

## Analisis Clustering Data Penyakit Pasien Berbasis Algoritma K-Means Pada RSUD Majene

SitiAulia Rachmini\*<sup>1</sup>, Irfan AP<sup>2</sup>, Milasari<sup>3</sup>, Gusna Mayang Sari<sup>4</sup>

<sup>1,2,3,4</sup>Program Studi Teknik Informatika, Universitas Sulawesi Barat

E-mail: \*<sup>1</sup>[sitiaulia.rachmini@unsulbar.ac.id](mailto:sitiaulia.rachmini@unsulbar.ac.id),

<sup>2</sup>[irfan\\_ap@unsulbar.ac.id](mailto:irfan_ap@unsulbar.ac.id), <sup>3</sup>[milasari@unsulbar.ac.id](mailto:milasari@unsulbar.ac.id), <sup>4</sup>[gusnamayangsari@gmail.com](mailto:gusnamayangsari@gmail.com)

### Abstrak

Penelitian ini didasari oleh permasalahan yang dihadapi oleh RSUD Majene terkait peningkatan jumlah data pasien di RSUD Majene yang belum dimanfaatkan secara optimal untuk menghasilkan informasi strategis, sehingga menyulitkan pihak rumah sakit dalam mengidentifikasi pola penyakit serta penentuan prioritas pelayanan kesehatan. Oleh karena itu, diperlukan suatu pendekatan berbasis data mining untuk mengelompokkan data penyakit pasien secara efektif. Penelitian ini menggunakan metode K-Means Clustering dengan data sebanyak 1.417 pasien tahun 2023 yang mencakup variabel diagnosis penyakit, usia, dan jenis kelamin. Tahapan penelitian meliputi data cleaning, transformasi data, serta proses clustering. Penentuan jumlah cluster optimal dilakukan menggunakan metode Elbow dengan nilai SSE pada  $k=3$  sebesar 6.835.033,99 yang menunjukkan hasil paling optimal. Hasil penelitian menghasilkan tiga cluster utama, yaitu cluster pertama didominasi oleh penyakit pulpitis irreversibel dan nekrosis pulpa pada usia 5–24 tahun, cluster kedua didominasi oleh periodontitis kronis dan pericoronitis pada usia 25–64 tahun, serta cluster ketiga mencakup kasus penyakit dengan tingkat kompleksitas tinggi pada usia lanjut ( $\geq 65$  tahun). Hasil ini menunjukkan bahwa metode K-Means mampu mengidentifikasi pola distribusi penyakit pasien secara efektif. Dengan demikian, penelitian ini penting sebagai dasar pengambilan keputusan dalam peningkatan kualitas layanan kesehatan dan perencanaan strategi pencegahan penyakit secara lebih tepat sasaran.

**Kata kunci**— Data Mining, K-Means Clustering, Pengelompokan Data, Pola Penyakit, Pelayanan Kesehatan

### Abstract

This study is based on the problems faced by RSUD Majene related to the increasing volume of patient data that has not been optimally utilized to generate strategic information, making it difficult for the hospital to identify disease patterns and determine healthcare service priorities. Therefore, a data mining-based approach is needed to effectively cluster patient disease data. This study applies the K-Means Clustering method using 1,417 patient records from 2023, which include variables such as disease diagnosis, age, and gender. The research stages consist of data cleaning, data transformation, and clustering processes. The optimal number of clusters was determined using the Elbow method, with the best result at  $k=3$  and a Sum of Squared Error (SSE) value of 6,835,033.99. The results produce three main clusters: the first cluster is

*dominated by diseases such as irreversible pulpitis and pulp necrosis in the 5–24 age group; the second cluster is dominated by chronic periodontitis and pericoronitis in the 25–64 age group; and the third cluster represents cases with higher disease complexity in the elderly group ( $\geq 65$  years). These results indicate that the K-Means method effectively identifies patterns in the distribution of patients' diseases. Therefore, this study is important as a basis for decision-making in improving the quality of healthcare services and developing more targeted disease prevention strategies.*

**Keywords**—Data Mining, K-Means Clustering, Data Clustering, Disease Patterns, Healthcare Services

## 1. PENDAHULUAN

Perkembangan teknologi informasi telah mendorong pemanfaatan data dalam berbagai bidang, termasuk sektor kesehatan. Rumah sakit sebagai institusi pelayanan kesehatan menghasilkan data dalam jumlah besar setiap harinya, terutama data rekam medis pasien yang mencakup diagnosis penyakit, usia, dan jenis kelamin. Data rekam medis elektronik memiliki potensi besar dalam mendukung pengambilan keputusan klinis dan analisis kesehatan berbasis data [1]. Namun, dalam praktiknya, data tersebut umumnya masih digunakan untuk keperluan administratif dan pelaporan sehingga potensi data untuk menghasilkan informasi strategis belum dimanfaatkan secara optimal [2]. Selain itu, kompleksitas dan volume data kesehatan yang terus meningkat menyebabkan perlunya penerapan teknik analisis data untuk menghasilkan informasi yang lebih bermakna dan mendukung pengambilan keputusan di bidang kesehatan [3].

