

BAB 1

PENDAHULUAN

1.1 Latar Belakang

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

Namun, pada kenyataannya banyak institusi kesehatan yang belum mampu memanfaatkan data tersebut secara optimal. Data yang tersimpan seringkali hanya digunakan sebagai arsip administratif tanpa dilakukan analisis lebih lanjut untuk menggali informasi penting yang terkandung di dalamnya [2]. Hal ini menyebabkan potensi besar dari data medis tidak dimanfaatkan secara maksimal, terutama dalam memahami pola kunjungan pasien dan karakteristik penyakit.

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

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

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

Salah satu algoritma *clustering* yang paling banyak digunakan adalah K-Means. Algoritma ini bekerja dengan membagi data ke dalam sejumlah cluster berdasarkan jarak terdekat terhadap titik pusat (*centroid*) yang telah ditentukan. K-Means memiliki keunggulan dalam hal kemudahan implementasi, kecepatan proses,

serta kemampuan dalam mengolah data dalam jumlah besar secara efisien. Beberapa penelitian sebelumnya menunjukkan bahwa algoritma K-Means dapat diterapkan secara efektif dalam bidang kesehatan. Penelitian yang dilakukan oleh peneliti sebelumnya menunjukkan bahwa K-Means mampu mengelompokkan data rekam medis pasien berdasarkan pola penyakit sehingga dapat membantu dalam analisis distribusi penyakit. Selain itu, penelitian lain juga menunjukkan bahwa metode ini dapat digunakan untuk mengidentifikasi kelompok pasien dengan karakteristik tertentu yang dapat digunakan sebagai dasar dalam perencanaan layanan kesehatan [6].

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

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

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

1.2 Rumusan Masalah

Adapun rumusan masalah dalam penelitian ini adalah:

1. Bagaimana menerapkan algoritma K-Means dalam segmentasi pasien?
2. Bagaimana mengelompokkan pasien berdasarkan riwayat kunjungan dan diagnosis?
3. Bagaimana hasil *clustering* dapat digunakan untuk analisis pola pasien?

1.3 Tujuan dan Manfaat

Tujuan dari penelitian ini adalah menerapkan metode *K-Means Clustering* untuk melakukan segmentasi pasien berdasarkan riwayat kunjungan dan diagnosis penyakit, sehingga dapat digunakan sebagai dasar dalam memahami pola pasien serta membantu dalam pengambilan keputusan di bidang pelayanan kesehatan.

Manfaat dari penelitian ini yaitu:

1. Untuk mengelompokkan pasien berdasarkan karakteristik riwayat kunjungan dan diagnosis penyakit sehingga dapat diketahui pola kesamaan antar pasien.
2. Sebagai pendukung pengambilan keputusan dalam meningkatkan kualitas pelayanan kesehatan berdasarkan hasil segmentasi pasien.
3. Sebagai bahan analisis bagi pihak institusi kesehatan dalam menentukan strategi pelayanan yang lebih efektif dan efisien.
4. Membantu dalam mengidentifikasi kelompok pasien dengan tingkat kunjungan tinggi maupun rendah sehingga dapat dilakukan penanganan yang lebih tepat.
5. Sebagai referensi bagi penelitian selanjutnya yang berkaitan dengan penerapan data mining khususnya metode *clustering* dalam bidang kesehatan.

1.4 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. *Dataset* yang digunakan dalam penelitian ini diperoleh dari platform Kaggle dan bukan merupakan data langsung dari rumah sakit.
2. Data yang digunakan terbatas pada atribut yang berkaitan dengan riwayat kunjungan pasien dan diagnosis penyakit.
3. Metode yang digunakan dalam penelitian ini adalah algoritma *K-Means Clustering* sebagai teknik pengelompokan data.
4. Proses analisis hanya berfokus pada segmentasi atau pengelompokan pasien dan tidak membahas prediksi atau klasifikasi penyakit.
5. Data yang digunakan akan melalui tahap *preprocessing* seperti *cleaning data*, transformasi data, dan normalisasi sebelum dilakukan proses *clustering*.
6. *Tools* yang digunakan dalam penelitian ini adalah bahasa pemrograman PHP dan menggunakan software uji KNIME.

1.5 Keterbaharuan

Penelitian berjudul “Pengelompokkan Penyakit Pasien Menggunakan Algoritma K-Means” yang terbit di JURIKOM tahun 2022 menunjukkan bahwa K-Means dapat dipakai untuk mengelompokkan penyakit pasien pada layanan kesehatan, sehingga data diagnosis bisa dipetakan ke dalam kelompok yang lebih mudah dianalisis [8].

Penelitian “Penerapan Algoritma K-Means Untuk Klasterisasi Pasien Berdasarkan Riwayat Kesehatan dan Jenis Layanan Kesehatan” di INCODING tahun 2022 lebih dekat dengan topik kamu karena tidak hanya memakai satu atribut, tetapi

menggabungkan riwayat kesehatan dan jenis layanan kesehatan sebagai dasar segmentasi pasien [9].

