SKRIPSI

IMPLEMENTASI ALGORITMA *K-MEANS CLUSTERING* UNTUK PENGELOMPOKKAN DATA REKAM MEDIS RAWAT JALAN PADA PKU MUHAMMADIYAH GRABAG

AMANDHA SHAFA NABILA
NPM. 20.0504.0004

PROGRAM STUDI TEKNIK INFORMATIKA S1
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH MAGELANG
AGUSTUS, 2024

BAB 1
PENDAHULUAN

1.1 Latar Belakang Permasalahan

Di era globalisasi ini, teknologi informasi telah mengalami perkembangan yang sangat besar dan mempengaruhi hampir seluruh aspek kehidupan masyarakat. Teknologi kini mencakup berbagai sektor baik di ranah swasta maupun pemerintahan, seperti komunikasi, pendidikan, hiburan, ekonomi, kesehatan, dan lainnya. Salah satu penerapan penting dari teknologi informasi adalah dalam pengembangan sistem informasi. Penggunaan sistem informasi diartikan sebagai pemanfaatan teknologi informasi dan perangkat lunak untuk mengotomatisasi, mengintegrasikan, serta meningkatkan berbagai proses dalam suatu organisasi. Tujuannya adalah untuk meningkatkan efisiensi, kualitas, dan mengoptimalkan pengambilan keputusan (Prasetyo *et al.*, 2023).

Kemajuan teknologi informasi telah mengubah cara kita hidup dan bekerja dalam banyak hal, termasuk di bidang kesehatan yang sangat dipengaruhi oleh teknologi ini. Seiring dengan perkembangan tersebut, penggunaan teknologi informasi di rumah sakit menjadi sangat penting dalam pengelolaan data dan informasi pasien. Teknologi informasi mempermudah akses dan penggunaan data pasien secara cepat, yang salah satunya adalah untuk pengelolaan rekam medis. Selain itu, meski teknologi informasi mempermudah komunikasi dan pertukaran informasi, kemajuannya juga menimbulkan tantangan dalam hal pengelolaan data dan informasi pasien yang sensitif dan bersifat pribadi. Oleh karena itu, penting untuk memastikan keamanan dan kerahasiaan data pasien dalam penggunaan teknologi kesehatan. Pada akhirnya rekam medis berbasis teknologi dapat membantu dalam pengembangan pedoman klinis karena dokter dan tenaga kesehatan dapat mengakses data pasien lebih mudah (Sarce Joel *et al*., 2023).

Pencatatan rekam medis adalah hal wajib bagi dokter-dokter rumah sakit yang melakukan tindakan medis kepada pasien sesuai dengan aturan sehingga tidak ada alasan bagi dokter untuk tidak membuat rekam medik tersebut. Begitu juga dengan pelayanan medis. Pelayanan medis menjadi salah satu hal yang harus ditingkatkan. Dengan meningkatnya pelayanan medis maka ilmu rekam medis

sangat penting untuk dipelajari, sedangkan banyak diantara masyarakat masih belum memahami tentang rekam medis. Rekam medis merupakan berkas yang berisi data pasien, data pasien pada rumah sakit mengandung data yang cukup besar karena setiap pasien berkunjung atau mengecek kesehatan pada rumah sakit akan di jadikan rekam medis yang membuat data tersebut menumpuk yang menyebabkan berbagai masalah yang sering terjadi seperti tempat penyimpanan yang besar, hilangnya rekam medis, pengeluaran data yang dibutuhkan, dan lain-lain (Putri[1] *et al.*, n.d.).

Salah satu Puskesmas yang ada di Kabupaten Magelang adalah Pusat Kesehatan Umat Muhammadiyah Grabag. Pusat Kesehatan Umat (PKU) Muhammadiyah Grabag menerima pasien sebanyak 8 pasien s/d 10 pasien dalam satu hari, dalam satu bulan bisa mencapai 100 pasien s/d 251 pasien, sedangkan dalam satu tahun Pusat Kesehatan Umat (PKU) Muhammadiyah Grabag bisa menerima pasien kurang lebih sekitar 1.704 pasien dalam satu tahun. Pada tahun 2024 data pasien yang yang peneliti dapatkan ada sekitar 1.704 pasien. Oleh sebab itu, pengelolaan data rekam medis wajib dilakukan dengan baik serta efektif dan salah satu teknologi data yang bisa digunakan untuk pengelolaan data rekam medis merupakan dengan memanfaatkan *data mining.*

Data mining menggambarkan suatu proses yang memakai cara statistik, matematika, kecerdasan buatan, serta mesin pembelajaran yang mengekstraksi serta mengenali data yang berguna dan pengetahuan yang terkait dari berbagai data yang sangat besar. Salah satu guna data mining merupakan *k-means clustering*. Dalam konteks ini, penerapan *data mining* dengan menggunakan metode *k-means clustering* dapat menjadi solusi yang efektif untuk mengelola data rekam medis di rumah sakit. Metode *k-means clustering* menggambarkan proses pengelompokan titik-titik informasi ke dalam 2 kelompok ataupun lebih sehingga titik-titik informasi yang tercantum di dalam kelompok yang sama lebih mirip satu sama lain (Surya Nanda *et al*., n.d.)

