

KLASTERISASI DATA REKAM MEDIS PASIEN RAWAT INAP MENGUNAKAN METODE K-MEANS PADA RUMAH SAKIT DR. TADJUDDIN CHALID

Nur Fadilah¹, Muh. Fajri Raharjo², Amhar Davi Dewantara³

^{1,2,3} Teknik Informatika dan Komputer, Politeknik Negeri Ujung Pandang, Indonesia

Info Artikel

Riwayat artikel:

Received, (28 Mei 2026)

Revised, (10 Juni 2026)

Accepted, (22 Juni 2026)

Kata kunci:

data mining;

K-Means;

rekam medis;

silhouette index;

web dashboard

ABSTRACT

Electronic Medical Records (EMR) in hospitals generate large volumes of data that have not been optimally utilized for management decision-making. This study applies the K-Means algorithm to cluster inpatient medical record data at Dr. Tadjuddin Chalid Hospital and visualizes the results through an interactive dashboard. The dataset covers 1,848 patient records with attributes of diagnosis, residential area, age, and gender. The research stages include preprocessing, data transformation, K-Means clustering, and validation using the Silhouette Index. The results show that the optimal model is formed at K=2 with a Silhouette Index value of 0.8030 (strong structure). Cluster 1 is dominated by local patients (close proximity) aged adults to elderly with primary diagnoses of typhoid fever and dyspepsia. Conversely, cluster 2 is dominated by pediatric patients from outside Makassar City with a primary diagnosis of neonatal jaundice. This study also successfully developed a web-based dashboard to visualize the clustering results interactively to support management in data-driven decision-making.

ABSTRAK

Rekam Medis Elektronik (RME) di rumah sakit menghasilkan volume data besar yang belum dimanfaatkan secara optimal untuk pengambilan keputusan manajemen. Penelitian ini menerapkan algoritma K-Means untuk melakukan klusterisasi data rekam medis rawat inap di Rumah Sakit Dr. Tadjuddin Chalid dan memvisualisasikan hasilnya melalui *dashboard* interaktif. Dataset yang digunakan mencakup 1.848 data pasien dengan atribut diagnosis, wilayah tinggal, usia, dan jenis kelamin. Tahapan penelitian meliputi *preprocessing*, transformasi data, klusterisasi K-Means, dan validasi dengan *Silhouette Index*. Hasil pengujian menunjukkan model optimal terbentuk pada K=2 dengan nilai *Silhouette Index* sebesar 0,8030 (struktur sangat kuat). Klaster 1 didominasi pasien lokal (domisili dekat) usia dewasa hingga lansia dengan diagnosis utama *typhoid fever* dan *dyspepsia*. Sebaliknya, klaster 2 didominasi pasien anak-anak dari luar Kota Makassar dengan diagnosis utama *neonatal jaundice*. Penelitian ini juga berhasil mengembangkan *dashboard* berbasis web untuk memvisualisasikan hasil klusterisasi secara interaktif guna mendukung pihak manajemen dalam pengambilan keputusan berbasis data.

Penulis Korespondensi:

Nur Fadilah

Teknik Informatika dan Komputer, Politeknik Negeri Ujung Pandang, Jl. Perintis Kemerdekaan KM 10 Makassar

Email: nurfadilahh038@gmail.com

1. PENDAHULUAN

Perkembangan teknologi informasi saat ini telah mendorong berbagai instansi dan lembaga daerah untuk melakukan transformasi tata kelola data melalui implementasi aplikasi manajemen fungsional guna meningkatkan efisiensi layanan publik [20]. Sektor pelayanan kesehatan menjadi salah satu domain utama yang mengalami akselerasi transformasi ini, di mana pemerintah mendorong pengelolaan rekam medis konvensional menjadi rekam medis elektronik (RME) sebagaimana diatur dalam Peraturan Menteri Kesehatan Nomor 24 Tahun 2022 [11]. Melalui RME, data pasien tersimpan secara digital dan terintegrasi sehingga memudahkan proses pencatatan, penyimpanan, dan pengelolaan informasi kesehatan. Implementasi rekam medis elektronik secara kualitatif terbukti mampu meningkatkan efisiensi dan mutu pelayanan, serta mendukung keselamatan pasien melalui ketersediaan data yang cepat dan akurat [1], [5].

Setiap pelayanan medis yang diberikan kepada pasien menghasilkan data rekam medis baru yang tersimpan di dalam basis data RME. Akumulasi data tersebut terus meningkat dari waktu ke waktu dan membentuk kumpulan data dalam jumlah besar. Akumulasi data yang melimpah tersebut berpotensi menyimpan pola dan informasi yang bernilai bagi rumah sakit apabila dilakukan pengolahan lebih lanjut. Informasi yang diperoleh dapat dimanfaatkan sebagai dasar penyusunan kebijakan, perencanaan program promosi kesehatan, pengendalian penyakit menular, serta pengalokasian sumber daya medis secara lebih efektif dan efisien [3], [5].

