BAB III
METODE PENELITIAN

1.1 Objek Penelitian

Objek penelitian ini adalah data demografi pasien yang menderita penyakit mata. Fokus penelitian ini adalah untuk mengelompokkan pasien berdasarkan data demografi, termasuk usia, jenis kelamin, dan riwayat kesehatan, serta jenis penyakit mata yang diderita.

a. Tempat Penelitian.

Data diambil dari rekam medis pasien di rumah sakit di Cikampek Jawa Barat, pasien yang memiliki catatan lengkap mengenai penyakit mata. Pengumpulan data dilakukan dengan izin dari pihak yang berwenang dan dengan memperhatikan etika penelitian.

Berikut contoh tabel data pada pasien yang akan diteliti :

Tabel 3.1 Data Pasien

NO	NO.RM	N A M A	A L A M A T	L/P	U M U R (THN)	P O L I	DOKTER	DIAGNOSA
1	012341	Pasien 1	Jl. Sore-sore No.44, Cikampek.	P	67	Mata	dr. Senang ,Sp.M	BLEPHARITIS
2	012344	Pasien 2	Jl. Ramai Siang RT/RW 20/77, Purwakarta.	L	25	Mata	dr. Hadir ,Sp.M	CHALAZION
3	012343	Pasien 3	Perum Indah Sekali Blok Z7 No. 22, Subang	P	20	Mata	dr. Satusatu, Sp.M	MYOPIA
4	012345	Paien 4	Jl. Siang Sore No. 40, Karawang.	P	59	Mata	dr. Satusatua, Sp.M	LACRIMAL GLAND AND DUCT
5	012342	Pasien 5	Perum Kedua Belas Blok K3 No. 102, Purwakarta	P	48	Mata	dr. Hadir ,Sp.M	PRESBYOPIA
.	.....	.....	.....	...	...	.....	.....	.....
.	.....	.....	.....	...		.....	.....	.....
.	.....	.....	.....	...		.....	.....	.....
.	.....	.....	.....	...	...	.....	.....	.....
2703	012347	Pasien 2703	Jl. Kemana-mana RT/RW 77/19, Subang.	P	76	Mata	dr. Senang ,Sp.M	PRESBYOPIA