Kondisi ini juga terjadi di RSUD Majene, di mana peningkatan jumlah data pasien belum diiringi dengan analisis data yang memadai untuk mengidentifikasi pola penyakit serta mendukung pengambilan keputusan [4]. Akibatnya, pihak rumah sakit mengalami kesulitan dalam menentukan prioritas pelayanan kesehatan dan strategi pencegahan penyakit secara tepat [5]. Permasalahan serupa juga ditemukan pada berbagai fasilitas pelayanan kesehatan lainnya, di mana data rekam medis masih lebih banyak dimanfaatkan untuk kebutuhan administrasi dibandingkan sebagai sumber informasi strategis dalam pengambilan keputusan kesehatan [6]. Selain itu, kurang optimalnya pemanfaatan sistem informasi rumah sakit menyebabkan proses analisis data dan penentuan strategi pelayanan kesehatan menjadi kurang efektif [7]. Penelitian lain juga menyebutkan bahwa pemanfaatan data mining pada data rekam medis dapat membantu rumah sakit dalam mengidentifikasi pola penyakit pasien dan mendukung proses pengambilan keputusan secara lebih akurat dan sistematis [8].

Kondisi ini juga terjadi di RSUD Majene, di mana peningkatan jumlah data pasien belum diiringi dengan analisis data yang memadai untuk mengidentifikasi pola penyakit serta pendukung pengambilan keputusan. Akibatnya, pihak rumah sakit mengalami kesulitan dalam menentukan prioritas pelayanan kesehatan dan strategi pencegahan penyakit secara tepat [4][5]. Salah satu pendekatan yang dapat digunakan untuk mengatasi permasalahan tersebut adalah data mining. Data mining merupakan proses ekstraksi pengetahuan atau pola tersembunyi dari kumpulan data dalam jumlah besar yang dapat digunakan untuk mendukung pengambilan keputusan [9]. Salah satu teknik dalam data mining adalah clustering, yaitu metode pengelompokan data berdasarkan tingkat kemiripan karakteristik tertentu tanpa adanya label sebelumnya [10]. Metode K-Means Clustering merupakan algoritma clustering yang banyak digunakan karena kesederhanaan dan efisiensinya dalam mengelompokkan data ke dalam beberapa cluster berdasarkan jarak terdekat terhadap centroid. Dengan penerapan metode ini, data pasien dapat dikelompokkan sehingga pola distribusi penyakit dapat diidentifikasi secara lebih sistematis [10].

Sejumlah penelitian terdahulu menunjukkan bahwa metode K-Means Clustering efektif

dalam pengelompokan data kesehatan. Penelitian lain menunjukkan bahwa algoritma K-Means mampu mengelompokkan data secara efisien dan menghasilkan pola yang jelas pada data berdimensi besar [11]. Penerapan K-Means pada data rekam medis juga dapat membantu dalam mengidentifikasi pola penyakit dan meningkatkan proses pengambilan keputusan di bidang kesehatan [10]. Selanjutnya, peneliti lain melakukan kajian bahwa metode K-Means mampu mengelompokkan data pasien berdasarkan karakteristik tertentu sehingga memudahkan analisis distribusi penyakit [6]. Penelitian yang lain juga menyimpulkan bahwa teknik clustering, khususnya K-Means, efektif dalam mengekstraksi pola dari data dalam jumlah besar yang dapat digunakan sebagai dasar pengambilan keputusan di sektor kesehatan [13].

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk menerapkan metode K-Means Clustering dalam mengelompokkan data penyakit pasien di RSUD Majene berdasarkan variabel diagnosis penyakit, usia, dan jenis kelamin. Hasil dari penelitian ini diharapkan dapat memberikan informasi mengenai pola distribusi penyakit yang lebih akurat dan bermanfaat sebagai dasar dalam pengambilan keputusan, peningkatan kualitas pelayanan kesehatan, serta penyusunan strategi pencegahan penyakit yang lebih efektif dan tepat sasaran.

## 2. METODE

### 2.1. Analisis Masalah

RSUD Majene sebagai institusi pelayanan kesehatan menghasilkan data pasien dalam jumlah besar setiap tahunnya, bahkan jumlah kunjungan pasien dapat mencapai puluhan ribu per tahun. Berdasarkan data Badan Pusat Statistik Kabupaten Majene, jumlah kunjungan pasien ke rumah sakit umum daerah pada tahun 2023 mencapai sekitar 55.347 kunjungan, yang terdiri dari 45.822 pasien rawat jalan dan 9.525 pasien rawat inap [14]. Selain itu, penelitian lain juga menunjukkan bahwa tingkat kunjungan pasien di RSUD Majene mencapai sekitar 78.000 kunjungan pada tahun 2023, yang mengalami peningkatan dibandingkan dengan tahun sebelumnya. Namun, dalam praktiknya, data tersebut masih disimpan dalam bentuk cetak dan belum sepenuhnya terkomputerisasi, sehingga memerlukan proses konversi sebelum dapat dianalisis lebih lanjut [5].

Permasalahan utama yang dihadapi adalah belum tersedianya sistem yang mampu mengolah dan mengelompokkan data pasien secara otomatis untuk menemukan pola penyakit berdasarkan karakteristik tertentu seperti diagnosis penyakit, usia, dan jenis kelamin [15]. Selain itu, keterbatasan dalam analisis data menyebabkan pihak rumah sakit kesulitan dalam mengidentifikasi penyakit dominan, kelompok usia rentan, serta dalam menentukan prioritas pelayanan kesehatan dan strategi pencegahan penyakit secara tepat [16].