Data mining merupakan proses eksplorasi dan analisis data untuk menemukan pola, hubungan, maupun informasi yang sebelumnya tidak diketahui [14], [19]. Salah satu teknik *data mining* yang banyak digunakan adalah *clustering*, yaitu metode pengelompokan data berdasarkan tingkat kemiripan karakteristik tertentu [14]. Algoritma K-Means merupakan salah satu metode *clustering* yang populer karena memiliki proses komputasi yang relatif sederhana, mudah diimplementasikan, dan mampu menghasilkan kelompok data yang representatif [6], [9]. Informasi hasil klasterisasi dapat digunakan sebagai dasar dalam memahami karakteristik data dan mendukung pengambilan keputusan berbasis data [9], [12].

Rumah Sakit Dr. Tadjuddin Chalid merupakan salah satu fasilitas pelayanan kesehatan di Kota Makassar yang telah menerapkan sistem rekam medis elektronik [16]. Namun, pemanfaatan data rekam medis yang tersedia masih berfokus pada kebutuhan pelayanan pasien secara individual. Padahal, data tersebut memiliki potensi untuk dianalisis lebih lanjut guna menghasilkan informasi strategis yang dapat mendukung proses pengambilan keputusan di tingkat manajemen rumah sakit [16]. Oleh karena itu, diperlukan suatu metode yang mampu mengidentifikasi pola tersembunyi dari data rekam medis dalam jumlah besar [6].

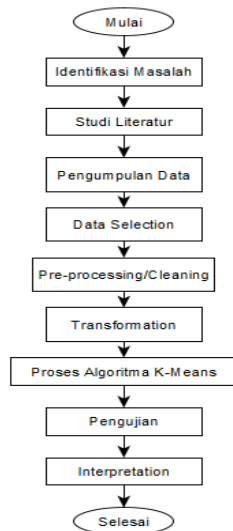
Beberapa penelitian terdahulu telah menerapkan teknik klasterisasi pada berbagai bidang untuk menemukan pola tersembunyi dari suatu dataset. Dilawati dkk. [6] menerapkan algoritma K-Means untuk melakukan pengelompokan data rekam medis pasien guna melihat pola tren kunjungan secara makro. Di sisi lain, untuk memahami karakteristik performa komputasi metode ini, Golalipour dkk. [8] melakukan tinjauan mendalam (*review*) mengenai perkembangan algoritma pembentukan klaster dari pendekatan partisi hingga metode ansambel.

Pada domain kesehatan lingkungan, Apriani dan Tedja Tresnaning [4] menerapkan *Fuzzy Adaptive Clustering* untuk mengetahui pengaruh faktor lingkungan hidup terhadap tingkat keluhan kesehatan masyarakat. Untuk mengevaluasi kualitas dan validitas model pembentukan kelompok tersebut, secara teoritis didukung oleh kajian Tomasini dkk. [14] yang membahas studi komparasi fungsional antara indeks validasi internal (seperti *Silhouette Index*) dan indeks eksternal pada algoritma klasterisasi. Selain itu, Yin dkk. [15] dalam penelitian terbarunya memberikan tinjauan cepat (*rapid review*) mengenai pemetaan perkembangan berbagai algoritma klasterisasi secara global.

Penelitian-penelitian tersebut menunjukkan bahwa metode klasterisasi mampu menghasilkan informasi yang sangat bermanfaat dalam ekstraksi pengetahuan dari data yang besar. Berdasarkan permasalahan tersebut, penelitian ini menerapkan algoritma K-Means untuk melakukan klasterisasi data rekam medis pasien rawat inap di Rumah Sakit Dr. Tadjuddin Chalid [16]. Selain menghasilkan kelompok pasien yang memiliki karakteristik serupa, penelitian ini juga mengembangkan *dashboard* berbasis web untuk memvisualisasikan hasil klasterisasi secara interaktif. Kontribusi penelitian ini terletak pada integrasi proses klasterisasi data rekam medis dengan *dashboard* visualisasi interaktif sehingga informasi yang dihasilkan tidak hanya mampu mengidentifikasi pola sebaran pasien, tetapi juga lebih mudah dipahami dan dimanfaatkan oleh pihak rumah sakit dalam mendukung pengambilan keputusan yang cepat, tepat, dan berbasis data [18].

2. METODE

Penelitian ini menerapkan pendekatan *Knowledge Discovery in Databases* (KDD) yang terdiri atas tahapan seleksi data, *preprocessing*, transformasi data, *data mining* menggunakan algoritma K-Means, dan evaluasi menggunakan *Silhouette Index* untuk mengolah data rekam medis pasien rawat inap menjadi informasi yang dapat mendukung pengambilan keputusan di Rumah Sakit Dr. Tadjuddin Chalid.