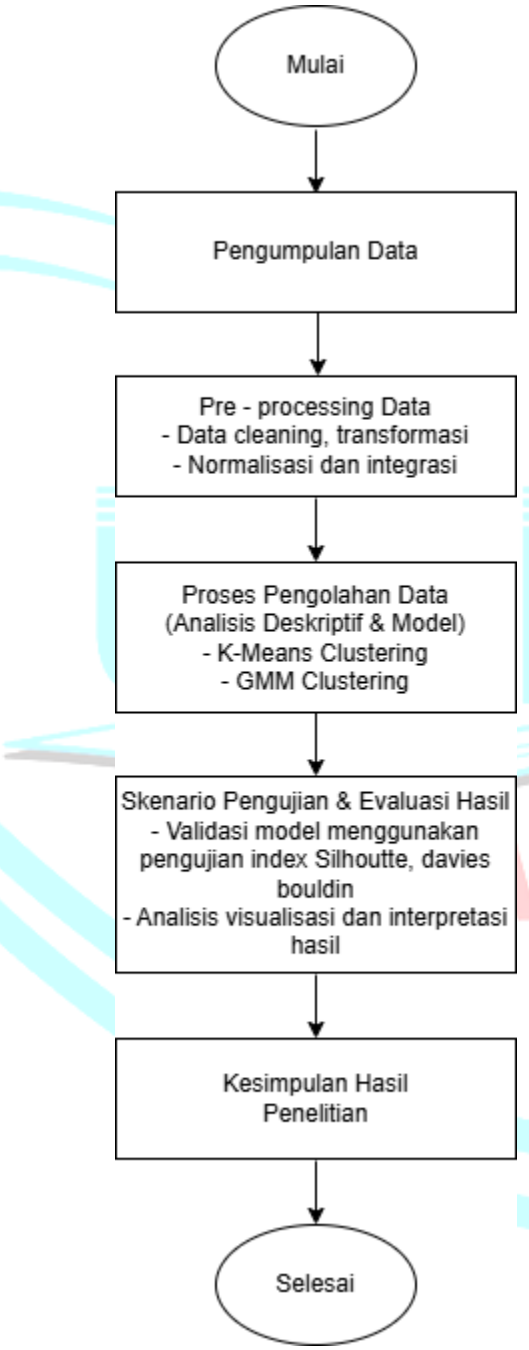
b. Tabel Waktu Penelitian

Tabel 3.2 Waktu Penelitian

No	Deskripsi	2024								2025											
		Des				Jan				Feb				Mar				April			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Identifikasi masalah																				
2	Studi literatur																				
3	Pengumpulan data																				
4	Pre-processing data																				
5	Penerapan algoritma clustering																				
6	Evaluasi model																				
7	Analisis model																				
8	Kesimpulan hasil penelitian																				

3.2 Diagram Alur Proses Penelitian

Berikut adalah diagram alur proses penelitian yang menggambarkan tahapan yang akan dilakukan:



Gambar 3.2 Diagram Alur Proses Penelitian

3.2.1 Tahapan Penelitian

Prosedur penelitian ini terdiri dari beberapa tahapan yang sistematis, yaitu:

1. Pengumpulan Data
   a. Data demografi pasien dikumpulkan dari informasi rekam medis yang ada di rumah sakit.
   b. Sebelum pengumpulan data, peneliti akan mendapatkan izin dari pihak rumah sakit serta memastikan bahwa semua prosedur mengikuti etika penelitian.
2. *Preprocessing* Data

   Data yang telah dikumpulkan akan melalui tahap *pre-processing* untuk mengatasi *missing value*, inkonsistensi, dan melakukan normalisasi agar skala data seragam. Tahapan pra-pemrosesan ini meliputi:
   a. Pembersihan Data : Menghapus atau memperbaiki data yang tidak konsisten atau duplikat.
   b. Transformasi Data : Mengubah format data agar sesuai untuk analisis.
   c. Normalisasi : Menstandarkan nilai atribut agar memiliki rentang yang sama sehingga tidak terjadi bias dalam proses *clustering*. Dengan menggunakan *Z-Score* atau *Robust Scaling.*
3. Proses Pengolahan Data (*Analisis Clustering*)

   Setelah data siap, peneliti melakukan analisis deskriptif untuk memahami distribusi dan karakteristik data demografi pasien. Kemudian, data dianalisis menggunakan dua metode *clustering*, yaitu:
   a. *K-Means* : Algoritma ini mengelompokkan data ke dalam sejumlah *cluster* berdasarkan jarak antara data dan centroid. Proses ini melibatkan penetapan jumlah *cluster* awal, inisialisasi centroid, dan iterasi perhitungan jarak hingga konvergensi. Penentuan jumlah *cluster* (K) akan dilakukan dengan menggunakan metode *elbow*, di mana grafik dari variasi dalam *cluster* (*inertia*) diplot terhadap jumlah *cluster* untuk menemukan titik di mana penurunan *inertia* mulai melambat.

b. *Gaussian Mixture Models* (GMM): Metode probabilistik ini mengasumsikan bahwa data merupakan campuran dari beberapa distribusi *Gaussian*. GMM menggunakan *Expectation-Maximization* (EM) untuk mendekati parameter distribusi sehingga setiap data dapat diatribusikan ke *cluster* dengan probabilitas tertentu. Alur model ini juga mencakup pemilihan atribut, penentuan parameter inisialisasi, evaluasi iteratif, hingga perbandingan hasil akhir kedua metode untuk menentukan model yang paling tepat. Model GMM akan dievaluasi menggunakan kriteria seperti *Akaike Information Criterion* (AIC) dan *Bayesian Information Criterion* (BIC) untuk menentukan model yang paling sesuai.