Dengan demikian, penerapan data *mining* dengan memanfaatkan metode *k-means clustering* dalam mengelola data rekam medis rawat jalan pada PKU Muhammadiyah Grabag mempunyai relevansi yang besar dalam menaikkan efisiensi, mutu pelayanan, serta penelitian kedokteran di industri kesehatan dikala ini.

Berdasarkan latar belakang diatas yang maka diterapkan penggunaan data mining dengan menggunakan metode *k-means clustering* dengan memanfaatkan data rekam medis rawat jalan pada PKU Muhammadiyah Grabag, tujuan dari clustering data rekam medis rawat jalan di PKU Muhammadiyah Grabag berdasarkan variabel unit pelayanan dan jenis kelamin adalah untuk mengelompokkan pasien berdasarkan karakteristik tertentu, sehingga dapat mengidentifikasi pola penggunaan layanan kesehatan dan distribusi pasien antar unit pelayanan. Dengan demikian, hasil *clustering* ini diharapkan dapat membantu manajemen puskesmas dalam meningkatkan efisiensi pelayanan, serta mengoptimalkan sumber daya rumah sakit sesuai dengan kebutuhan pasien.

1.2 Rumusan Masalah

Berdasarkan latar belakang permasalahan di atas, maka masalah penelitian ini dirumuskan sebagai berikut :

Bagaimana metode *k-means clustering* diterapkan untuk mengelompokkan data rekam medis rawat jalan pada PKU Muhammadiyah Grabag?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah, maka tujuan penelitian yang akan dicapai adalah :

Untuk mengevaluasi efektivitas metode *k-means clustering* dalam mengelola data rekam medis dengan jumlah besar dan kompleks di Puskemas PKU Muhammadiyah Grabag.

1.4 Manfaat Penelitian

Berdasarkan tujuan penelitian yang telah disebutkan di atas, maka hasil penelitian ini dapat memberikan manfaat sebagai berikut:

Penerapan teknologi informasi dalam pengelolaan rekam medis di PKU Muhammadiyah Grabag dapat meningkatkan efisiensi operasional dengan

mempermudah akses data pasien, mempercepat proses diagnosis dan perawatan, serta memungkinkan dokter membuat diagnosis yang lebih akurat dan konsisten. Selain itu, metode data *mining* seperti *k-means clustering* dapat membantu mengelola dan menganalisis data rekam medis yang besar dan kompleks, sehingga mengurangi risiko kehilangan data dan memudahkan pencarian informasi. penelitian ini dapat berkontribusi pada pengembangan ilmu kedokteran dengan memberikan wawasan tentang pemanfaatan teknologi informasi dan metode data mining dalam pengelolaan rekam medis, serta menjadi referensi bagi penelitian selanjutnya dan bahan ajar dalam pendidikan kedokteran mengenai pentingnya pengelolaan data rekam medis yang efisien dan akurat.

BAB 2
TINJAUAN PUSTAKA

2.1 Penelitian Relevan

Berikut ini adalah beberapa hasil penelitian sebelumnya yang relevan atau berhubungan dengan penelitian yang akan dilakukan oleh peneliti.:

Menurut (Ordila *et al.,* 2020), dengan judul "Penerapan Data *Mining* untuk Pengelompokkan Data Rekam Medis Pasien Berdasarkan Jenis Penyakit dengan Algoritma *Clustering*" menyatakan bahwa metode data mining algoritma *k-means* efektif dalam mengelompokkan data rekam medis pasien polo klinik PT. Inecda berdasarkan wilayah, jenis kelamin, dan usia. Pasien terbanyak adalah dewasa dengan jumlah 4.912 pasien, diikuti oleh pasien anak-anak dengan jumlah 1.262 pasien, dan balita ada 144 pasien. Penyakit terbanyak adalah ISPA dengan jumlah pasien 1.985 pasien dikarenakan lingkungan perumahan PT.Inecda merupakan perkebunan kelapa sawit, dan ada juga pasien terbanyak dengan penyakit lain-lain seperti jatuh dari motor, cek kolesterol, kontrol kehamilan, cek tensi dan lain lain sebanyak 2141 pasien.