Gambar 1. Tahapan Penelitian

2.1 Dataset dan Preprocessing

Penelitian ini dilaksanakan di Rumah Sakit Dr. Tadjuddin Chalid pada periode November 2024 hingga April 2025. Data yang digunakan merupakan data rekam medis pasien rawat inap periode Januari 2024–Februari 2025 yang diperoleh dari sistem Rekam Medis Elektronik (RME) rumah sakit. *Dataset* awal terdiri atas 2.934 baris data dengan sepuluh variabel, yaitu nopen, nomor, tanggal masuk, kode, diagnosis, jenis kelamin, kelurahan, usia, jenis kunjungan, dan ruangan.

Dataset yang digunakan terdiri atas atribut diagnosis, kelurahan, usia, dan jenis kelamin pasien. Keempat atribut tersebut dipilih karena dianggap mampu merepresentasikan karakteristik pasien serta mendukung proses pengelompokan data berdasarkan tingkat kemiripan karakteristiknya.

Pada tahap *preprocessing*, dilakukan seleksi atribut untuk memilih variabel yang relevan dengan tujuan penelitian. Selanjutnya dilakukan proses pembersihan data (*data cleaning*) untuk menghilangkan data yang tidak lengkap (*missing value*), data duplikat, serta data yang berpotensi menimbulkan kesalahan pada proses analisis. Setelah proses *cleaning* diperoleh 1.848 data yang lengkap dan digunakan pada tahap klusterisasi. Hasil dari tahap ini menghasilkan dataset yang siap digunakan dalam proses klusterisasi.

Tabel 1. Atribut yang Digunakan dalam Penelitian

No	Atribut	Tipe Data
1	Diagnosis	Kategorikal
2	Jenis Kelamin	Kategorikal
3	Kelurahan	Kategorikal
4	Usia	Numerik

2.2 Transformasi Data

Algoritma K-Means memerlukan data dalam bentuk numerik sehingga atribut kategorikal perlu ditransformasikan terlebih dahulu. Pada penelitian ini, atribut diagnosis, jenis kelamin, dan kelurahan dikonversi ke dalam bentuk numerik menggunakan teknik *ordinal encoding*, sedangkan atribut usia digunakan dalam bentuk numerik tanpa proses transformasi tambahan.

Transformasi dilakukan agar seluruh atribut dapat diproses menggunakan perhitungan jarak pada algoritma K-Means. Selain itu, atribut kelurahan tidak diberikan kode secara acak, tetapi dikonversi berdasarkan urutan kedekatan wilayah terhadap lokasi rumah sakit sehingga nilai numerik yang lebih besar merepresentasikan wilayah yang relatif lebih jauh.

Variabel diagnosis dan jenis kelamin dikodekan berdasarkan frekuensi kemunculan data, sedangkan variabel kelurahan dikodekan berdasarkan urutan jarak terhadap rumah sakit. Pada penelitian ini tidak dilakukan proses normalisasi karena hasil transformasi tetap mempertahankan informasi urutan kedekatan wilayah pada atribut kelurahan yang digunakan sebagai karakteristik utama dalam analisis kluster. Tabel 2 menunjukkan contoh data hasil transformasi yang digunakan sebagai input proses klusterisasi.

Tabel 2. Contoh Hasil Transformasi Data Rekam Medis

Diagnosis	Kelurahan	Usia	Jenis Kelamin
1	241	1	1
1	1	8	0
2	22	2	1
2	0	5	0
2	7	71	0
0	1	5	1
2	1	24	0
1	85	8	1
1	7	89	0
2	0	22	1

2.3 Klasterisasi Menggunakan Algoritma K-Means

Proses pengelompokan data dilakukan menggunakan algoritma K-Means. Algoritma ini dipilih karena memiliki proses komputasi yang sederhana, mudah diimplementasikan, serta banyak digunakan dalam penelitian klasterisasi data kesehatan [7], [9].

K-Means bekerja dengan membagi data ke dalam sejumlah klaster berdasarkan tingkat kemiripan karakteristik data. Setiap data ditempatkan pada klaster yang memiliki pusat klaster (*centroid*) terdekat. Selanjutnya, nilai *centroid* diperbarui secara iteratif hingga diperoleh kondisi yang stabil dan tidak terjadi perubahan anggota klaster secara signifikan.

Pada penelitian ini, proses klasterisasi dilakukan menggunakan bahasa pemrograman Python pada lingkungan Google Colaboratory. Variabel yang digunakan dalam proses klasterisasi meliputi diagnosis, kelurahan, usia, dan jenis kelamin pasien.

