliki karakteristik kondisi kesehatan relatif baik dengan frekuensi kunjungan ke fasilitas layanan kesehatan yang rendah. *Cluster* 1 terdiri atas 593 pasien (29,7%) yang menunjukkan kondisi kesehatan sedang dengan pola kunjungan yang lebih rutin. Sementara itu, *Cluster* 2 terdiri atas 822 pasien (41,1%) yang memiliki risiko kesehatan lebih tinggi dengan frekuensi kunjungan ke fasilitas layanan kesehatan yang lebih tinggi dibandingkan *cluster* lainnya. Hasil ini menunjukkan bahwa algoritma *K-Means* mampu mengelompokkan pasien berdasarkan tingkat kemiripan karakteristik kesehatan dan intensitas kunjungan sehingga menghasilkan segmentasi pasien yang lebih informatif untuk mendukung pengambilan keputusan di bidang pelayanan kesehatan. Selanjutnya, nilai centroid dari masing-masing *cluster* disajikan pada Tabel 4.

Tabel 4. Nilai Centroid Setiap Cluster

Variable	**Cluster 0**	**Cluster 1**	**Cluster 2**
Age	32,64	58,23	58,11
BMI	29,70	30,73	31,49
Num_Chronic_Conditions	0,15	1,48	1,45
Annual_Visits	2,52	6,66	6,70
Avg_Billing_Amount	2.612,39	4.597,07	4.557,46
Days_Since_Last_Visit	175,05	181,09	182,95
Preventive_Care_Flag	0,34	0,52	0,51

Berdasarkan Tabel 4, nilai centroid menunjukkan karakteristik masing-masing *cluster*. *Cluster* 0 memiliki rata-rata usia,

jumlah penyakit kronis, frekuensi kunjungan, dan biaya pelayanan yang lebih rendah sehingga merepresentasikan pasien dengan kondisi kesehatan relatif baik. *Cluster* 1 menunjukkan pasien dengan kondisi kesehatan sedang dan kunjungan yang lebih rutin, sedangkan *Cluster* 2 memiliki karakteristik risiko kesehatan lebih tinggi dengan frekuensi kunjungan yang lebih tinggi dibandingkan *cluster* lainnya.

E. Evaluasi *Clustering*

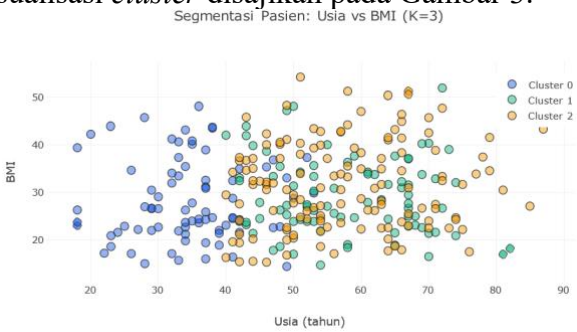
Evaluasi hasil *clustering* dilakukan menggunakan metode *Silhouette Score* untuk mengukur kualitas pengelompokan data yang dihasilkan oleh algoritma *K-Means*. Persamaan *Silhouette Score* yang digunakan dalam penelitian ini adalah sebagai berikut.

$$S(i) = \frac{b(i) - a(i)}{max(a(i), b(i))}$$

Hasil evaluasi menunjukkan bahwa proses *clustering* memperoleh *Silhouette Score* sebesar 0,1364. Nilai tersebut menunjukkan bahwa kualitas pemisahan antar-*cluster* masih tergolong rendah, sehingga masih terdapat kemiripan karakteristik antaranggota pada beberapa *cluster* [14]. Kondisi ini dapat disebabkan oleh adanya overlap karakteristik data pasien, seperti usia, BMI, jumlah penyakit kronis, dan frekuensi kunjungan yang memiliki rentang nilai yang saling berdekatan. Selain itu, atribut yang digunakan belum sepenuhnya merepresentasikan kondisi kesehatan pasien secara menyeluruh sehingga memengaruhi kualitas pemisahan *cluster*. Meskipun demikian, algoritma *K-Means* berhasil membentuk tiga *cluster* yang merepresentasikan kelompok pasien dengan kondisi kesehatan relatif baik, kondisi kesehatan sedang, dan risiko kesehatan tinggi. Untuk meningkatkan kualitas hasil *clustering* pada penelitian selanjutnya, dapat dilakukan penambahan atribut yang lebih representatif serta membandingkan algoritma *K-Means* dengan metode *clustering* lain agar diperoleh kualitas pemisahan *cluster* yang lebih baik.

F. Visualisasi Hasil *Clustering*

Visualisasi hasil *clustering* dilakukan menggunakan *scatter plot* dengan atribut Age sebagai sumbu X dan BMI sebagai sumbu Y. Visualisasi ini bertujuan untuk memperlihatkan persebaran data pasien berdasarkan hasil pengelompokan yang diperoleh dari algoritma *K-Means*. Hasil visualisasi *cluster* disajikan pada Gambar 3.



Gambar 3. Visualisasi Hasil Clustering (Age vs BMI)

Berdasarkan Gambar 3, terlihat bahwa data pasien telah terbagi ke dalam tiga *cluster* yang ditandai dengan warna yang berbeda. Persebaran data menunjukkan bahwa *Cluster* 0, *Cluster* 1, dan *Cluster* 2 memiliki karakteristik yang berbeda berdasarkan kombinasi usia dan nilai BMI. Visualisasi hasil *clustering* membantu memperjelas perbedaan karakteristik antar-*cluster* sehingga memudahkan interpretasi hasil segmentasi pasien sebagai dasar pendukung pengambilan keputusan dalam pelayanan kesehatan[15]. Meskipun masih terdapat beberapa titik yang berdekatan, secara umum hasil visualisasi menunjukkan bahwa proses *clustering* berhasil mengelompokkan pasien sesuai dengan karakteristik yang dimiliki.