Untuk mengatasi permasalahan tersebut, diperlukan suatu pendekatan berbasis data mining yang mampu mengekstraksi pola tersembunyi dari data dalam jumlah besar. Teknik clustering, khususnya metode K-Means Clustering, dipilih karena kemampuannya dalam mengelompokkan data berdasarkan kemiripan karakteristik secara efisien serta banyak digunakan dalam analisis data kesehatan [13]. Dengan penerapan metode ini, diharapkan dapat dihasilkan informasi berupa pola distribusi penyakit yang dapat mendukung pengambilan keputusan secara lebih efektif.

### 2.2. Desain Sistem

Desain sistem dalam penelitian ini dibangun berdasarkan kerangka sistem yang terdiri dari tiga komponen utama, yaitu input, proses, dan output. Secara umum, sistem dirancang untuk mengolah data pasien menjadi informasi berupa hasil pengelompokan penyakit menggunakan metode K-Means Clustering. Kerangka sistem dalam penelitian ini dapat dilihat pada Gambar 1.

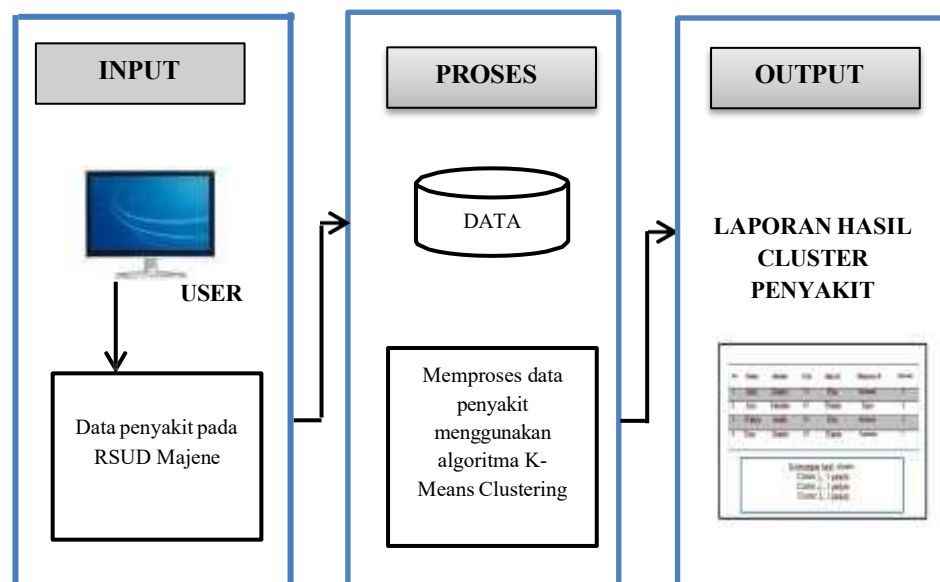
#### 2.2.1. Diagram Pemrosesan Data

Secara konseptual, alur pemrosesan data dalam penelitian ini mengikuti tahapan data mining (KDD), yaitu dimulai dari data mentah hingga menghasilkan informasi yang bermakna. Data input berupa data pasien RSUD Majene tahun 2023 yang terdiri dari variabel diagnosis

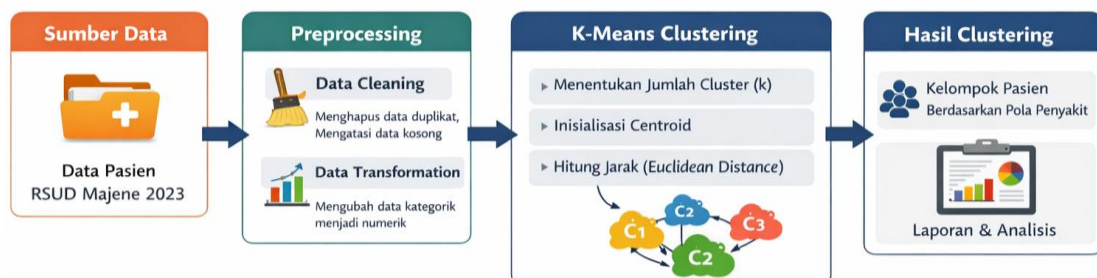
penyakit, usia, dan jenis kelamin. Selanjutnya, data tersebut melalui tahap preprocessing yang meliputi data cleaning dan data transformation.

Pada tahap data cleaning, dilakukan proses pemeriksaan terhadap data untuk memastikan tidak terdapat duplikasi, inkonsistensi, maupun data yang tidak lengkap. Berdasarkan hasil penelitian, seluruh data yang digunakan telah memenuhi kriteria sehingga tidak ada data yang dihapus. Tahap berikutnya adalah transformasi data, di mana data kategorik seperti jenis kelamin dan diagnosis penyakit diubah ke dalam bentuk numerik agar dapat diproses oleh algoritma K-Means.

Setelah preprocessing, data diproses menggunakan algoritma K-Means Clustering. Proses ini dimulai dengan penentuan jumlah cluster ( $k$ ), dimana penentuan jumlah cluster optimum dilakukan menggunakan metode elbow. Selanjutnya dilakukan inisialisasi centroid, perhitungan jarak menggunakan Euclidean Distance, serta pengelompokan data ke dalam cluster berdasarkan jarak terdekat. Proses ini dilakukan secara iteratif hingga mencapai kondisi konvergen, yaitu ketika tidak terjadi perubahan pada anggota cluster. Hasil akhir dari proses ini berupa kelompok data pasien yang memiliki karakteristik serupa. Diagram pemrosesan data dapat dilihat pada Gambar 2.