2.4 Validasi Klaster

Untuk menentukan kualitas hasil klasterisasi, dilakukan validasi menggunakan metode *Silhouette Index* (SI). Metode ini digunakan untuk mengukur kualitas pengelompokan berdasarkan tingkat kedekatan data dalam satu klaster (*cohesion*) dan tingkat keterpisahan data antar klaster (*separation*) [16], [17].

Nilai *Silhouette Index* dihitung menggunakan Persamaan (1).

$$SI(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (1)$$

dengan:

- $a(i)$ adalah rata-rata jarak data ke seluruh anggota dalam klaster yang sama;
- $b(i)$ adalah rata-rata jarak minimum data terhadap anggota pada klaster lain.

Nilai SI berada pada rentang -1 hingga 1. Nilai yang semakin mendekati 1 menunjukkan bahwa objek berada pada klaster yang tepat dan memiliki kualitas pengelompokan yang baik. Sebaliknya, nilai yang mendekati 0 menunjukkan adanya tumpang tindih antar klaster, sedangkan nilai negatif mengindikasikan kemungkinan kesalahan penempatan objek pada klaster. Hasil perhitungan *Silhouette Index* digunakan sebagai dasar untuk mengevaluasi kualitas klaster yang terbentuk dan memastikan bahwa hasil pengelompokan memiliki tingkat pemisahan yang baik antar klaster.

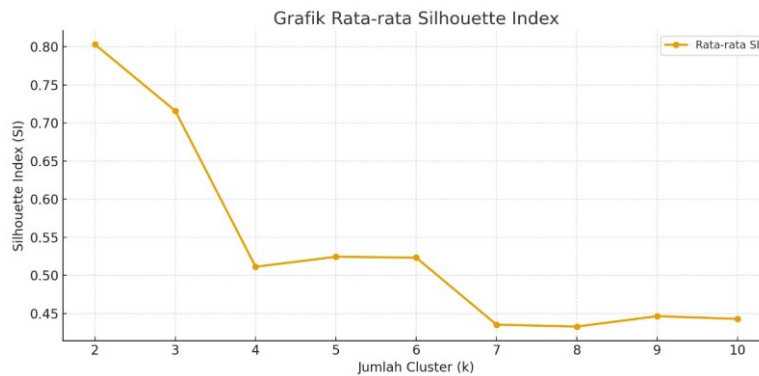
3. HASIL DAN PEMBAHASAN

3.1. Penentuan Jumlah Klaster Optimal

Penentuan jumlah klaster optimal dilakukan menggunakan metode *Silhouette Index* (SI). Pengujian dilakukan terhadap beberapa alternatif jumlah klaster, yaitu K=2 hingga K=10. Nilai *Silhouette Index* digunakan untuk mengevaluasi kualitas hasil klasterisasi berdasarkan tingkat kedekatan data dalam satu klaster (*cohesion*) dan tingkat keterpisahan antar klaster (*separation*). Semakin tinggi nilai SI yang diperoleh, semakin baik kualitas pengelompokan yang dihasilkan.

Tabel 3. Nilai *Silhouette Index* pada Berbagai Jumlah Klaster

Jumlah K	SI	Waktu Komputasi	Rata-rata Iterasi
2	0,8030	0,5656	5
3	0,7158	0,0179	8
4	0,5112	0,0174	11
5	0,5243	0,0183	12
6	0,5232	0,0156	15
7	0,4352	0,0185	21
8	0,4327	0,0221	30
9	0,4463	0,0335	18
10	0,4428	0,0335	29

Gambar 2. Grafik *Silhouette Index* Setiap Nilai K pada Data Rekam Medis

Berdasarkan Tabel 3 dan Gambar 2, nilai *Silhouette Index* tertinggi diperoleh pada K=2 dengan nilai sebesar 0,8030. Nilai tersebut mendekati nilai maksimum SI yaitu 1, yang menunjukkan bahwa hasil klasterisasi memiliki kualitas pengelompokan yang sangat baik. Sementara itu, nilai SI pada K=3 hingga K=10 mengalami penurunan yang cukup signifikan dibandingkan K=2.

Hasil tersebut menunjukkan bahwa pembagian data ke dalam dua klaster menghasilkan tingkat kemiripan yang tinggi antaranggota dalam klaster yang sama serta tingkat pemisahan yang baik terhadap klaster lainnya. Oleh karena itu, jumlah klaster optimal yang digunakan dalam penelitian ini adalah dua klaster.

Selain menghasilkan nilai *Silhouette Index* tertinggi, penggunaan dua klaster juga memberikan proses komputasi yang relatif efisien dengan rata-rata iterasi sebanyak lima kali. Dengan demikian, K=2 dipilih sebagai jumlah klaster terbaik untuk proses analisis data rekam medis pasien rawat inap di Rumah Sakit Dr. Tadjuddin Chalid.