Sehingga sistem pengelolaan data sangat dibutuhkan dalam pengelompokkan data rekam medis rawat jalan di PKU Muhammadiyah Grabag, karena sistem ini dirancang untuk mengatasi masalah pengelompokan data pasien. Perberdaan dari sistem ini menggunakan metode *k-means clustering* untuk menyederhanakan dan mempercepat proses pengelompokan data dengan bantuan teknik *data mining*. Metode *Elbow* dan *Silhoutte* digunakan untuk menentukan jumlah cluster yang optimal. Hasil yang diperoleh dari penelitian ini adalah terciptanya sebuah sistem pengelompokan data rekam medis berdasarkan variabel jenis kelamin dan unit pelayanan, yang dapat memberikan wawasan lebih dalam untuk analisis kesehatan pasien. Namun, sistem ini hanya dirancang untuk digunakan oleh unit pelayanan terkait, sehingga bagian lain di rumah sakit tidak dapat mengakses data hasil pengelompokkan ini.

Selanjutnya menurut (Ma'rifatin, 2020) yang berjudul " Implementasi Algoritma *K-Means* untuk Pengelompokkan Penyakit Pasien pada Puskesmas Warujayeng" yaitu tentang implementasi algortima *k-means* untuk pengelompokan penyakit pasien pada puskesmas warujayeng. Kesimpulannya adalah bahwa

pengelompokan status kesehatan suatu wilayah masih dilakukan dengan secara manual menggunakan teknik statistik dasar, yang menghasilkan tidak optimal untuk menimbulkan masalah dengan konsistensi data untuk setiap layanan kesehatan. Dari 1000 data, *cluster* akut terdiri dari 36 item dan non akut terdiri dari 624 item. Ini disebabkan oleh banyaknya data di poliklinik, rumah sakit, puskesmas dan layanan kesehatan. Untuk menemukan informasi tersembunyi dalam data multimediasi dan meningkatkan konsistensi data layanan kesehatan, teknik pengelompokan data seperti *clustering k-means* dapat digunakan.

Oleh (Azkar & Kusrini, 2024) yang berjudul " Clusterisasi Pasien Rawat Jalan di Puskesmas dengan Menggunakan Metode Algoritma *Clustering* K-Means" menyatakan bahwa hasil pengolahan dan analisis penerapan algoritma *clustering k-means* pasien rawat jalan di Puskesmas Aikmel Utara Kabupaten Lombok Timur, maka diperoleh kesimpulan yaitu pengelompokan pasien rawat jalan dengan atribut jenis kelamin, kelompok usia, status kepesertaan BPJS, alamat pasien, dan diagnosa penyakit pasien menggunakan clustering k-means dengan jumlah 1889 data pasien, menunjukkan bahwa jumlah cluster yang dibentuk sebanyak 2 cluster, dengan jumlah data masing-masing cluster yaitu cluster 1 sebanyak 1570 data dan cluster 2 sebanyak 319 data. Cluster 1 didominasi oleh pasien perempuan sebanyak 1074 (68,4%), pasien peserta BPJS sebanyak 819 (52,2%), dengan kelompok usia terbanyak usia dewasa sebanyak 883 (56,2%), dan terbanyak dari desa Toya sebanyak 488 (31,1%), serta diagnose penyakit terbanyak ISPA akut (J06) sebanyak 223 (14,2%). Sedangkan cluster 2 didominasi oleh pasien perempuan sebanyak 205 (64,3%), pasien peserta BPJS sebanyak 202 (63,3%), dengan kelompok usia terbanyak usia dewasa sebanyak 191 (59,9%), dan terbanyak dari desa Toya sebanyak 108 (33,9%), serta diagnose penyakit terbanyak pemeriksaan kehamilan (Z34) sebanyak 29 (9,1%). Dari hasil cluster ini diperoleh informasi bahwa pengunjung rawat jalan di Puskesmas Aikmel Utara Kabupaten Lombok Timur terbanyak adalah dari desa Toya dan didominasi kelompok usia dewasa dan peserta BPJS kesehatan, serta penyakit terbanyak adalah ISPA akut, sehingga diharapkan dapat membantu pihak Puskesmas dalam pengambilan keputusan atau kebijakan yang berkaitan dengan program-program kesehatan di wilayah kerjanya.

Setelah melihat beberapa penelitian, dapat disimpulkan bahwa penelitian ini berfokus pada pengelompokan data rekam medis rawat jalan di PKU Muhammadiyah Grabag dengan variabel jenis kelamin dan unit pelayanan. Penelitian ini menggunakan metode *k-means clustering* yang merupakan salah satu teknik data *mining* untuk mengelompokkan data secara efektif. Metode *elbow* dan *silhoutte* digunakan untuk menentukan jumlah *cluster* yang optimal, sehingga pengelompokan data menjadi lebih akurat dan terstruktur. Dengan menerapkan metode ini, diharapkan sistem dapat memberikan informasi yang lebih mendalam dan berguna bagi pihak rumah sakit dalam menganalisis data pasien, sehingga dapat membantu dalam pengambilan keputusan yang lebih tepat terkait pelayanan kesehatan. Perbedaan utama antara penelitian ini dengan penelitian relevan sebelumnya adalah fokus pada pengelompokan data rekam medis berdasarkan variabel yang lebih spesifik, serta pemanfaatan metode *k-means* untuk menghasilkan *cluster* yang optimal dalam konteks pelayanan kesehatan di rumah sakit.