Gambar 1 Kerangka Sistem



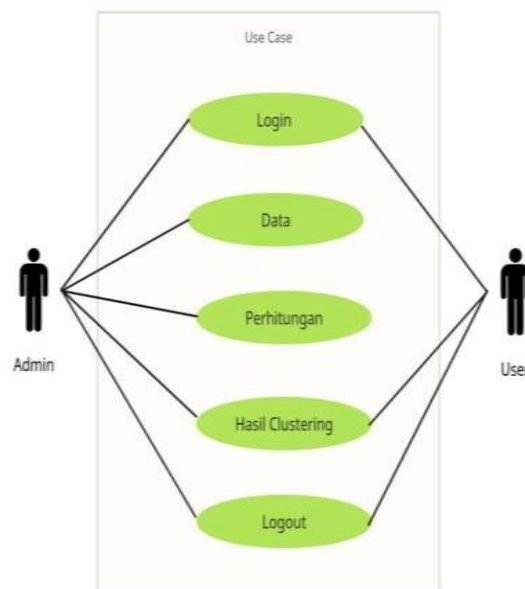
Gambar 2 Diagram Pemrosesan Data

### 2.2.2. Diagram Arsitektur Sistem

Arsitektur sistem dalam penelitian ini berbasis aplikasi web yang terdiri dari beberapa komponen utama, yaitu pengguna (user), sistem aplikasi, basis data, dan modul pemrosesan.

Pengguna sistem terdiri dari dua jenis, yaitu petugas rekam medis dan kepala rekam medis, yang memiliki hak akses berbeda dalam mengoperasikan sistem. *Use case diagram* pada sistem ini dapat dilihat pada Gambar 3. Admin pada sistem adalah petugas rekam medis dan user dalam sistem adalah kepala rekam medis.

Sistem aplikasi berfungsi sebagai antarmuka yang dibangun menggunakan bahasa pemrograman PHP dan dijalankan pada server lokal menggunakan XAMPP. Basis data menggunakan MySQL yang berfungsi untuk menyimpan data pasien serta hasil pengolahan clustering. Modul pemrosesan merupakan bagian inti sistem yang mengimplementasikan algoritma K-Means Clustering untuk mengelompokkan data secara otomatis. Hasil dari proses ini ditampilkan dalam bentuk laporan clustering yang dapat diakses oleh pengguna untuk keperluan analisis dan pengambilan keputusan.



Gambar 3 *Use Case Diagram*

#### 2.2.3. Diagram Alur Sistem

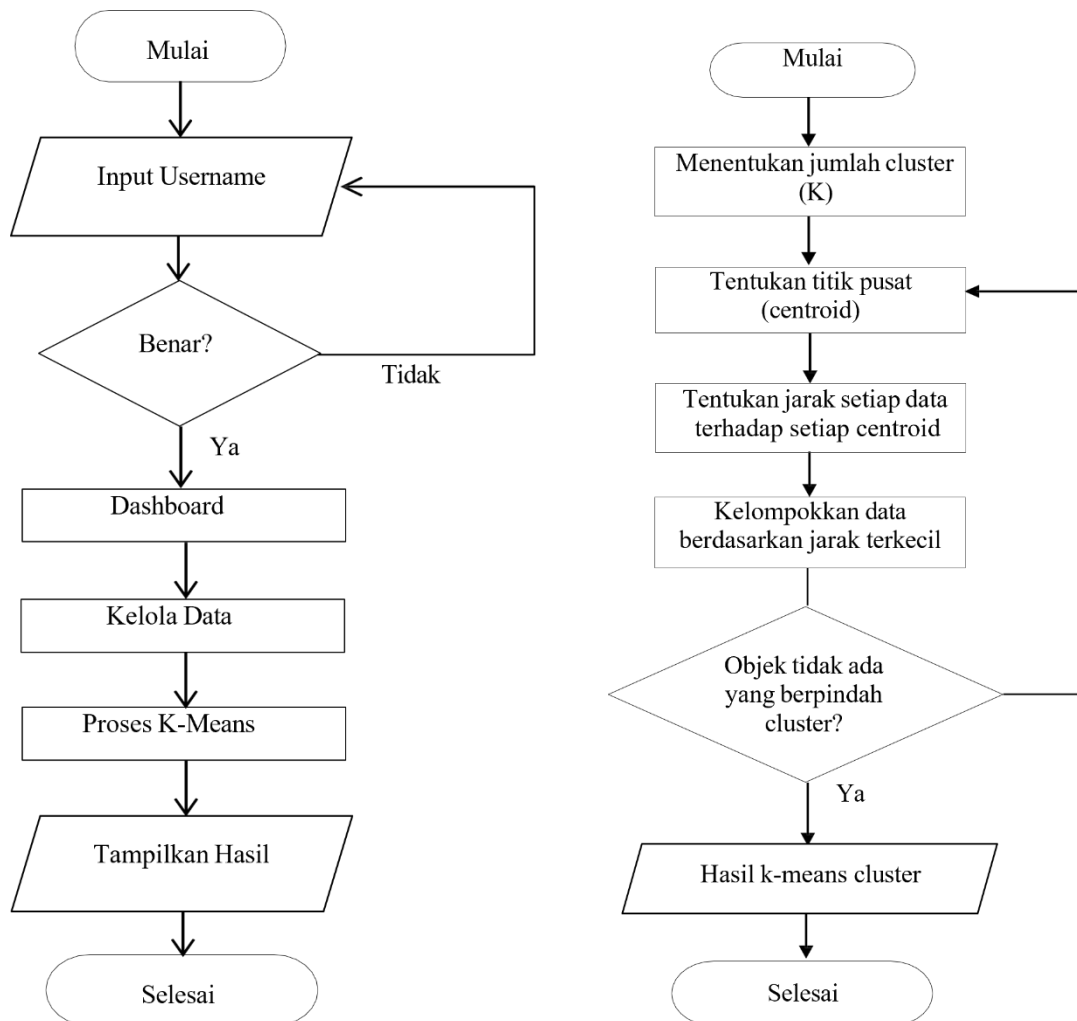
Flowchart sistem menggambarkan alur kerja sistem secara keseluruhan, dimulai dari proses login pengguna, pengelolaan data pasien, hingga proses perhitungan clustering dan penampilan hasil sesuai dengan yang ada pada Gambar 4 (a). Setelah pengguna berhasil login, sistem akan menampilkan dashboard yang berisi menu utama, seperti menu data pasien, proses perhitungan, dan hasil clustering. Pada menu data pasien, pengguna dapat menambahkan atau mengelola data yang kemudian akan diproses menggunakan algoritma K-Means.

