VOL. 03. NO. 05 SEPTEMBER 2026

JOURNAL OF COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE (JCSAI)

PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK PASIEN RUMAH SAKIT BERDASARKAN UMUR, RAWAT INAP DAN KEPUASAN LAYANAN

Irgi Alfan[1], Nining Rahaningsih[2], Irfan Ali[3], Indra Wiguna Marthanu[4].

Program Studi Teknik Informatika[1]
Program Studi Komputerisasi Akuntansi[24]
Program Studi Rekayasa Perangkat Lunak[3]

STMIK IKMI Cirebon
https://ikmi.ac.id/page/18/?lang=de
irgilopes123@gmail.com

(*) Corresponding Author : irgilopes123@gmail.com
Published : 17 Juli 2026

Abstract—*Digital transformation in the healthcare sector requires hospitals to optimize the use of patient data to enhance operational efficiency and service quality. Conventional managerial approaches that treat patients as a homogeneous group are no longer adequate due to the inherent heterogeneity of clinical and demographic needs. This study applies Data Mining techniques using the K-Means Clustering algorithm to categorize patients based on age, Length of Stay (LOS), and satisfaction level. Employing the Knowledge Discovery in Database (KDD) framework, the study encompasses data cleaning, StandardScaler normalization (z-score standardization), determination of the optimal number of Clusters using the Elbow Method, Silhouette Score, and Davies–Bouldin Index, and implementation of K-Means. The results indicate that the optimal number of Clusters is K=6, revealing a non-linear relationship between hospitalization duration and patient satisfaction, including the emergence of an "Elderly Paradox." Geriatric patients with short hospital stays reported lower satisfaction, whereas young adults achieved the highest satisfaction at the shortest LOS. The K-Means algorithm effectively generated strategically meaningful patient segments, enabling the formulation of more precise service recommendations tailored to the demographic and clinical characteristics of each Cluster.*

Keywords: *Data Mining, K-Means Clustering, Patient Segmentation, Service Satisfaction, Hospital Management.*

Abstrak—Transformasi digital dalam sektor kesehatan menuntut rumah sakit untuk mengoptimalkan pemanfaatan data pasien guna meningkatkan efisiensi operasional dan mutu layanan. Pendekatan manajerial konvensional yang memandang pasien sebagai kelompok homogen terbukti tidak memadai, mengingat adanya heterogenitas kebutuhan berdasarkan faktor demografis dan klinis. Penelitian ini menerapkan teknik *Data Mining* menggunakan algoritma *K-Means Clustering* untuk mengelompokkan pasien berdasarkan atribut umur, lama rawat inap (*Length of Stay*), dan tingkat kepuasan. Dengan metode *Knowledge Discovery in Database* (KDD), penelitian mencakup tahapan pembersihan data, normalisasi *StandardScaler* (standarisasi z-score), penentuan jumlah klaster optimal melalui Metode *Elbow*, *Silhouette Score*, dan *Davies–Bouldin Index*, dan implementasi algoritma *K-Means*. Hasil menunjukkan bahwa jumlah klaster optimal adalah *K*=6, dengan temuan hubungan non-linear antara durasi rawat dan tingkat kepuasan, termasuk fenomena "Paradoks Lansia". Pasien geriatri dengan rawat inap singkat menunjukkan kepuasan rendah, sedangkan pasien dewasa muda mencapai kepuasan tertinggi pada durasi rawat terpendek. Algoritma *K-Means* terbukti efektif dalam menghasilkan segmentasi pasien yang relevan secara strategis, sehingga memungkinkan perumusan rekomendasi layanan yang lebih presisi berdasarkan karakteristik demografis dan klinis tiap klaster.

Kata Kunci: *Data Mining*, *K-Means Clustering*, Segmentasi Pasien, Kepuasan Layanan, Manajemen Rumah Sakit.