2.2 Landasan Teori

2.2.1. Rekam Medis

Menurut PERMENKES RI No 269/MENKES/PER/III/2008 pengertian rekam medis didefinisikan sebagai berkas yang berisikan catatan dan dokumen mengenai identitas pasien, hasil pemeriksaan, pengobatan, tindakan medis, serta pelayanan lain yang telah diberikan kepada pasien dalam rangka pelayanan kesehatan. Rekam medis disusun secara tertib dan sistematis, baik dalam bentuk tertulis, elektronik, maupun kombinasi keduanya. Tujuan penyusunan rekam medis adalah untuk menjamin ketertelusuran pelayanan kesehatan yang diberikan, serta sebagai alat bukti hukum, media pendidikan, bahan penelitian, dan dasar untuk memantau kualitas pelayanan kesehatan. Oleh karena itu, rekam medis memiliki peran penting dalam menjaga kualitas dan kesinambungan pelayanan kesehatan, serta dalam mendukung berbagai fungsi administratif dan legal di fasilitas kesehatan (Amran et al., 2022).

2.2.2. Data *Mining*

Data *mining* adalah proses untuk menemukan pola, hubungan, atau informasi yang berguna dari kumpulan data yang besar dan kompleks. Teknik ini memanfaatkan berbagai metode analisis data, termasuk statistik, pembelajaran mesin, dan teknik berbasis basis data, untuk mengidentifikasi pola tersembunyi dan hubungan yang signifikan di dalam data. Tujuan utama dari data *mining* adalah untuk mengubah data mentah menjadi informasi yang bermakna yang dapat digunakan untuk pengambilan keputusan strategis, analisis prediktif, dan peningkatan efisiensi dalam berbagai bidang seperti bisnis, kesehatan, pemasaran dan penelitian ilmiah (Purba et al., 2023).

2.2.3 *Clustering*

Clustering adalah metode data *mining* yang digunakan untuk mengelompokkan data ke dalam beberapa kelompok atau cluster berdasarkan kemiripan karakteristik di antara data tersebut. Dalam *clustering*, data yang memiliki karakteristik serupa akan dikelompokkan bersama dalam satu *cluster*, sedangkan data yang berbeda akan ditempatkan di *cluster* yang berbeda. Teknik ini sering digunakan untuk menemukan pola atau struktur tersembunyi dalam data yang besar dan kompleks, tanpa memerlukan label atau kategori sebelumnya. *Clustering* memiliki aplikasi luas dalam berbagai bidang, termasuk segmentasi pasar, pengelompokan pelanggan, analisis citra, dan pengelompokan data medis, serta membantu dalam memahami dan menganalisis data dengan lebih efisien (Suryadi & Supriatna, 2019).

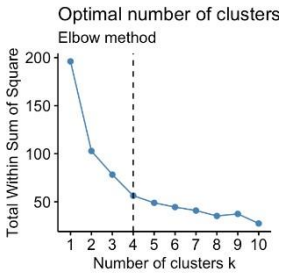
2.2.4 Algoritma *K-Means*

K-Means Clustering adalah yaitu salah satu dari metode pengelompokan data nonhierarki (sekatan) yang dapat mempartisi data kedalam bentuk dua kelompok ataupun lebih. Metoda tersebut akan mempartisi data kedalam suatu kelompok dimana data yang berkarakteristik sama akan dimasukkan kedalam satu kelompok sama sedangkan data yang memiliki karakteristik yang berbeda akan dikelompokan kedalam kelompok lainnya. Tujuan dari pengelompokan yaitu untuk

meminimalkan dari fungsi objektif yang diset dalam proses pengelompokan, pada umumnya akan berusaha meminimalkan variasi didalam suatu kelompok dan memaksimalkan variasi antar kelompok (Gustientiedina et al., 2019).

2.2.5 *Elbow Method*

Elbow Method adalah teknik dalam analisis cluster untuk menentukan jumlah cluster optimal dalam suatu dataset. Teknik ini mengevaluasi variasi atau inersia dalam cluster yang terbentuk dengan jumlah cluster yang berbeda. Langkah pertama dalam metode ini adalah menguji berbagai nilai untuk jumlah *cluster* yang mungkin, biasanya dari 1 hingga nilai tertentu yang sesuai dengan karakteristik data dan tujuan analisis. Algoritma *clustering* seperti *K-Means Clustering* digunakan pada dataset dengan jumlah *cluster* yang berbeda-beda. Setelah proses *clustering*, inersia atau variasi dalam setiap *cluster* dihitung. Kemudian, grafik *Elbow* dibuat dengan jumlah cluster pada sumbu X dan inersia pada sumbu Y (Atira & Nurina Sari, 2023).