#### 2.2.4. Diagram Algoritma K-Means

Proses K-Means Clustering dalam penelitian ini dilakukan melalui beberapa tahapan yang terstruktur untuk menghasilkan pengelompokan data yang optimal. Tahap awal dimulai dengan menentukan jumlah cluster ( $k$ ) yang paling sesuai menggunakan metode Elbow, yaitu dengan menganalisis nilai *Sum of Squared Error* (SSE) untuk setiap kemungkinan jumlah cluster dan memilih titik yang menunjukkan penurunan signifikan [17]. Setelah nilai  $k$  ditentukan, langkah berikutnya adalah menginisialisasi centroid awal secara acak sebagai pusat cluster. Selanjutnya, dilakukan perhitungan jarak antara setiap data dengan centroid menggunakan metode Euclidean Distance untuk menentukan kedekatan antar data [12].

Berdasarkan hasil perhitungan jarak tersebut, setiap data akan dikelompokkan ke dalam cluster yang memiliki centroid terdekat. Setelah proses pengelompokan awal, dilakukan perhitungan ulang centroid dengan cara mengambil nilai rata-rata dari seluruh data dalam

masing-masing cluster. Proses ini kemudian diulang secara iteratif, yaitu dengan menghitung kembali jarak data ke centroid yang baru dan memperbarui keanggotaan klaster, hingga mencapai kondisi konvergen. Kondisi konvergen ditandai dengan tidak adanya perubahan signifikan pada anggota klaster atau nilai centroid, sehingga diperoleh hasil pengelompokan yang stabil dan representatif terhadap struktur data [6]. Diagram algoritma K-Means dapat dilihat pada Gambar 4(a).



(a) Flow Chart Sistem

(b) Flow Chart K-Means

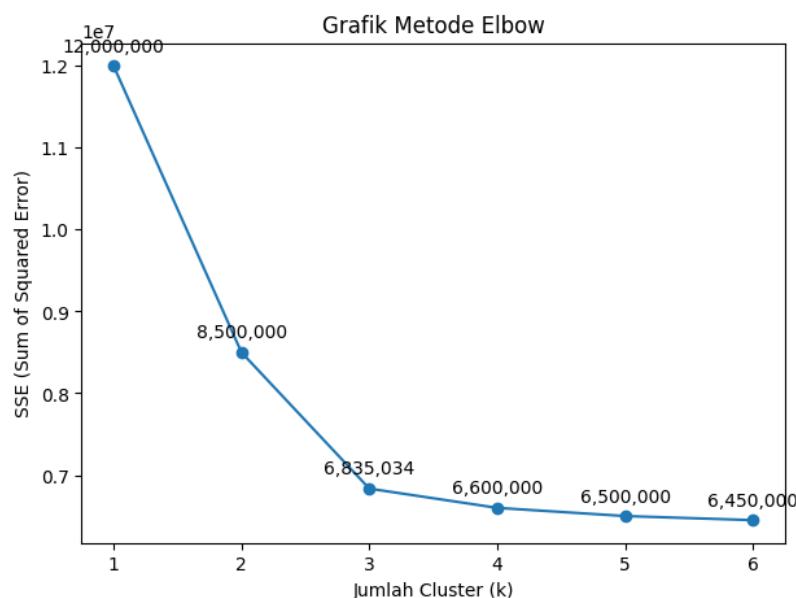
Gambar 4 Flow Chart Sistem Clustering Data penyakit Pasien

### 3. HASIL DAN PEMBAHASAN

Hasil penelitian ini diperoleh dari proses pengolahan data pasien RSUD Majene tahun 2023 sebanyak 1.417 data menggunakan metode K-Means Clustering. Tahapan awal penelitian dilakukan melalui proses preprocessing yang meliputi data cleaning dan transformasi data. Pada tahap *cleaning* data, dilakukan proses pemeriksaan terhadap data untuk memastikan tidak terdapat duplikasi, inkonsistensi, maupun data yang tidak lengkap. Berdasarkan hasil pemeriksaan, seluruh data dinyatakan valid sehingga tidak ada data yang dihapus dan seluruh data sebanyak 1.417 entri digunakan dalam proses clustering. Pada tahap transformasi data, data yang masih berbentuk kategori diubah ke dalam bentuk numerik agar dapat diproses menggunakan algoritma K-Means.