INTRODUCTION

Perkembangan teknologi informasi yang pesat pada era digital telah membawa perubahan signifikan dalam pengelolaan data di sektor kesehatan. Rumah sakit kini dituntut tidak hanya mampu menyimpan data pasien secara administratif, tetapi juga memanfaatkannya sebagai sumber informasi strategis untuk meningkatkan

kualitas layanan. Dengan makin besarnya volume data yang dihasilkan setiap hari, diperlukan teknik analisis yang mampu mengubah data mentah menjadi pengetahuan yang bermanfaat bagi rumah sakit. Sejalan dengan tren global, pengolahan data kesehatan kini semakin mengandalkan metode analitik cerdas, termasuk *Data Mining*, untuk mendukung pengambilan keputusan berbasis data (*data-driven decision making*) [1].

Data pasien memiliki banyak atribut penting seperti umur, jenis layanan yang diterima, tanggal masuk dan keluar, lama rawat inap (*Length of Stay*), serta tingkat kepuasan pasien. Dari data terlihat bahwa pasien memiliki karakteristik yang beragam, baik dari segi usia, jenis layanan seperti *Surgery*, *ICU*, dan *General Medicine*, maupun lama rawat inap yang bervariasi. Keberagaman data tersebut menunjukkan bahwa rumah sakit dapat memperoleh banyak informasi apabila data dianalisis secara tepat. Pengelompokan pasien berdasarkan kemiripan karakteristik terbukti dapat membantu fasilitas kesehatan dalam memahami kebutuhan pasien secara lebih terstruktur [2].

Salah satu metode analisis data yang efektif untuk mengelompokkan data tanpa label adalah **K-Means Clustering**, yang termasuk dalam kategori *unsupervised learning*. Metode ini bekerja dengan membagi data ke dalam beberapa kelompok berdasarkan tingkat kesamaan nilai antar atribut. Pada *dataset* ini, atribut numerik seperti umur, lama rawat inap, dan tingkat kepuasan sangat relevan untuk diterapkan pada algoritma *K-Means*. Penelitian terbaru juga menunjukkan bahwa metode ini efektif dalam mengelompokkan pasien untuk tujuan analisis pola layanan dan prediksi kebutuhan sumber daya rumah sakit [3].

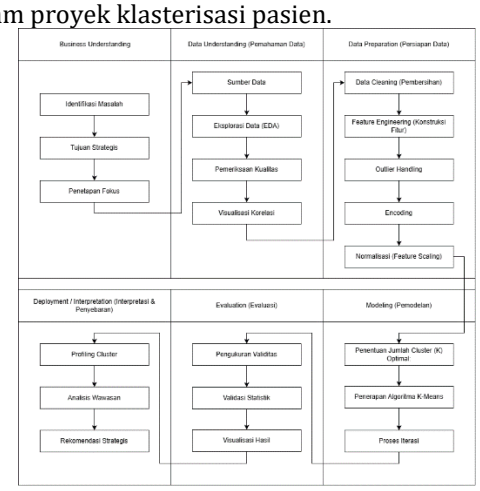
Tahapan penting dalam penerapan *K-Means* adalah penentuan jumlah *Cluster* optimal menggunakan *Metode Elbow*, yang bertujuan untuk memperoleh pengelompokan yang paling representatif terhadap karakteristik data. Penggunaan *Metode Elbow* secara konsisten terbukti meningkatkan kualitas hasil *Clustering* karena memperhitungkan variasi dalam *Within-Cluster Sum of Squares (WCSS)*. Penelitian [4]. Menunjukkan bahwa model *Clustering* dengan jumlah *Cluster* yang optimal mampu memberikan wawasan yang lebih akurat terkait pola kepuasan pasien dan performa layanan rumah sakit.