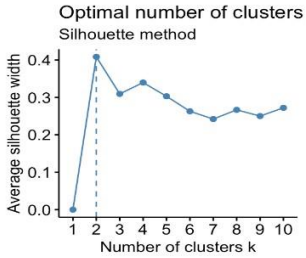


Gambar 2. 1 Diagram *Elbow Method*

2.2.6 *Silhoutte Method*

Silhoutte Coefficient adalah metode untuk mengevaluasi kualitas *cluster* dalam analisis data. Metode ini mengukur seberapa baik data sesuai dengan *cluster* tempat mereka berada dibandingkan dengan *cluster* lainnya. Nilai dihitung untuk setiap objek dalam *cluster*, menunjukkan seberapa dekat objek tersebut dengan objek lain dalam cluster yang sama dibandingkan dengan objek di cluster lain. Nilai *Silhouette Coefficient* berkisar antara -1 hingga 1, dengan nilai positif menunjukkan kecocokan objek dengan *cluster*nya, dan nilai negatif menunjukkan objek mungkin lebih cocok dengan cluster lain. Metode ini sering digunakan bersama *elbow*

method untuk menentukan jumlah *cluster* optimal dalam sebuah dataset (Sholeh & Aeni, 2023).



Gambar 2. 2 Diagram *Silhoutte Method*

2.2.7 *Python*

Selain itu, penelitian ini menggunakan bahasa pemgrograman *Python. Python* adalah salah bahasa pemrograman yang pertama kali dikembangkan oleh Guido van Rossum pada tahun 1991. Bahasa ini dirancang dengan penekanan pada kejelasan, kesederhanaan, dan kemudahan pemahaman. *Python* terkenal karena sintaksnya yang ekspresif, membuatnya mudah digunakan. Penggunaan *Python* telah berkembang luas di berbagai bidang, termasuk pengembangan perangkat lunak, analisis data, pengembangan web, kecerdasan buatan, serta pemrosesan gambar dan suara. *Python* juga dilengkapi dengan *library* yang menawarkan berbagai fitur dan fungsi yang siap pakai tanpa perlu mengunduh puastaka tambahan (Alfarizi et al., 2023).

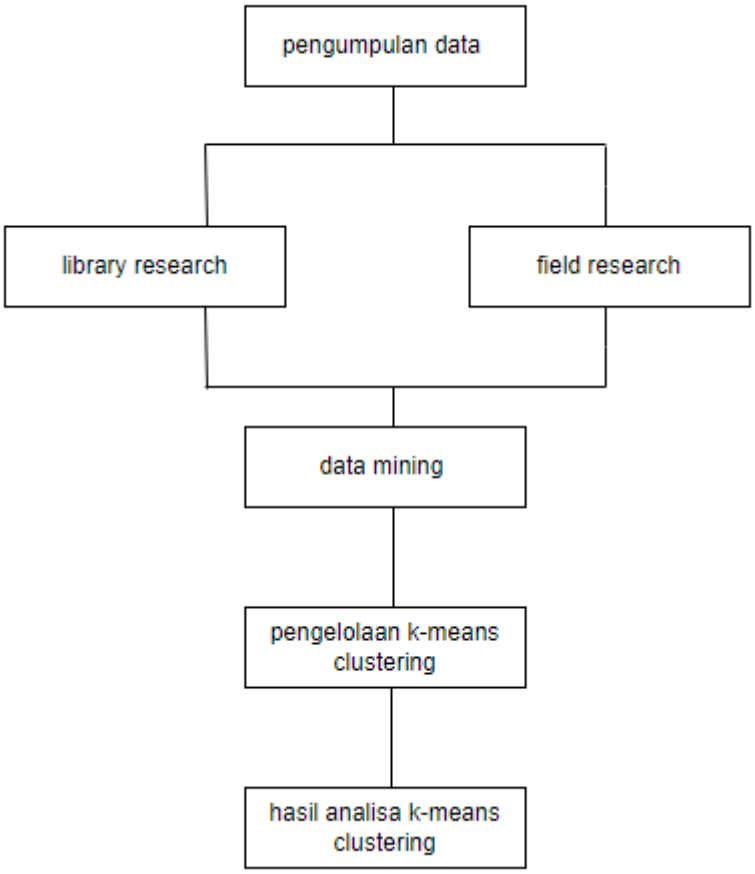
2.2.8 *Wss*

Within-Cluster Sum of Squares (WSS) atau sering disebut sebagai inersia, adalah metrik kunci yang digunakan dalam perhitungan *elbow method* untuk menentukan jumlah cluster yang optimal. WSS mengukur seberapa dekat titik-titik data dalam setiap *cluster* terhadap *centroid*-nya. WSS dihitung sebagai jumlah dari kuadrat jarak antara setiap titik data dan *centroid* dari *cluster* di mana titik tersebut berada. *Elbow method* kemudian mengukur WSS untuk berbagai jumlah *cluster* dan menampilkan grafik dari WSS jumlah *cluster*. Tujuannya adalah untuk menemukan titik di mana penurunan WSS mulai melambat (yaitu, titik *elbow*). (SAPUTRA et al., 2020)