IV. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma *K-Means Clustering* untuk melakukan segmentasi pasien berdasarkan kondisi kesehatan dan frekuensi kunjungan ke fasilitas layanan kesehatan menggunakan metode *Knowledge Discovery in Databases (KDD)*. Berdasarkan hasil *Elbow Method*, jumlah *cluster* yang optimal adalah 3 cluster, yaitu kelompok pasien dengan kondisi kesehatan relatif baik, kondisi kesehatan sedang, dan kelompok pasien dengan risiko

kesehatan tinggi. Hasil segmentasi ini menunjukkan bahwa algoritma *K-Means* mampu mengelompokkan pasien berdasarkan kemiripan karakteristik sehingga dapat digunakan sebagai dasar analisis karakteristik pasien.

Hasil evaluasi menggunakan *Silhouette Score* sebesar 0,1364 menunjukkan bahwa kualitas pemisahan antar-*cluster* masih tergolong rendah. Kondisi ini diduga dipengaruhi oleh adanya kemiripan karakteristik antar data pasien (*overlap*) serta keterbatasan atribut yang digunakan sehingga belum sepenuhnya merepresentasikan kondisi kesehatan pasien. Meskipun demikian, hasil segmentasi tetap memberikan informasi yang bermanfaat untuk membantu fasilitas kesehatan dalam mengidentifikasi kelompok pasien, menentukan prioritas pelayanan, serta mendukung perencanaan alokasi sumber daya secara lebih efektif.

Penelitian ini masih memiliki keterbatasan, yaitu menggunakan dataset simulasi dengan jumlah atribut yang terbatas serta hanya menerapkan algoritma *K-Means* sebagai metode *clustering*. Oleh karena itu, penelitian selanjutnya disarankan menggunakan dataset yang lebih beragam, menambahkan atribut yang lebih representatif, serta membandingkan performa *K-Means* dengan metode *clustering* lain, seperti *DBSCAN, Hierarchical Clustering,* atau *Gaussian Mixture* Model, untuk memperoleh kualitas segmentasi yang lebih baik.

REFERENSI

[1] 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.

[2] S. Adinda, A. Andrianingsih, and S. Informasi, "KLASTERISASI DATA PENYAKIT JANTUNG MENGGUNAKAN K-MEANS," vol. 2, no. 3, pp. 300–312, 2025.

[3] P. Fr, S. Sieranoja, K. Wikstr, and T. Laatikainen, "Clustering Diagnoses From 58 Million Patient Visits in Finland Between 2015 and 2018 Corresponding Author :," vol. 10, 2022, doi: 10.2196/35422.

[4] A. Gartner, R. Daniel, C. Slyne, and K. E. Nnoaham, "How predictive of future healthcare utilisation and mortality is data-driven population segmentation based on healthcare utilisation and chronic condition comorbidity ?," pp. 1–10, 2024.

[5] J. Peng, X. Zeng, J. Townsend, G. Liu, and Y. Huang, "A Machine Learning Approach to Uncovering Hidden Utilization Patterns of Early Childhood Dental Care Among Medicaid-Insured," vol. 8, no. January, pp. 1–10, 2021, doi: 10.3389/fpubh.2020.599187.

[6] A. Campione *et al.*, "Data-driven segmentation of type 2 diabetes mellitus patients : an observational study on health care utilisation prior to an emergency department visit in Germany," no. May, pp. 1–16, 2025, doi: 10.3389/fmed.2025.1509220.

[7] D. J. Anthonimuthu, N. N. Holm, A. Stockmarr, A. D. Zwisler, and F. W. Udsen, "Healthcare utilization and chronic condition clusters in multimorbidity patients using weighted k - means : a register - based study in Denmark," 2026.

[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] W. B. Laksono, Y. Syahidin, and Y. Yunengsih, "Implementasi Data Mining Klasterisasi Data Pasien Rawat Inap dengan Algoritma K-Means Clustering," vol. 7, no. 2, pp. 621–627, 2024, doi: 10.32493/jtsi.v7i2.39354.

[10] J. Pendidikan, "Segmentasi Pasien Rumah Sakit Berdasarkan Pola Kunjungan Menggunakan Algoritma K-Means Clustering untuk Optimasi Layanan Medis," vol. 5, no. 1, pp. 59–69, 2025.

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

[12] 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.

[13] R. D. Putri and R. Muliono, "Penerapan Algoritma K-Means Untuk Klasterisasi Pasien Berdasarkan Riwayat Kesehatan dan Jenis Layanan Kesehatan Application of K-Means Algorithm for Clustering of Patients Based on Health History and Type of Health Service," vol. 5, no. 2, pp. 250–268, 2025, doi: 10.34007/incoding.v5i2.973.

[14] M. D. A. ANIS FITRI NUR MASRURIYAH, MARDIAH and K. N. MALIK, "Pendekatan Unsupervised learning dalam Segmentasi Kesehatan: Perbandingan K-Means dan DBSCAN," 2025.

[15] W. Sukarjiyah, R. Jalan, and S. Pasien, "Strategi Segmentasi , Targeting , dan Positioning Berbasis Clustering K-Means untuk Optimalisasi Layanan Poliklinik Rumah Sakit," pp. 176–189, 2022