3.2 Hasil Klasterisasi Data Rekam Medis

Proses klasterisasi dilakukan terhadap 1.848 data rekam medis pasien rawat inap menggunakan algoritma K-Means dengan jumlah klaster optimal sebanyak dua klaster berdasarkan hasil evaluasi *Silhouette Index*. Variabel yang digunakan dalam proses pengelompokan meliputi diagnosis, keluhan, usia, dan jenis kelamin pasien.

Tabel 4. Distribusi Data pada Setiap Klaster

Klaster	Jumlah Pasien
Klaster 1	1.632
Klaster 2	216
Total	1.848

Berdasarkan hasil klasterisasi, Klaster 1 terdiri atas 1.632 data pasien atau 88,56% dari total data, sedangkan Klaster 2 terdiri atas 216 data pasien atau 11,44% dari total data. Hasil tersebut menunjukkan bahwa sebagian besar pasien memiliki karakteristik yang terkonsentrasi pada satu kelompok utama.

Untuk mengetahui karakteristik masing-masing klaster, dilakukan analisis terhadap nilai *centroid* akhir yang dihasilkan oleh algoritma K-Means.

Tabel 5. Nilai *Centroid* Akhir Setiap Klaster

Klaster	Tabel <i>Centroid</i>			
	Diagnosis	Kelurahan	Usia	Jenis Kelamin
Klaster 1	0	8	26	1
Klaster 2	1	125	19	1

Berdasarkan Tabel 5, terdapat perbedaan nilai centroid pada setiap variabel yang digunakan. Variabel kelurahan menunjukkan perbedaan nilai yang paling mencolok antara kedua klaster dibandingkan variabel lainnya. Hasil ini menunjukkan bahwa lokasi tempat tinggal pasien menjadi faktor utama yang membedakan karakteristik kedua klaster yang terbentuk.

Perbedaan nilai *centroid* pada variabel kelurahan merupakan yang terbesar dibandingkan variabel lainnya sehingga menjadi faktor yang paling dominan dalam pembentukan klaster. Nilai *centroid* pada variabel diagnosis dan jenis kelamin merupakan nilai hasil *encoding* sehingga digunakan untuk menunjukkan kecenderungan kelompok, bukan untuk merepresentasikan kategori diagnosis atau jenis kelamin tertentu secara langsung.

Selain itu, perbedaan nilai *centroid* pada variabel usia menunjukkan adanya kecenderungan perbedaan kelompok umur antar klaster, meskipun pengaruhnya tidak sebesar variabel kelurahan. Sementara itu, variabel jenis kelamin memiliki nilai *centroid* yang relatif sama pada kedua klaster sehingga kontribusinya terhadap pembentukan klaster cenderung lebih kecil. Dengan demikian, hasil klasterisasi lebih merepresentasikan pola persebaran geografis pasien dibandingkan pola diagnosis atau karakteristik demografis lainnya.

3.3 Karakteristik dan Interpretasi Klaster

Untuk memahami pola yang terbentuk, dilakukan analisis terhadap distribusi pasien pada masing-masing klaster berdasarkan wilayah tempat tinggal.

Klaster 1

Klaster 1 merupakan kelompok dengan jumlah anggota terbesar, yaitu 1.632 pasien. Karakteristiknya adalah pasien didominasi oleh warga lokal pada wilayah dekat (Kelurahan Paccerakkang, Daya, Sudiang), dengan rentang usia dewasa hingga lansia, dan diagnosis penyakit terbanyak adalah *Typhoid Fever* (Tipes) dan *Dyspepsia* (Maag).

Tabel 6. Kelurahan Dominan pada Klaster 1

No	Kelurahan	Jumlah	%
1	Paccerakkang	263	15,73
2	Sudiang Raya	227	13,58
3	Sudiang	176	10,53
4	Laikang	113	6,76
5	Pai	103	6,16
6	Daya	96	5,74
7	Kapasa	80	4,78
8	Moncong Loe	76	4,55
9	Tamalanrea	70	4,19
10	Kelurahan lain	428	26,23
Total		1.632	100

Berdasarkan Tabel 6, pasien pada Klaster 1 didominasi oleh beberapa wilayah yang berada di sekitar rumah sakit, seperti Paccerakkang, Sudiang Raya, dan Sudiang. Tingginya jumlah pasien dari wilayah tersebut menunjukkan bahwa faktor kedekatan geografis masih menjadi pertimbangan utama masyarakat dalam memanfaatkan layanan rawat inap. Kondisi ini mengindikasikan bahwa aksesibilitas lokasi memiliki pengaruh yang signifikan terhadap pola kunjungan pasien.