BAB 3
METODOLOGI PENELITIAN

3.1 Prosedur Penelitian

Metode penelitian yang digunakan dalam menyelesaikan penelitian adalah dengan mengumpulkan informasi dari *library research* serta *field research*, setelah dikumpulkan dilanjutkan dengan Pengelolaan data *mining*, sesudah melakukan proses pengelolaan data *mining* berikutnya dilanjutkan dengan menggunakan *k-means clustering* untuk mencari informasi jenis perawatan yang sedang dirawat pada PKU Muhammadiyah Grabag. Proses penelitian terdapat pada Gambar 3.1



Gambar 3. 1 *Flowchart* tahapan penelitian

3.1.1 Pengumpulan Data

Pengambilan dataset yang dilakukan secara langsung di PKU Muhammadiyah Grabag dengan menggunakan surat perintah resmi. Dalam proses ini, peneliti secara fisik berada di PKU Muhammadiyah Grabag untuk mengumpulkan informasi yang relevan terkait rekam medis pasien rawat jalan. Proses pengumpulan data mencakup pengamatan langsung terhadap pengelolaan rekam medis, serta interaksi dengan staf yang bertanggung jawab atas administrasi rekam medis. Selain itu, peneliti juga dapat melakukan wawancara dengan petugas terkait untuk mendapatkan wawasan lebih lanjut mengenai praktik pengelolaan data medis. Dengan terjun langsung ke lapangan, pengambilan dataset ini memungkinkan peneliti untuk memperoleh data yang akurat dan pemahaman mendalam tentang pengelolaan rekam medis di PKU Muhammadiyah Grabag. Data yang digunakan dalam penelitian ini mencakup periode dari tahun 2023 hingga 2024. Variabel yang diambil dalam dataset meliputi jenis kelamin pasien dan unit pelayanan yang digunakan. Dengan mengumpulkan dan menganalisis data ini, peneliti dapat mengidentifikasi pola dalam penggunaan layanan kesehatan dan distribusi pasien berdasarkan karakteristik tertentu selama periode tersebut, yang kemudian dapat digunakan untuk meningkatkan efisiensi dan kualitas pelayanan di PKU Muhammadiyah Grabag.

Tabel 3.1 Data Teratas Rekam Medis Rawat Jalan Tahun 2023-2024

No.	No. RM	JK	Umur	Status	Tgl Reg	Cara Bayar	Unit Pelayanan
1	00003485	LakiLaki	71 Th	Lama	15/09/2023 11:26	BPJS / JKN	PoliSaraf
2	00001739	Perempuan	6 Th	Lama	20/09/2023 17:35	BPJS / JKN	PoliAnak
3	00001896	Perempuan	80 Th	Lama	22/09/2023 15:30	BPJS / JKN	PoliSaraf
4	00003262	LakiLaki	64 Th	Lama	22/09/2023 15:30	BPJS / JKN	PoliSaraf
5	00002670	Perempuan	63 Th	Lama	22/09/2023 15:38	BPJS / JKN	PoliSaraf
6	00003403	Perempuan	15 Th	Lama	22/09/2023 15:47	BPJS / JKN	PoliAnak

Tabel di atas menampilkan data pasien rawat jalan di PKU Muhammadiyah Grabag, yang mencakup variabel seperti nomor rekam medis (RM), jenis kelamin (JK), umur, status pasien, tanggal registrasi, metode pembayaran, dan unit pelayanan yang dikunjungi. Data ini mencakup enam pasien dengan rincian sebagai berikut: seorang laki-laki berusia 71 tahun yang mengunjungi Poli Saraf

pada tanggal 15 September 2023; seorang perempuan berusia 6 tahun yang mengunjungi Poli Anak pada tanggal 20 September 2023; seorang perempuan berusia 80 tahun yang mengunjungi Poli Saraf pada tanggal 22 September 2023; seorang laki-laki berusia 64 tahun yang juga mengunjungi Poli Saraf pada tanggal yang sama; serta dua pasien perempuan berusia 63 tahun dan 15 tahun yang masing-masing mengunjungi Poli Saraf dan Poli Anak pada tanggal 22 September 2023. Semua pasien ini tercatat sebagai pasien lama dan menggunakan BPJS/JKN sebagai metode pembayaran. Data ini memberikan gambaran mengenai distribusi kunjungan pasien berdasarkan usia, jenis kelamin, dan unit pelayanan yang dikunjungi, serta menunjukkan preferensi pasien terhadap metode pembayaran BPJS/JKN di PKU Muhammadiyah Grabag.