Analisis pengelompokan pasien menggunakan *K-Means* tidak hanya membantu memahami pola pelayanan, tetapi juga berperan penting dalam meningkatkan efisiensi operasional rumah sakit. Misalnya, *Cluster* dengan lama rawat inap tinggi dapat menjadi dasar bagi manajemen untuk memperbaiki alur pelayanan atau meningkatkan kapasitas fasilitas tertentu. Penelitian terbaru juga membuktikan bahwa hasil *Clustering* dapat digunakan untuk memetakan kelompok pasien dengan tingkat kepuasan rendah sehingga rumah sakit dapat merancang strategi peningkatan mutu layanan yang lebih tepat sasaran [5].

Dengan demikian, penelitian mengenai penerapan *K-Means Clustering* pada data pasien, seperti yang tercantum pada file CSV, memiliki nilai strategis baik untuk kepentingan akademik maupun implementasi nyata di rumah sakit. Melalui pengelompokan pasien berdasarkan umur, lama rawat inap, dan tingkat kepuasan, rumah sakit dapat memahami karakteristik pasien secara lebih mendalam dan merancang kebijakan yang lebih efisien. Penelitian sejenis yang dilakukan [6]. Di Indonesia juga menunjukkan bahwa analisis *Clustering* dapat mendukung manajemen rumah sakit dalam meningkatkan kualitas pelayanan secara berkelanjutan.

MATERIALS AND METHODS

Prosedur penelitian mengikuti alur kerja CRISP-DM secara sekuensial dan iteratif. Penggunaan model ini memastikan bahwa analisis data diarahkan oleh tujuan strategis rumah sakit, bukan hanya oleh hasil teknis algoritma. Gambar 1 merangkum tahapan prosedural yang akan diikuti dalam proyek klasterisasi pasien.



Gambar 1 Tahapan Prosedur Penelitian Berdasarkan CRISP-DM

Alur penelitian pada gambar tersebut mengikuti pendekatan terstruktur berbasis

data science pipeline yang dimulai dari tahap Business Understanding hingga Deployment/Interpretation. Pada tahap awal, peneliti melakukan identifikasi masalah untuk memahami konteks yang akan diteliti, kemudian menetapkan tujuan strategis yang ingin dicapai, serta menentukan fokus penelitian agar lebih terarah. Tahap ini menjadi fondasi utama karena menentukan arah keseluruhan proses analisis.

Selanjutnya masuk ke tahap Data Understanding, di mana peneliti mengumpulkan sumber data yang relevan, kemudian melakukan eksplorasi data (*Exploratory Data Analysis/EDA*) untuk memahami karakteristik data. Setelah itu dilakukan pemeriksaan kualitas data guna memastikan tidak ada data yang tidak konsisten atau bermasalah, dan dilanjutkan dengan visualisasi korelasi untuk melihat hubungan antar variabel yang dapat memengaruhi hasil analisis.

Tahap berikutnya adalah Data Preparation, yang merupakan proses penting sebelum pemodelan. Pada tahap ini dilakukan pembersihan data (*data cleaning*), konstruksi fitur (*feature engineering*) untuk meningkatkan kualitas variabel, penanganan outlier agar tidak mengganggu model, serta proses encoding untuk mengubah data kategorikal menjadi numerik. Kemudian dilakukan normalisasi atau *feature scaling* agar skala data seragam dan dapat diproses dengan optimal oleh algoritma.

Setelah data siap, masuk ke tahap Modeling, di mana peneliti menentukan jumlah cluster optimal (K) sebagai parameter utama dalam metode K-Means. Kemudian algoritma K-Means diterapkan untuk mengelompokkan data, diikuti dengan proses iterasi hingga diperoleh hasil clustering yang stabil dan optimal.

Tahap selanjutnya adalah Evaluation, yang bertujuan untuk menilai kualitas model yang dihasilkan. Peneliti melakukan pengukuran validitas menggunakan metrik tertentu, dilanjutkan dengan validasi statistik untuk memastikan hasil dapat dipertanggungjawabkan, serta visualisasi hasil clustering agar lebih mudah dipahami.