Klaster 2

Klaster 2 terdiri atas 216 pasien dan merepresentasikan kelompok pasien yang berasal dari wilayah dengan jarak relatif lebih jauh dari rumah sakit. Karakteristiknya adalah pasien yang berasal dari luar wilayah Kota Makassar

(daerah *hinterland*/kabupaten tetangga), didominasi oleh pasien anak-anak/balita, dengan diagnosis medis utama terkait *Neonatal Jaundice* (bayi kuning) atau gangguan pernapasan akut.

Tabel 7. Kelurahan Dominan pada Klaster 2

No	Kelurahan	Jumlah	%
1	Turikale	4	1,85
2	Pabentengang	3	1,39
3	Kalabbirang	3	1,39
4	Balang Baru	3	1,39
5	Parang Tambung	2	0,93
6	Pangkajene	2	0,93
7	Tondo	2	0,93
8	Bontorannu	2	0,93
9	Samalewa	2	0,93
10	Mattoangin	2	0,93
11	Petobo	2	0,93
12	Kanaungan	2	0,93
13	Baruga	2	0,93
14	Balleanging	2	0,93
15	Kelurahan lain	158	73,15
Total		216	100

Distribusi pasien pada Klaster 2 cenderung lebih tersebar dibandingkan Klaster 1. Tidak terdapat satu wilayah yang sangat dominan sehingga menunjukkan bahwa pasien dalam klaster ini berasal dari berbagai daerah dengan karakteristik lokasi yang lebih beragam.

Temuan tersebut menunjukkan bahwa Rumah Sakit Dr. Tadjuddin Chalid tidak hanya melayani masyarakat di wilayah sekitar rumah sakit, tetapi juga menjadi tujuan pelayanan kesehatan bagi pasien yang berasal dari daerah yang lebih jauh.

3.4 Pembahasan

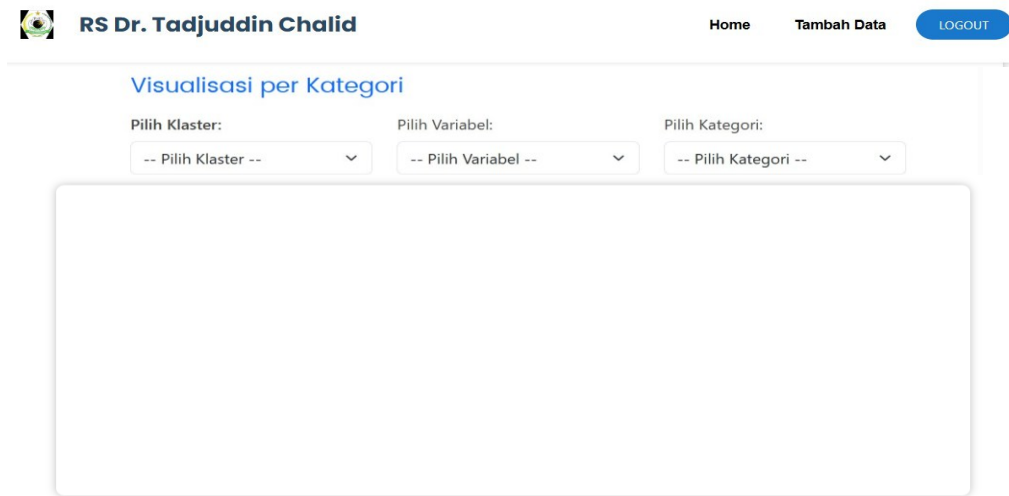
Hasil penelitian menunjukkan bahwa algoritma K-Means mampu mengidentifikasi pola tersembunyi pada data rekam medis pasien rawat inap dan mengelompokkannya ke dalam dua klaster utama dengan kualitas pengelompokan yang baik. Nilai *Silhouette Index* sebesar 0,8030 menunjukkan bahwa struktur klaster yang terbentuk memiliki tingkat pemisahan yang jelas dan tingkat kemiripan internal yang tinggi.

Dominasi variabel kelurahan terjadi karena rentang nilainya jauh lebih besar dibandingkan variabel diagnosis dan jenis kelamin sehingga memberikan kontribusi terbesar terhadap perhitungan *Euclidean Distance*. Analisis karakteristik klaster menunjukkan bahwa variabel kelurahan menjadi faktor yang paling berpengaruh dalam pembentukan kelompok pasien. Hasil ini mengindikasikan bahwa faktor geografis masih memiliki peran penting dalam pemanfaatan layanan kesehatan. Pasien yang berdomisili di wilayah sekitar rumah sakit cenderung membentuk kelompok terbesar, sedangkan pasien dari wilayah yang lebih jauh membentuk kelompok yang lebih kecil dan tersebar.