3.1.2 *Library Research*

Tahapan ini melibatkan metode penelitian literatur atau studi pustaka. Peneliti melakukan kajian terhadap berbagai sumber ilmiah, seperti buku, jurnal, dan artikel, untuk memahami teori, metode, dan teknik yang relevan dengan topik penelitian. Tujuan dari *library research* adalah untuk memperoleh landasan teori yang kuat dan mengetahui perkembangan terbaru dalam bidang yang diteliti.

3.1.3 *Field Research*

Metode pengamatan (*field research*) pada penelitian ini menggunakan penelitian lapangan. Pada penelitian ini dataset yang digunakan berasal dari data rekam medis rawat jalan PKU Muhammadiyah Grabag. Dataset ini diambil dari tahun 2023 hingga 2024, data ini mempunyai 7 kolom serta 1704 baris. Yang dimana data ini berisikan informasi data pasien yang tengah dirawat jalan.

3.1.4 Data *Mining*

Pada tahap ini setelah data terkumpul dan dilakukan kajian literatur, tahap berikutnya adalah data *mining*. Data mining adalah proses penemuan pola atau informasi yang berguna dari kumpulan data yang besar. Dalam konteks ini, teknik-teknik data *mining* digunakan untuk mengekstraksi informasi penting yang akan digunakan dalam proses pengelompokan atau *clustering*.

3.1.5 *K-Means Clustering*

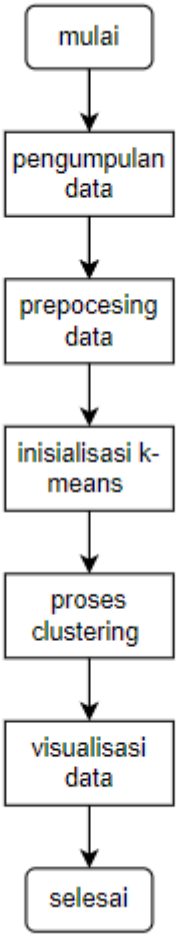
Pada tahap ini *K-Means Clustering* digunakan untuk mengelompokkan data yang memiliki kesamaan karakteristik atau pola tertentu ke dalam suatu *cluster* dengan memilih sejumlah titik k *cluster* sebagai *centroid*, kemudian objek data akan dikelompokkan ke dalam kelompok yang memiliki *centroid* terdekat dengannya berdasarkan *distance space*.

3.2 Hipotesa

Hipotesis pada penelitian ini didasarkan pada asumsi bahwa penerapan algoritma *K-Means Clustering* dalam pengelompokan data rekam medis rawat jalan di PKU Muhammadiyah Grabag, berdasarkan variabel jenis kelamin dan unit pelayanan, akan menghasilkan model yang efektif dalam mengidentifikasi pola penggunaan layanan kesehatan. Dengan memanfaatkan teknik *clustering* ini, diharapkan dapat terbentuk kelompok-kelompok pasien dengan karakteristik yang serupa secara akurat, sehingga hasil pengelompokan ini dapat memberikan informasi yang berguna dalam mendukung pengambilan keputusan yang lebih tepat dan efisien terkait perencanaan serta optimalisasi pelayanan kesehatan di PKU Muhammadiyah Grabag.

3.3 Tahapan Perancangan Percobaan

Tahap rancangan percobaan dalam penelitian ini dijelaskan pada Gambar 3.2 tahap rancangan percobaan berikut:



Gambar 3.2 *Flowchart* Tahap Rancangan Percobaan

Tahap perancangan percobaan yang pertama adalah tahap pengumpulan data primer yang diperoleh langsung dari PKU Muhammadiyah Grabag. Data yang berhasil dikumpulkan berjumlah 1704 data pasien dengan variabel jenis kelamin dan unit pelayanan yang tercatat dari tahun 2023 hingga 2024. Pada tahap kedua, dilakukan proses preprocessing data, yang meliputi pembersihan data untuk mengidentifikasi dan mengatasi inkonsistensi, serta normalisasi data untuk memastikan bahwa data berada dalam skala yang tepat sebelum masuk ke proses pemodelan. Setelah itu, pada tahap ketiga, dilakukan inisialisasi algoritma *K-Means Clustering* dengan menetapkan jumlah cluster (k) yang diinginkan. Jumlah cluster ditentukan menggunakan metode *Wss* dan divisualisasikan menggunakan *Elbow*

Method, yang membantu mengidentifikasi jumlah *cluster* optimal dengan menganalisis variasi inersia dalam setiap *cluster*.