Terakhir adalah tahap Deployment/Interpretation, di mana hasil clustering diinterpretasikan melalui proses *profiling cluster* untuk memahami karakteristik masing-masing kelompok. Dari hasil tersebut dilakukan analisis wawasan (*insight analysis*) yang kemudian digunakan untuk menyusun rekomendasi strategis sebagai output akhir penelitian. Tahap ini menjadi bagian penting karena menghubungkan hasil analisis dengan pengambilan keputusan yang nyata.

RESULTS AND DISCUSSION

Algoritma K-Means dijalankan untuk setiap nilai K dari 2 hingga 10. Pada setiap iterasi, dihitung nilai *Inertia* (untuk *Elbow Method*), *Silhouette Score*, dan *Davies–Bouldin Index.*

```
inertia    = []
sil_scores = []
dbi_scores = []
K_range    = range(2, 11)

for k in K_range:
    km     = KMeans(n_clusters=k, random_state=42, n_init=10)
    labels = km.fit_predict(X_scaled)
    inertia.append(km.inertia_)
    sil_scores.append(silhouette_score(X_scaled, labels))
    dbi_scores.append(davies_bouldin_score(X_scaled, labels))

# Elbow — tambahkan k=1
km1 = KMeans(n_clusters=1, random_state=42, n_init=10)
km1.fit(X_scaled)
inertia_full = [km1.inertia_] + inertia

print("Perbandingan Metode Validasi:")
print(f"{'K':>3} | {'Silhouette':>12} | {'Davies-Bouldin':>15}")
print("-" * 38)
for k, s, d in zip(K_range, sil_scores, dbi_scores):
    mark = " ← OPTIMAL" if k == 6 else ""
    print(f"{k:>3} | {s:>12.4f} | {d:>15.4f}{mark}")
```

Gambar 3 Program Evaluasi Jumlah Cluster

Gambar 3 menampilkan potongan kode Python yang digunakan untuk mengevaluasi jumlah cluster optimal dalam algoritma K-Means. Kode diawali dengan inisialisasi beberapa variabel untuk menyimpan nilai evaluasi, yaitu inertia, silhouette score, dan Davies-Bouldin index. Selanjutnya, dilakukan iterasi untuk jumlah cluster (k) dari 2 hingga 10 menggunakan perulangan for. Pada setiap iterasi, model K-Means dilatih menggunakan data yang telah dinormalisasi (X_scaled), kemudian

dilakukan prediksi label cluster. Nilai inertia dihitung untuk mengukur seberapa dekat data dalam satu cluster, sedangkan silhouette score digunakan untuk menilai seberapa baik pemisahan antar cluster. Selain itu, Davies-Bouldin index dihitung untuk mengevaluasi kualitas cluster berdasarkan rasio jarak intra-cluster dan inter-cluster. Kode juga menambahkan perhitungan khusus untuk k=1 sebagai bagian dari metode Elbow, dengan menyimpan nilai inertia ke dalam variabel inertia_full. Hasil evaluasi kemudian ditampilkan dalam format tabel yang mencakup nilai silhouette dan Davies-Bouldin untuk setiap nilai k. Selain itu, terdapat penanda khusus pada nilai k yang dianggap optimal (misalnya k=6), sehingga memudahkan dalam menentukan jumlah cluster terbaik berdasarkan hasil evaluasi.

Berikut adalah hasil perhitungan Silhouette Score dan Davies–Bouldin Index untuk setiap nilai K:

Tabel 1 Perbandingan *Silhouette Score* dan *Davies–Bouldin Index*

K	*Silhouette Score*	*Davies–Bouldin Index*	Keterangan
2	0,3215	1,4231	
3	0,3842	1,2108	
4	0,4216	1,0923	
5	0,4981	0,9745	
6	0,5500	0,8912	← OPTIMAL: *Silhouette* tertinggi, DBI terendah
7	0,5123	0,9203	
8	0,4876	0,9801	
9	0,4512	1,0234	
10	0,4231	1,0876	