4. Skenario Pengujian

Skenario pengujian dirancang untuk mengukur validitas dan reliabilitas hasil *clustering*. Langkah-langkah pengujian antara lain:

a. Validasi Internal: Menggunakan metrik seperti :
   1. *Silhouette score* untuk mengukur kualitas *cluster*.
   2. *Davies-Bouldin index* untuk mengukur rata-rata kesamaan antara setiap cluster dengan cluster yang paling mirip dengannya, di mana nilai yang lebih rendah menunjukkan cluster yang lebih baik karena memiliki jarak antar cluster yang lebih besar dan variansi dalam cluster yang lebih kecil.

b. Analisis Visualisasi: Membandingkan hasil *clustering* secara visual melalui plot sebar (*scatter plot*) atau diagram sebar tiga dimensi.

c. Pengujian Parameter: Melakukan eksperimen dengan variasi parameter (misalnya jumlah *cluster* untuk *K-Means* atau jumlah komponen pada GMM) untuk memastikan konsistensi dan stabilitas hasil.

5. Kesimpulan Hasil Penelitian

Dari hasil evaluasi model clustering menggunakan *indeks Silhouette* dan *Davies-Bouldin*, dapat disimpulkan bahwa metode *K-Means* dan *GMM* memberikan hasil pengelompokan yang cukup baik dengan tingkat kohesi dan pemisahan *cluster* yang memadai. Nilai indeks *Silhouette* yang relatif tinggi menunjukkan bahwa data berhasil dikelompokkan secara konsisten, sementara nilai *Davies-Bouldin* yang rendah mengindikasikan *cluster* yang terbentuk cukup rapat dan terpisah dengan jelas. Visualisasi hasil *clustering* juga memperkuat temuan ini dengan menampilkan pola-pola yang bermakna dan mudah diinterpretasikan. Temuan ini menegaskan bahwa pendekatan yang digunakan efektif dalam

mengidentifikasi struktur data yang tersembunyi dan dapat digunakan sebagai dasar analisis lanjutan.

3.3 Model Analisis

Apabila data memiliki label *ground truth*, evaluasi model dapat diperluas dengan menggunakan *confusion matrix* dan metrik evaluasi klasifikasi seperti *accuracy, precision, recall,* dan *F1-score. Confusion Matrix* adalah alat yang digunakan untuk mengevaluasi kinerja algoritma klasifikasi dengan membandingkan hasil prediksi dengan label sebenarnya. Meskipun penelitian ini berfokus pada *clustering*, di mana tidak ada label yang jelas untuk setiap *cluster*, tetapi dapat menggunakan pendekatan ini untuk mengevaluasi seberapa baik model *clustering* dalam mengelompokkan data berdasarkan kategori yang telah ditentukan sebelumnya.

A. Proses Analisis Menggunakan *Confusion Matrix*

1. Definisi Kategori:
   Sebelum menerapkan *Confusion Matrix*, perlu mendefinisikan kategori atau label yang akan digunakan untuk evaluasi. Dalam konteks penelitian ini, kategori dapat berupa jenis penyakit mata yang telah diketahui sebelumnya.
2. Pengelompokan Data:
   Setelah menerapkan algoritma *K-Means* dan GMM, data akan dikelompokkan ke dalam *cluster*. Setiap *cluster* akan memiliki sejumlah pasien yang terkelompok berdasarkan karakteristik demografi mereka.
3. Penentuan *Label Cluster*:
   Untuk setiap cluster yang dihasilkan, perlu menentukan label yang paling umum atau dominan berdasarkan data yang ada. Misalnya, jika cluster 1 terdiri dari 70% pasien dengan penyakit katarak, maka dapat memberi label "Katarak" pada *cluster* tersebut.
4. Membuat *Confusion Matrix*:
   Setelah label ditentukan, dapat membuat *Confusion Matrix* dengan membandingkan label yang diprediksi (hasil *clustering*) dengan label sebenarnya (data yang telah diketahui).