Variabel yang digunakan dalam penelitian ini meliputi diagnosis penyakit, usia, dan jenis kelamin. Transformasi dilakukan dengan memberikan kode numerik pada setiap kategori variabel. Diagnosis penyakit terdiri atas delapan kategori, yaitu pulpitis irreversible, nekrosis pulpa, impaksi, pericoronitis, akar gigi tertinggal, kelainan bentuk mahkota, supernumerary teeth, dan periodontitis kronis. Transformasi usia dibagi ke dalam tujuh kelompok usia, sedangkan jenis kelamin diklasifikasikan menjadi laki-laki dan perempuan. Tabel transformasi data diagnosis penyakit, usia, serta jenis kelamin dapat dilihat masing-masing pada Tabel 1, 2, dan 3.

Setelah proses transformasi dan *cleaning* data selesai dilakukan, tahap berikutnya adalah menentukan jumlah cluster optimal menggunakan metode Elbow. Metode ini dilakukan dengan menghitung nilai *Sum of Squared Error* (SSE) pada beberapa nilai cluster. Hasil pengujian menunjukkan bahwa nilai SSE mengalami penurunan signifikan hingga  $k = 3$  dengan nilai sebesar 6.835.033,99, sehingga jumlah cluster optimal yang digunakan dalam penelitian ini adalah tiga cluster. Visualisasi hasil metode Elbow menunjukkan bahwa penurunan nilai SSE mulai melandai setelah  $k = 3$ , sehingga nilai tersebut dipilih sebagai jumlah cluster terbaik dalam penelitian ini. Grafik Elbow pada penelitian ini dapat dilihat pada Gambar 4.



Gambar 4 Grafik Metode Elbow

Tabel 1 Transformasi Data Diagnosis Penyakit

| No | Nama Penyakit           | Kode |
|----|-------------------------|------|
| 1  | Pulpitis irreversible   | 1    |
| 2  | Nekrosis pulpa          | 2    |
| 3  | Impaksi                 | 3    |
| 4  | Pericoronitis           | 4    |
| 5  | Akar gigi tertinggal    | 5    |
| 6  | Kelainan bentuk mahkota | 6    |
| 7  | Supernumerary teeth     | 7    |
| 8  | Periodontitis kronis    | 8    |

Tabel 2 Transformasi Data Usia

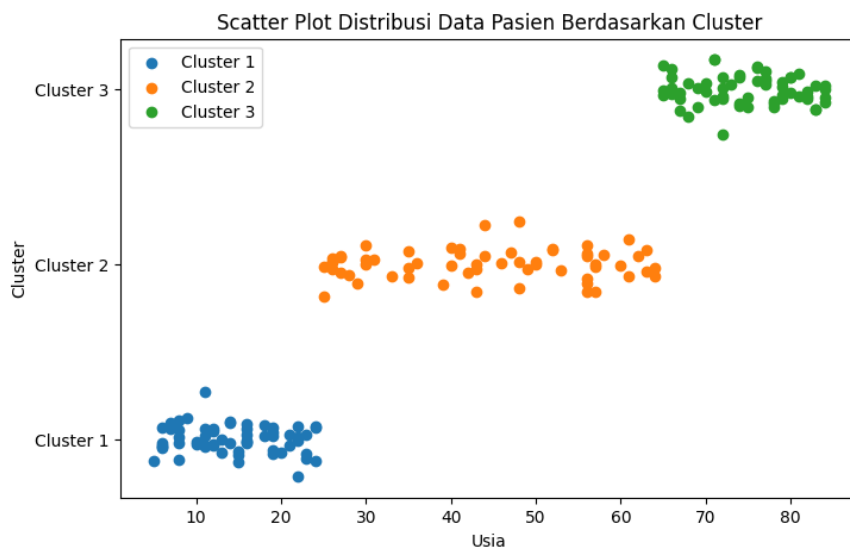
| No | Usia      | Kode |
|----|-----------|------|
| 1  | < 1 tahun | 1    |

| No | Usia        | Kode |
|----|-------------|------|
| 2  | 1–4 tahun   | 2    |
| 3  | 5–14 tahun  | 3    |
| 4  | 15–24 tahun | 4    |
| 5  | 25–44 tahun | 5    |
| 6  | 45–64 tahun | 6    |
| 7  | ≥65 tahun   | 7    |

Tabel 3 Transformasi Data Jenis Kelamin

| No | Jenis Kelamin | Kode |
|----|---------------|------|
| 1  | Laki-laki     | 0    |
| 2  | Perempuan     | 1    |