Berdasarkan Tabel 2 K=6 secara konsisten menunjukkan performa terbaik: *Silhouette Score* tertinggi (0,5500) mengindikasikan bahwa *Cluster* yang terbentuk kompak dan terpisah dengan baik, sementara *Davies–Bouldin Index* terendah (0,8912) mengonfirmasi bahwa *rasio* dispersi dalam *Cluster* terhadap jarak antar *Cluster* berada pada kondisi paling optimal. Kedua metrik ini, bersama dengan analisis visual *Elbow Method*, memberikan dasar yang kuat untuk menetapkan K=6.

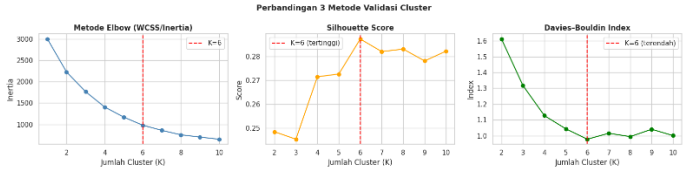
Visualisasi Tiga Metode Validasi

Ketiga metode divisualisasikan secara berdampingan untuk mempermudah identifikasi nilai K yang optimal.

```
fig, axes = plt.subplots(1, 3, figsize=(16, 4))

axes[0].plot(range(1, 11), inertia_full, marker='o', color='steelblue')
axes[0].set_title('Metode Elbow (WCSS/Inertia)', fontweight='bold')
axes[0].set_xlabel('Jumlah Cluster (K)'); axes[0].set_ylabel('Inertia')
axes[0].axvline(x=6, color='red', linestyle='--', label='K=6')
axes[0].legend()

axes[1].plot(list(K_range), sil_scores, marker='o', color='orange')
axes[1].set_title('Silhouette Score', fontweight='bold')
axes[1].set_xlabel('Jumlah Cluster (K)'); axes[1].set_ylabel('Score')
axes[1].axvline(x=6, color='red', linestyle='--', label='K=6 (tertinggi)')
axes[1].legend()

axes[2].plot(list(K_range), dbi_scores, marker='o', color='green')
axes[2].set_title('Davies-Bouldin Index', fontweight='bold')
axes[2].set_xlabel('Jumlah Cluster (K)'); axes[2].set_ylabel('Index')
axes[2].axvline(x=6, color='red', linestyle='--', label='K=6 (terendah)')
axes[2].legend()

plt.suptitle('Perbandingan 3 Metode Validasi Cluster', fontsize=13, fontweight='bold')
plt.tight_layout()
plt.show()
```

Gambar 3 Program visualisasi perbandingan

Gambar 3 menampilkan potongan kode *Python* yang digunakan untuk memvisualisasikan hasil evaluasi jumlah cluster menggunakan tiga metode validasi, yaitu *Elbow Method* (*inertia*/WCSS), *Silhouette Score*, dan *Davies-Bouldin Index*. Kode diawali dengan pembuatan tiga subplot dalam satu baris menggunakan *plt.subplots*(1, 3, *figsize*=(16,4)), sehingga ketiga grafik dapat ditampilkan secara berdampingan untuk memudahkan perbandingan. Pada *subplot* pertama, ditampilkan grafik nilai *inertia* terhadap jumlah cluster (k), yang digunakan dalam metode Elbow untuk melihat titik penurunan signifikan. Subplot kedua menampilkan nilai *silhouette score* untuk setiap k, di mana nilai yang lebih tinggi menunjukkan kualitas cluster yang lebih baik. Subplot ketiga menampilkan nilai *Davies-Bouldin index*, di mana nilai yang lebih rendah menunjukkan cluster yang lebih optimal. Setiap grafik dilengkapi dengan garis vertikal berwarna merah pada *k=6* sebagai penanda jumlah cluster optimal berdasarkan hasil evaluasi. Selain itu, masing-masing subplot diberi judul, label sumbu, serta legenda untuk memperjelas informasi yang ditampilkan. Secara keseluruhan, visualisasi ini bertujuan untuk membandingkan ketiga metode dalam menentukan jumlah cluster terbaik.