Temuan penelitian ini menunjukkan bahwa algoritma K-Means mampu secara akurat mengidentifikasi pola tersembunyi pada data rekam medis pasien rawat inap berdasarkan karakteristik wilayah dan diagnosis. Hasil klasterisasi geografis dan sebaran demografis ini diperkuat oleh konsep ilmu epidemiologi dan kesehatan masyarakat yang dipaparkan oleh Amirah dan Ahmaruddin [2] serta Hasmi [10], di mana distribusi penyakit selalu berkaitan erat dengan faktor tempat (*place*), karakteristik umur, dan aksesibilitas masyarakat ke fasilitas kesehatan utama [17]. Kelengkapan dan kualitas input rekam medis elektronik di Rumah Sakit Dr. Tadjuddin Chalid memegang peranan krusial untuk menghasilkan klasterisasi yang valid, sejalan dengan premis Amran dkk. [3] serta Ariani [5] yang menegaskan bahwa data rekam medis yang andal merupakan instrumen strategis dalam meningkatkan efisiensi operasional dan mutu manajemen pelayanan rumah sakit.

Lebih lanjut, keberadaan pelayanan Rumah Sakit Dr. Tadjuddin Chalid dalam melayani sebaran pasien lokal dari kelurahan terdekat hingga pasien luar daerah terekam secara historis dalam kajian Fatmawati [7] mengenai profil pelayanan rumah sakit tersebut terhadap penanganan kesehatan wilayah di Kecamatan Biringkanaya. Efektivitas penggunaan K-Means dalam pemetaan atribut institusional ini juga selaras dengan penerapan komputasi *data mining* oleh Handayani [9] serta Prastyo dkk. [12], yang membuktikan bahwa algoritma K-Means handal untuk mengelompokkan profil objek dalam jumlah besar guna mendukung perancangan strategi manajemen yang berbasis data (*data-driven decision making*).

Sebagai implementasi hasil penelitian, hasil klasterisasi divisualisasikan dalam bentuk dashboard berbasis web yang dikembangkan menggunakan bahasa pemrograman PHP. *Dashboard* menampilkan informasi distribusi pasien berdasarkan klaster, diagnosis, usia, jenis kelamin, dan lokasi tempat tinggal sehingga memudahkan pengguna dalam melakukan eksplorasi data secara dinamis.



Gambar 3. Tampilan *Dashboard* Klasterisasi Data Rekam Medis

Dashboard memungkinkan pengguna melakukan eksplorasi distribusi pasien berdasarkan diagnosis, usia, jenis kelamin, dan lokasi tempat tinggal melalui filter interaktif sehingga informasi dapat dianalisis secara lebih cepat dibandingkan laporan konvensional. Informasi tersebut dapat dimanfaatkan oleh pihak manajemen rumah sakit untuk memantau karakteristik pasien dan mendukung perencanaan layanan kesehatan.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penerapan teknik data mining menggunakan algoritma K-Means berhasil mengelompokkan 1.848 data rekam medis pasien rawat inap di Rumah Sakit Dr. Tadjuddin Chalid dengan sangat optimal. Pengujian validitas menggunakan metode *Silhouette Index* menghasilkan nilai tertinggi sebesar 0,8030 pada jumlah cluster $K=2$, yang menunjukkan bahwa struktur klaster yang terbentuk memiliki kualitas pengelompokan yang sangat kuat (*strong structure*).

Klaster 1 didominasi oleh pasien lokal (domisili dekat) dengan diagnosis penyakit seperti *typhoid fever* dan *dyspepsia*, sedangkan Klaster 2 mengelompokkan pasien dari luar wilayah Makassar yang didominasi usia anak-anak dengan diagnosis *neonatal jaundice*. Selain itu, penelitian ini telah berhasil mengembangkan sebuah *dashboard* berbasis web interaktif yang berfungsi efektif untuk memvisualisasikan hasil klasterisasi tersebut, sehingga dapat digunakan oleh pihak manajemen rumah sakit sebagai sistem pendukung keputusan strategis dalam optimalisasi pelayanan kesehatan. Untuk penelitian selanjutnya, disarankan menambahkan variabel klinis yang lebih beragam serta membandingkan hasil dengan metode klasterisasi lainnya.