Berdasarkan hasil proses clustering menggunakan algoritma K-Means, diperoleh tiga kelompok utama dengan karakteristik yang berbeda. Cluster pertama didominasi oleh pasien dengan diagnosis pulpitis irreversibel dan nekrosis pulpa pada rentang usia 5–24 tahun. Cluster kedua didominasi oleh pasien dengan diagnosis periodontitis kronis dan pericoronitis pada rentang usia 25–64 tahun. Sementara itu, cluster ketiga terdiri atas pasien dengan tingkat kompleksitas penyakit yang lebih tinggi yang umumnya berada pada kelompok usia lanjut, yaitu  $\geq 65$  tahun. Berdasarkan karakteristik data, cluster ini cenderung mencakup kasus penyakit seperti impaksi, akar gigi tertinggal, kelainan bentuk mahkota, dan supernumerary teeth yang memerlukan penanganan medis lebih lanjut dibandingkan penyakit dasar lainnya. Kondisi ini menunjukkan bahwa peningkatan usia berpengaruh terhadap meningkatnya kompleksitas permasalahan kesehatan gigi dan mulut pasien. Hasil clustering direpresentasikan dalam bentuk grafik distribusi cluster sesuai pada Gambar 5 untuk memperjelas persebaran data pasien pada masing-masing kelompok. Grafik menunjukkan bahwa cluster kedua memiliki jumlah data paling banyak dibandingkan dengan cluster lainnya, yang menunjukkan bahwa kelompok usia dewasa merupakan kelompok dengan tingkat kasus penyakit tertinggi di RSUD Majene.



Gambar 5 Grafik Sebaran Hasil Clustering

Hasil penelitian ini menunjukkan bahwa metode K-Means Clustering mampu mengelompokkan data pasien berdasarkan tingkat kemiripan karakteristik secara efektif. Nilai SSE yang relatif

rendah pada  $k = 3$  menunjukkan bahwa data pasien yang berada dalam satu cluster memiliki karakteristik yang cukup homogen, baik berdasarkan diagnosis penyakit, usia, maupun jenis kelamin. Penyakit pada kelompok usia muda cenderung didominasi oleh gangguan kesehatan gigi dasar, sedangkan kelompok usia dewasa hingga lanjut menunjukkan kecenderungan penyakit dengan tingkat kompleksitas yang lebih tinggi. Berdasarkan hasil clustering yang dihasilkan, persentase penyakit pada masing-masing cluster dapat dilihat pada Tabel 4. Hasil tersebut menunjukkan bahwa sebagian besar kasus penyakit pasien di RSUD Majene berada pada kelompok usia produktif atau dewasa. Selain itu, distribusi data pada setiap cluster memperlihatkan bahwa metode K-Means mampu memisahkan data pasien berdasarkan karakteristik penyakit dan usia secara cukup jelas dan terstruktur.

Tabel 4 Persentasi Penyakit Setiap Cluster

| Cluster   | Usia Karakteristik | Penyakit Dominan                                                            | Presentase |
|-----------|--------------------|-----------------------------------------------------------------------------|------------|
| Cluster 1 | 5–24 tahun         | Pulpitis irreversible,<br>Nekrosis pulpa                                    | 36,13%     |
| Cluster 2 | 25-54 tahun        | Periodontitis kronis,<br>Pericoronitis                                      | 45,38%     |
| Cluster 3 | $\geq 65$ tahun    | Impaksi, Akar gigi tertinggal, Kelainan bentuk mahkota, Supernumerary teeth | 18,49%     |

#### 4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa metode K-Means Clustering berhasil diterapkan untuk mengelompokkan data penyakit pasien di RSUD Majene tahun 2023 sebanyak 1.417 data pasien. Proses penelitian dilakukan melalui tahapan preprocessing data yang meliputi data cleaning dan transformasi data, kemudian dilanjutkan dengan proses clustering menggunakan algoritma K-Means. Penentuan jumlah cluster optimal dilakukan menggunakan metode Elbow dan diperoleh jumlah cluster terbaik sebanyak tiga cluster dengan nilai *Sum of Squared Error* (SSE) sebesar 6.835.033,99.

Hasil pengelompokan menunjukkan bahwa Cluster 1 didominasi oleh penyakit pulpitis irreversible dan nekrosis pulpa pada kelompok usia 5–24 tahun. Cluster 2 didominasi oleh penyakit periodontitis kronis dan pericoronitis pada kelompok usia 25–64 tahun, sedangkan Cluster 3 mencakup penyakit dengan tingkat kompleksitas yang lebih tinggi pada kelompok usia lanjut ( $\geq 65$  tahun). Distribusi data menunjukkan bahwa Cluster 2 memiliki jumlah pasien terbanyak, yaitu 643 pasien atau 45,38% dari total data.

Secara keseluruhan, hasil penelitian membuktikan bahwa metode K-Means Clustering mampu mengidentifikasi pola distribusi penyakit pasien berdasarkan karakteristik usia, jenis kelamin, dan diagnosis penyakit secara efektif. Informasi yang dihasilkan dapat dimanfaatkan oleh pihak RSUD Majene sebagai dasar dalam mendukung pengambilan keputusan, menentukan prioritas pelayanan kesehatan, serta menyusun strategi pencegahan penyakit yang lebih tepat sasaran guna meningkatkan kualitas layanan kesehatan.