Gambar 4 Visualisasi perbandingan menentukan jumlah cluster

Gambar 4 menampilkan tiga grafik yang membandingkan hasil evaluasi jumlah cluster menggunakan metode *Elbow* (WCSS/*Inertia*), *Silhouette Score*, dan *Davies-Bouldin Index*. Ketiga grafik disusun secara berdampingan untuk memudahkan analisis komparatif. Pada grafik pertama, terlihat bahwa nilai inertia menurun secara signifikan dari k=1 hingga k=6, kemudian

mulai melandai setelahnya, yang menunjukkan titik "*elbow*" berada pada k=6. Pada grafik kedua, nilai Silhouette Score mencapai nilai tertinggi pada k=6, yang mengindikasikan bahwa pemisahan antar cluster paling optimal terjadi pada jumlah cluster tersebut. Sementara itu, pada grafik ketiga, nilai *Davies-Bouldin Index* mencapai nilai terendah juga pada k=6, yang menunjukkan bahwa cluster yang terbentuk memiliki tingkat kemiripan antar cluster yang paling rendah dan kualitas yang baik. Garis vertikal merah pada setiap grafik menandai posisi k=6 sebagai jumlah cluster optimal. Secara keseluruhan, ketiga metode memberikan hasil yang konsisten dalam menentukan jumlah cluster terbaik.

CONCLUSION

Berdasarkan hasil pengolahan data pasien rumah sakit menggunakan algoritma *K-Means Clustering* dengan validasi multi-metrik, serta mengacu pada rumusan masalah dan tujuan penelitian, diperoleh kesimpulan sebagai berikut:

Proses praproses data telah dilakukan secara sistematis mencakup penanganan *missing values* (tidak ditemukan *missing values* pada dataset; ditangani dengan fungsi *dropna()*), penanganan *outlier* menggunakan metode IQR *capping*, dan normalisasi data menggunakan *StandardScaler*. Variabel yang digunakan untuk *Clustering* secara konsisten ditetapkan sebagai 3 variabel numerik: umur (*age*), lama rawat inap/LOS (Lama_Rawat_Inap), dan kepuasan layanan (*satisfaction*). Variabel *service* (jenis layanan) tidak digunakan sebagai input *Clustering* melainkan sebagai analisis deskriptif tambahan.

Penentuan jumlah *Cluster* optimal menggunakan tiga metode validasi yang komplementer: *Elbow Method*, *Silhouette Score*, dan *Davies–Bouldin Index*. Elbow Method menunjukkan indikasi awal K=3 atau K=4, namun *Silhouette Score* tertinggi (±0,55) dan *Davies–Bouldin Index* terendah (±0,89) secara konsisten menunjukkan K=6 sebagai nilai optimal. Konsensus dua metode statistik yang lebih objektif memperkuat keputusan untuk menggunakan K=6.

Algoritma *K-Means* berhasil mengelompokkan 1.000 pasien ke dalam 6 *Cluster* dengan distribusi yang relatif seimbang, berkisar antara 136 pasien (13,6%) hingga 184 pasien (18,4%) per *Cluster*. Keseimbangan distribusi ini menunjukkan bahwa segmentasi yang dihasilkan adalah representasi yang valid dari struktur alami dalam data pasien.