Penelitian “Implementasi Algoritma *K-Means Clustering* Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Pulo Brayan” tahun 2023 memperlihatkan bahwa data penyakit di layanan primer dapat diproses dengan K-Means untuk mengetahui penyakit yang paling sering muncul pada pasien [10].

Penelitian “Implementasi Data Mining Menggunakan Algoritma *K-Means Clustering* Penyakit Pasien Rawat Jalan Pada Klinik Dr. Atirah Desa Sioyong, Sulteng” tahun 2023 menunjukkan bahwa data pasien rawat jalan juga dapat dikelompokkan berdasarkan pola penyakit menggunakan K-Means. [11]

Penelitian “Klasterisasi Pasien Rawat Jalan di Puskesmas dengan Menggunakan Metode Algoritma *Clustering K-Means*” tahun 2024 membuktikan bahwa pola kunjungan pasien di puskesmas dapat diolah menjadi beberapa cluster yang membantu melihat karakteristik pasien layanan primer [12].

Penelitian “Klasterisasi Data Rekam Medis Pasien Menggunakan Metode *K-Means Clustering* Di Rumah Sakit Widodo Ngawi” tahun 2024 menunjukkan bahwa rekam medis rumah sakit dapat diolah menjadi cluster yang bermanfaat untuk membaca pola data pasien secara lebih terstruktur [13].

Penelitian “Klasterisasi Pasien pada RSUD Ciamis Menggunakan Metode K-Means” tahun 2024 mengidentifikasi klaster penyakit dari data rawat inap dan menekankan bahwa *preprocessing* seperti seleksi, pembersihan, dan transformasi data sangat berpengaruh pada kualitas cluster [14].

Penelitian “Penerapan Metode *K-Means Clustering* Dalam Pengelompokan Data Pasien Rawat Inap Peserta BPJS Di Rumah Sakit Umum Daerah Kabupaten Kaur” tahun 2024 menghasilkan pengelompokan pasien berdasarkan umur, jenis kelamin, kelas BPJS, dan penyakit, sehingga menunjukkan bahwa atribut klinis dan administratif bisa digabung dalam satu model *clustering* [15].

Penelitian “Penerapan Metode K-Means Untuk Mengelompokkan Rekam Medis Pasien Berdasarkan Diagnosa Penyakit Guna Menentukan Diagnosa Tertinggi Pada Suatu Periode (Study Kasus: Klinik Dokter Kita)” tahun 2024 menemukan tiga diagnosis dominan, yaitu gastritis, ISPA, dan myalgia, sehingga K-Means terbukti dapat mengungkap penyakit yang paling menonjol dalam periode tertentu [16].

Penelitian “Implementasi Algoritma K-Means Untuk *Clustering* Data Penyakit Di Puskesmas Kotagede 2 Yogyakarta” tahun 2024 memperlihatkan bahwa data penyakit di puskesmas bisa dipetakan menjadi cluster yang mendukung analisis pelayanan kesehatan [17].

Penelitian “Pengelompokan Data Rekam Medis Pada Pasien Penyakit Dalam Untuk Meningkatkan Manajemen Informasi Kesehatan Berdasarkan Wilayah Kota Binjai Menggunakan Algoritma *Clustering K-Means*” tahun 2024 menambahkan unsur wilayah, usia, dan jenis penyakit sehingga hasil cluster tidak hanya klinis tetapi juga membantu manajemen informasi kesehatan rumah sakit [18].

Penelitian “Penerapan *K-Means Clustering* untuk Pengelompokan Pasien Rumah Sakit berdasarkan Tingkat Keparahan Penyakit” tahun 2025 mengelompokkan 995 data pasien ke dalam kategori ringan, sedang, dan parah berdasarkan atribut klinis seperti usia, jenis kelamin, triase awal, tanda vital, dan hasil laboratorium [19].

Penelitian “Klasterisasi Pasien Rawat Inap BPJS pada RS Islam Assyifa Sukabumi menggunakan Metode *K-Means*” tahun 2025 membuktikan bahwa data pasien BPJS rawat inap dapat dikelompokkan untuk membantu analisis beban layanan dan karakteristik pasien [20].

Penelitian “Data Mining untuk Pola Perbaikan Pelayanan Pasien dengan Metode Klustering *K-Means* pada Rumah Sakit Mitra Sejati” tahun 2025 menggabungkan data pasien rawat inap dan rawat jalan untuk mendukung penyusunan strategi pelayanan yang lebih optimal [21].

Penelitian “Klasterisasi Data Rekam Medis Pasien Menggunakan Metode *K-Means Clustering* Di RSUD Muhammadiyah Bantul” tahun 2025 menggunakan atribut jenis kelamin, usia, alamat, diagnosis, durasi rawat inap, dan metode pembayaran, sehingga menghasilkan cluster yang lebih komprehensif [22].