#### REFERENSI

- [1] Nasarudin, N. A., Al Jasmi, F., Sinnott, R. O., Zaki, N., Al Ashwal, H., Mohamed, E. A., & Mohamad, M. S. (2024). A review of deep learning models and online healthcare databases for electronic health records and their use for health prediction. *Artificial Intelligence Review*, 57(249), 1–35. <https://doi.org/10.1007/s10462-024-10876-2>
- [2] Suyoko, S., Agiwahyunto, F., Ernawati, D., Prasetya, J., & Abiyasa, M. T. (2024). Analisis data rekam medis elektronik di rumah sakit (Studi perancangan dan evaluasi sistem informasi analisis kuantitatif). *J-REMI: Jurnal Rekam Medik dan Informasi Kesehatan*, 5(3), 219–233. <https://doi.org/10.25047/j-remi.v5i3.4666>
- [3] Kolling, M. L., Furstenau, L. B., Sott, M. K., Rabaioli, B., Ulmi, P. H., Bragazzi, N. L., & Tedesco, L. P. C. (2021). Data mining in healthcare: Applying strategic intelligence techniques to depict 25 years of research development. *International Journal of Environmental Research and Public Health*, 18(6), 3099. <https://doi.org/10.3390/ijerph18063099>
- [4] Dinas Kesehatan Kabupaten Majene. (2023). *Profil Kesehatan Kabupaten Majene Tahun 2023*. Dinas Kesehatan Kabupaten Majene.
- [5] Djalaluddin, N. A., Masniati, M., Hidayat, L., & Alim, M. C. (2024). *Pengaruh perceived value dan social culture terhadap pengambilan keputusan pasien pada layanan kesehatan di RSUD Majene*. *Journal of Public Health Science (JoPHS)*, 1(4). <https://journal.ppmi.web.id/index.php/jophs/article/view/1702>
- [6] Ardiansyah, A. A., Khaeroshi, E. A., & Wicaksono, R. R. (2025). Klasterisasi data rekam medis pasien menggunakan metode K-Means Clustering di RSUD Muhammadiyah Bantul. *Jurnal Teknik Informatika UNIS*, 13(1). <https://doi.org/10.33592/jutis.v13i1.6776>
- [7] Ordila, R., Wahyuni, R., Irawan, Y., & Sari, M. Y. (2020). Penerapan data mining untuk pengelompokan data rekam medis pasien berdasarkan jenis penyakit dengan algoritma clustering. *Jurnal Ilmu Komputer*, 9(2), 148–153. <https://doi.org/10.33060/JIK/2020/Vol9.Iss2.181>
- [8] Nanda, W. S., Pardede, A. M. H., & Simanjuntak, M. (2022). Analisis data mining untuk klasterisasi data rekam medis menggunakan algoritma K-Means pada Rumah Sakit Sylvani Binjai. *Indonesian Journal of Education and Computer Science*, 1(2), 45–53. <https://doi.org/10.60076/indotech.v1i2.43>
- [9] Kolling, M. L., Furstenau, L. B., Sott, M. K., Rabaioli, B., Ulmi, P. H., Bragazzi, N. L., & Tedesco, L. P. C. (2021). Data mining in healthcare: Applying strategic intelligence techniques to depict 25 years of research development. *International Journal of Environmental Research and Public Health*, 18(6), 3099. <https://doi.org/10.3390/ijerph18063099>
- [10] Kaur, M., Sharma, A., & Singh, P. (2021). A comprehensive study of clustering techniques in data mining. *International Journal of Advanced Science and Technology*, 29(3), 1123–1131.
- [11] Chiu, Y. C., Chen, Y. H., & Tsai, C. F. (2023). Clustering techniques and applications in healthcare data analytics: A systematic review. *Healthcare Analytics*, 3, 100165. <https://doi.org/10.1016/j.health.2023.100165>
- [12] Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., & Lin, C. T. (2021). A review of clustering techniques and developments. *Neurocomputing*, 267, 664–681. <https://doi.org/10.1016/j.neucom.2017.06.053>
- [13] Haraty, R. A., Dimishkieh, M., & Masud, M. (2015). An enhanced k-means clustering algorithm for pattern discovery in healthcare data. *International Journal of Distributed Sensor Networks*, 11(6), 615740. <https://doi.org/10.1155/2015/615740>
- [14] Badan Pusat Statistik Kabupaten Majene. (2023). *Jumlah kunjungan pasien ke rumah sakit umum daerah menurut bulan dan status perawatan di Kabupaten Majene tahun 2023*. <https://majenekab.bps.go.id/id/statistics-table/2/MTI0IzI%3D/number-of-patient-visits-to-regional-general-hospitals-by-month-and-treatment-status-in-majene-district.html>
- [15] Fajri, M. B., & Purnamasari, S. D. (2022). Klasterisasi pola penyebaran penyakit pasien berdasarkan usia pasien menggunakan K-Means Clustering. *Journal of Information*

- Technology Ampera*, 3(3), 317–334.  
<https://doi.org/10.51519/journalita.volume3.issue3.year2022.page317-334>
- [16] Laksono, B., Syahidin, Y., & Yunengsih, Y. (2023). Implementasi data mining klasterisasi data pasien rawat inap dengan algoritma K-Means Clustering. *Jurnal Teknologi Sistem Informasi dan Aplikasi*, 7(2). <https://doi.org/10.32493/jtsi.v7i2.39354>
- [17] Maori, N.A., Evanita. (2023). Metode Elbow Dalam Optimasi Jumlah Cluster Pada K-Means Clustering. *Jurnal SIMETRIS*. Vol 14 No 2.