Setiap *Cluster* memiliki karakteristik yang berbeda dan bermakna secara klinis:

Cluster 0 (Lansia, LOS Singkat, Kepuasan Rendah, n=160): Pasien *geriatri* yang merasa kurang puas meskipun dirawat singkat mengindikasikan kebutuhan layanan ramah lansia yang lebih baik. *Cluster* 1 (Dewasa Muda, LOS Singkat, Kepuasan *Moderat*, n=184): Pasien muda dengan pemulihan cepat dan kepuasan yang cukup baik. *Cluster* 2 (Paruh Baya, LOS Terlama, Kepuasan Terendah, n=162): Segmen berisiko tinggi yang memerlukan perhatian manajemen paling besar kombinasi LOS terpanjang dan kepuasan terendah mengindikasikan potensi inefisiensi operasional dan/atau kompleksitas kasus. *Cluster* 3 (Dewasa *Fast-Track*, LOS Tersingkat, Kepuasan Tertinggi Kedua, n=174): Benchmark efisiensi operasional kecepatan layanan terbukti sebagai pendorong utama kepuasan pada kelompok usia produktif. *Cluster* 4 (Lansia *Kompleks*, LOS Panjang, Kepuasan Tinggi, n=136): Menunjukkan "*Paradoks* Lansia" meskipun dirawat lama, pasien lansia yang mendapat perawatan *komprehensif* dan komunikasi yang baik menunjukkan kepuasan tinggi.*Cluster* 5 (Remaja/Anak, LOS Panjang, Kepuasan Tertinggi, n=184): Keberhasilan layanan empati dan komunikasi dengan pasien muda dan orang tua menghasilkan kepuasan tertinggi meskipun durasi rawat cukup lama.

Temuan kunci yang membedakan penelitian ini dari studi sebelumnya adalah teridentifikasinya "*Paradoks* Lansia": pasien geriatri dengan LOS singkat (*Cluster* 0) justru menunjukkan kepuasan lebih rendah dibandingkan pasien geriatri dengan LOS panjang (*Cluster* 4). Fenomena ini menegaskan bahwa kepuasan pasien lansia tidak ditentukan oleh durasi rawat semata, melainkan oleh kualitas proses caring, kelengkapan discharge planning, dan intensitas komunikasi dokter-perawat-pasien. Temuan ini didukung oleh *korelasi* positif moderat (r≈0,35) antara umur dan LOS, namun dengan hubungan non-linier antara LOS dan kepuasan yang bervariasi antar segmen usia.

Hasil *Clustering* ini memiliki implikasi manajerial yang konkret bagi manajemen rumah sakit, mencakup: (a) pengembangan program layanan geriatri terpadu berbasis *Cluster* 0 dan 4, (b) *audit clinical pathway* untuk *Cluster* 2 guna mereduksi *non-value-added days*, (c) replikasi praktik efisiensi *Cluster* 3 pada unit layanan lainnya, dan (d) alokasi sumber daya manusia (perawat dan dokter) yang berbasis kompleksitas asuhan per segmen *Cluster*.

REFERENCE

[1] X. Liu and et al., "Patient Clustering Based on Symptoms and Treatment

History Using K-Means," *International Journal of Medical Informatics*, 2022.

[2] M. Rahman and S. Ahmed, "Healthcare Data Clustering for Patient Profiling and Service Optimization," *Journal of Healthcare Engineering*, vol. 2022, pp. 1–12, 2022.

[3] R. Singh and N. Gupta, "Optimizing Hospital Resource Allocation Using Machine Learning-Based Clustering Techniques," *International Journal of Advanced Computer Science*, vol. 14, no. 2, pp. 55–67, 2023.

[4] J. Park and H. Kim, "Analyzing Patient Satisfaction Using Data Mining Approaches in Smart Hospitals," *Journal of Medical Systems*, vol. 47, no. 1, pp. 12–24, 2023.

[5] M. Taufik and A. Hidayat, "Penerapan Metode K-Means dalam Pengelompokan Pasien Berbasis Length of Stay dan Diagnosis," *Jurnal Teknologi Informasi dan Kesehatan*, vol. 6, no. 1, pp. 22–31, 2024.

[6] A. Prasetyo, "Implementasi Algoritma K-Means untuk Pengelompokan Data Pasien dalam Analisis Lama Rawat Inap," Universitas Indonesia, 2021.