REFERENSI

- [1] M. Amin, W. Setyonugroho, dan N. Hidayah, "Implementasi Rekam Medik Elektronik: Sebuah Studi Kualitatif," *Jurnal Manajemen Informasi Kesehatan Indonesia (JMIKI)*, vol. 9, no. 1, pp. 45–52, 2021.
- [2] A. Amirah dan S. Ahmaruddin, *Konsep dan Aplikasi Epidemiologi*. Yogyakarta: Deepublish, 2020.
- [3] R. Amran, A. Apriyani, dan N. P. Dewi, "Peran Penting Kelengkapan Rekam Medik di Rumah Sakit," *Baiturrahmah Medical Journal*, vol. 1, no. 2, pp. 69–76, 2022.
- [4] A. Apriani dan P. Tedja Tresnaning, "Fuzzy Adaptive Clustering Untuk Mengetahui Pengaruh Faktor Lingkungan Hidup Terhadap Tingkat Keluhan Kesehatan," *Kurvatek*, vol. 7, no. 2, pp. 35–40, 2022, doi: 10.33579/krvtk.v7i2.3769.
- [5] S. Ariani, "Analisis Keberhasilan Implementasi Rekam Medik Elektronik Dalam Meningkatkan Efisiensi Dan Mutu Pelayanan," *Jurnal Kesehatan dan Kedokteran*, vol. 2, no. 2, pp. 7–14, 2023, doi: 10.56127/jukeke.v2i2.720.
- [6] H. Dilawati, H. Widiyanto, dan A. Kuswiadji, "Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means," *Jurnal Teknologi dan Sistem Informasi*, vol. 5, no. 2, pp. 5–8, 2024.
- [7] Fatmawati, "Sistem Pelayanan Rumah Sakit Dr. Tadjuddin Chalid Terhadap Penyandang Kusta di Kecamatan Biringkanaya Kota Makassar," Skripsi, Jurusan Kesehatan Masyarakat, UIN Alauddin, Makassar, 2018.
- [8] K. Golalipour, E. Akbari, S. S. Hamidi, M. Lee, dan R. Enayatifar, "From clustering to clustering ensemble selection: A review," *Engineering Applications of Artificial Intelligence*, vol. 104, p. 104388, 2021, doi: 10.1016/j.engappai.2021.104388.
- [9] F. Handayani, "Aplikasi Data Mining Menggunakan Algoritma K-Means Clustering untuk Mengelompokkan Mahasiswa Berdasarkan Gaya Belajar," *Jurnal Teknologi dan Informasi*, vol. 12, no. 1, pp. 46–63, 2022, doi: 10.34010/jati.v12i1.6733.
- [10] Hasmi, *Metode Penelitian Epidemiologi*, A. Maftuhin, Ed. Jakarta: CV Trans Info Media, 2021.

- [11] Kementerian Kesehatan Republik Indonesia, Peraturan Menteri Kesehatan Republik Indonesia Nomor 24 Tahun 2022 tentang Rekam Medis. Jakarta: Kementerian Kesehatan RI, 2022.
- [12] P. H. Prastyo, Z. Saharuna, dan I. Abduh, "Analisis Data Atribut Mahasiswa untuk Menentukan Strategi Promosi Kampus Menggunakan Metode Data Mining," dalam Seminar Nasional Teknik Elektro dan Informatika (SNTEI), 2017, pp. 81–86.
- [13] R. F. Putra et al., Data Mining: Algoritma dan Penerapannya, Efitra dan Sepriano, Ed. Padang: PT Sonpedia Publishing Indonesia, 2023.
- [14] C. Tomasini, E. N. Borges, K. Machado, dan L. Emmendorfer, "A study on the relationship between internal and external validity indices applied to partitioning and density-based clustering algorithms," dalam Proceedings of the 19th International Conference on Enterprise Information Systems (ICEIS), 2017, vol. 1, pp. 89–98, doi: 10.5220/0006317000890098.
- [15] H. Yin, A. Aryani, S. Petrie, A. Nambissan, A. Astudillo, dan S. Cao, "A Rapid Review of Clustering Algorithms," arXiv preprint arXiv:2401.07389, 2024.
- [16] K. T. Yudita, H. H. Abbas, dan S. Suyuti, "Analisis Sistem Informasi Manajemen Rumah Sakit (SIMRS) pada Bagian Rekam Medis RSUP Dr. Tadjuddin Chalid Makassar," Window of Public Health Journal, vol. 3, no. 1, pp. 112–121, 2022.
- [17] Surahman dan S. Supardi, Pengantar Ilmu Kesehatan Masyarakat PKM. Jakarta: Kementerian Kesehatan RI, 2016.
- [18] I. B. Agung, A. Rioland, A. D. Dewantara, dan R. Nur, "Pengembangan Sistem Data Warehouse dan Web API Untuk Dukungan Pengambilan Keputusan Produksi Buah di Sulawesi Selatan," Journal of Informatics and Computer Engineering (JICER), vol. 3, no. 1, pp. 15–22, 2025.
- [19] M. R. Fajri, "Implementasi Data Mining Untuk Menentukan Strategi Pemasaran Produk Menggunakan Algoritma K-Means Clustering," Jurnal Teknologi dan Open Source, vol. 4, no. 1, pp. 45–52, 2021.
- [20] N. I. Sari, M. F. Raharjo, dan A. D. Dewantara, "Aplikasi Manajemen Dinas Pariwisata, Pemuda, dan Olahraga Kabupaten Pangkajene Kepulauan," Journal of Informatics and Computer Engineering Research (JICER), vol. 2, no. 1, pp. 9–17, 2025.