Setelah menetapkan jumlah *cluster*, tahap selanjutnya adalah menentukan *centroid* awal secara acak. *Centroid* ini menjadi acuan dalam proses pengelompokan data, di mana jarak antara setiap data pasien dan *centroid* dihitung menggunakan metrik jarak tertentu. Setiap pasien kemudian dikelompokkan ke dalam *cluster* dengan *centroid* terdekat. Proses ini berlanjut dengan memeriksa apakah ada perpindahan data dari satu *cluster* ke *cluster* lain. Jika tidak ada perpindahan, algoritma dianggap telah mencapai konvergensi. Namun, jika ada perpindahan, *centroid* diperbarui dengan menghitung ulang berdasarkan posisi data dalam *cluster*. Langkah ini diulang hingga tidak ada lagi perpindahan data, menandakan tercapainya konvergensi. Tahap terakhir dari perancangan percobaan ini adalah proses visualisasi hasil analisis data. Visualisasi ini nantinya akan digunakan oleh tenaga medis untuk mengidentifikasi kelompok unit pelayanan dengan jumlah pasien terbanyak selama satu tahun terakhir, membantu dalam pengambilan keputusan dan peningkatan layanan kesehatan di PKU Muhammadiyah Grabag.

3.4 Tahap Pengujian

Tahap pengujian untuk analisis data dengan algoritma *K-Means Clustering* melibatkan beberapa langkah penting. Pertama, jumlah cluster yang optimal ditentukan menggunakan *elbow method*, di mana metode ini membantu mengidentifikasi titik optimal yang dapat meminimalkan nilai dalam inersia *cluster*. Setelah jumlah *cluster* ditentukan, algoritma *K-Means Clustering* diterapkan dengan menggunakan jumlah *cluster* tersebut untuk melakukan pengelompokan data. Evaluasi kualitas *cluster* kemudian dilakukan menggunakan metode *Silhouette*, yang mengukur sejauh mana setiap objek data cocok dengan *cluster*-nya sendiri dibandingkan dengan *cluster* lainnya. Hasil evaluasi ini selanjutnya akan divisualisasikan melalui tabel untuk memberikan gambaran visual mengenai distribusi dan kualitas data pengelompokkan data.

BAB 5
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil dan pembahasan penelitian ini, dapat disimpulkan bahwa pengelompokan data pasien rawat jalan di Pusat Kesehatan Umat (PKU) Muhammadiyah Grabag berhasil dilakukan menggunakan metode *K-Means Clustering*. Dari analisis terhadap 1704 data pasien yang mencakup berbagai unit pelayanan, diperoleh empat *cluster* yang memberikan pemisahan optimal. Penentuan jumlah cluster optimal menggunakan metode *Elbow* dan *Silhouette* menunjukkan bahwa empat cluster merupakan jumlah yang tepat, dengan penurunan nilai *Within-Cluster Sum of Squares (WSS)* yang signifikan dan peningkatan kualitas cluster berdasarkan nilai *Silhouette Score*. Distribusi cluster menunjukkan karakteristik yang berbeda-beda, di mana *Cluster* 1 didominasi oleh pasien dari poli saraf, *Cluster* 2 oleh pasien dari poli dalam/interma, *Cluster* 3 oleh pasien dari poli umum, dan *Cluster* 4 oleh pasien dari poli anak. Selain itu, terdapat dominasi pasien perempuan di sebagian besar *cluster*, kecuali *Cluster* 3 yang memiliki distribusi jenis kelamin hampir seimbang.

Penelitian ini memberikan manfaat bagi manajemen PKU Muhammadiyah Grabag dalam memahami pola penggunaan layanan berdasarkan karakteristik pasien, seperti unit pelayanan dan jenis kelamin. Dengan adanya pembagian *cluster* ini, pihak manajemen dapat mengoptimalkan alokasi sumber daya dan meningkatkan kualitas pelayanan kesehatan secara lebih terarah dan efisien. Hasil penelitian ini juga membantu mengidentifikasi segmen pasien dengan kebutuhan yang berbeda, sehingga dapat mendukung pengambilan keputusan dalam pelayanan kesehatan yang lebih baik di masa mendatang.

5.2 Saran

a**.** Penelitian lanjutan dapat mempertimbangkan penambahan variabel lain seperti usia pasien, riwayat penyakit, atau jenis obat yang digunakan, untuk mendapatkan hasil *clustering* yang lebih komprehensif dan relevan bagi pengambilan keputusan klinis.

b. Untuk meningkatkan akurasi dan relevansi hasil penelitian, disarankan menggunakan data rekam medis yang lebih besar dan lebih bervariasi, mencakup lebih banyak tahun atau jenis layanan medis.