

BAB III

METODOLOGI PENELITIAN

3.1 Metode Pengumpulan Data

Metode pengumpulan data adalah langkah-langkah yang digunakan untuk memperoleh informasi yang relevan dan akurat dalam suatu penelitian. Tujuannya adalah untuk mengumpulkan data yang dapat digunakan untuk menjawab permasalahan atau tujuan penelitian[13]. Dalam penelitian ini, metode pengumpulan data dilakukan melalui tiga tahapan utama yang saling melengkapi untuk mendapatkan informasi yang akurat dan relevan terkait penyakit hipertensi dan risiko penyakit lainnya di Kecamatan Teluk Betung Utara sebagai berikut :

3.1.1 Tahap Pengamatan (Observation)

Pada tahap ini, peneliti akan memahami kondisi dan karakteristik data yang ada, seperti jumlah penderita hipertensi, serta kaitannya dengan penyakit kronis lainnya. Observasi ini bertujuan untuk memahami kondisi awal data dan masalah hipertensi di Puskesmas Sumur Batu yang beralamat Jl. Cut Mutia, Kelurahan Gulak Galik, Kecamatan Teluk Betung Utara, Kota Bandar Lampung, Provinsi Lampung. Hasil pengamatan ini akan menjadi dasar dalam merancang proses pengolahan data yang sesuai dengan tujuan penelitian.

3.1.2 Tahap Wawancara

Wawancara dilakukan dengan tenaga medis di UPT Puskesmas Sumur Batu, serta beberapa pasien yang bersedia memberikan informasi terkait kondisi kesehatan mereka, terutama mengenai penyakit hipertensi yang dialami. Wawancara ini bertujuan untuk mendalami informasi lebih lanjut mengenai faktor-faktor yang dapat memengaruhi kondisi hipertensi serta risiko penyakit yang muncul pada pasien.

3.1.3 Tahap Studi Pustaka

Peneliti akan melakukan studi pustaka untuk mencari dan mempelajari referensi, jurnal, artikel, serta penelitian sebelumnya yang berkaitan dengan penggunaan algoritma *K-Means Clustering* dalam bidang kesehatan, khususnya dalam menganalisis gejala penyakit hipertensi. Tinjauan Pustaka ini juga bertujuan untuk memahami teori-teori dasar mengenai hipertensi dan metode data mining yang digunakan dalam penelitian ini.

3.2 Tempat dan Waktu Penelitian

a. Tempat Penelitian

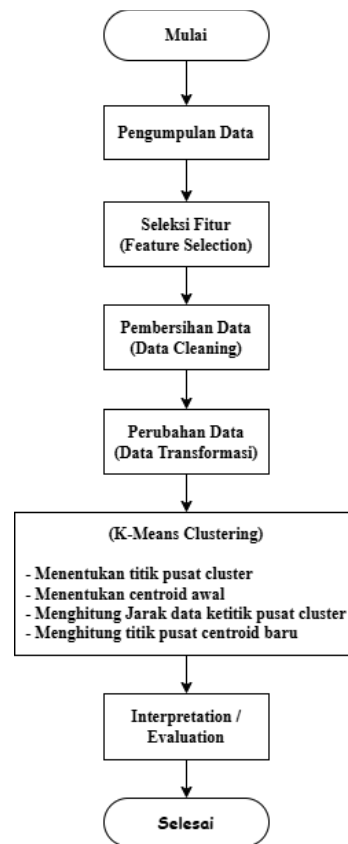
Penelitian akan dilaksanakan di Puskesmas Sumur Batu yang berada di Jl. Cut Mutia, Kelurahan Gulak Galik, Kecamatan Teluk Betung Utara, Kota Bandar Lampung, Provinsi Lampung.

b. Waktu Penelitian

Waktu penelitian akan dilaksanakan pada bulan Oktober 2024 – Februari 2025.

3.3 Tahapan Penelitian

Metode Penelitian yang digunakan dalam penelitian ini akan dijelaskan pada tahapann penelitian. Tahapan Penelitian ini akan menggambarkan alur penelitian dari awal hingga akhir penelitian, guna untuk mencapai tujuan penelitian, yaitu mengimplementasikan algoritma *K-Means Clustering* dalam mengklastrisasi penyakit hipertensi dengan risiko penyakit kronis lainnya. Tahapan-tahapan tersebut dapat dilihat pada Gambar 3.1:



Gambar 3.1 *Tahapan Penelitian*

Berdasarkan pada gambar di atas, tahapan yang dilakukan adalah sebagai berikut :

3.3.1 Pengumpulan Data

Dalam penelitian ini, data dikumpulkan dari rekam medis pasien di UPT Puskesmas Sumur Batu, Kecamatan Teluk Betung Utara. Data yang dikumpulkan meliputi informasi pasien seperti NIK, Nama, Kecamatan, Usia, Jenis Kelamin, Riwayat tekanan darah (sistol dan diastol), IMT, lingkar perut, gangguan penglihatan, pendengaran, riwayat katarak dan PUMA. Pengumpulan data mengacu pada dokumen rekam medis pasien. Peneliti memastikan bahwa data yang dikumpulkan akurat, lengkap, dan relevan dengan tujuan penelitian.

3.3.2 Preprocessing Data

Data mentah yang telah dikumpulkan diolah terlebih dahulu untuk memastikan kualitasnya. Proses ini meliputi Seleksi Data (*Data Selection*), Pembersihan Data (*Data Cleaning*), Perubahan Data (*Transformation Data*), menangani data yang sama (*Duplicate Data*), transformasi data dan Seleksi Fitur agar data siap digunakan dalam analisis.

1. Seleksi Fitur (*Feature Selection*)

Dari data yang diperoleh dari data kunjungan pasien hipertensi UPT Puskesmas Sumur Batu di Kecamatan Teluk Betung Utara, akan dilakukan seleksi data atau atribut yang relevan untuk penelitian ini. Atribut-atribut yang akan digunakan dalam analisis meliputi 'Usia', 'Jenis Kelamin (JK)', 'Hasil - Tekanan Darah (Sistol)', 'Hasil - Tekanan Darah (Diastol)', 'Hasil - Gula Darah Sewaktu', 'Hasil - Indeks Massa Tubuh (IMT)', 'Hasil - Lingkar Perut', 'Tajam Penglihatan', 'Tajam Pendengaran', 'Risiko Katarak', dan 'Skor PUMA'. Setelah menentukan atribut yang relevan, data tersebut yang akan menjadi indikator penentuan kelompok untuk mendukung proses klusterisasi menggunakan algoritma K-Means .

2. Pembersihan Data (*Data Cleaning*)

Pada tahap *Data Cleaning*, data yang telah dikumpulkan diperiksa dan dibersihkan untuk memastikan kualitasnya. Proses ini melibatkan memperbaiki data yang salah atau tidak lengkap, serta menghapus entri yang tidak relevan atau terdapat *coloum* yang kosong. Langkah ini sangat penting agar data yang digunakan menjadi lebih akurat dan siap untuk analisis menggunakan algoritma K-Means.

3. Perubahan Data (*Transformation Data*)

Pada tahap *Transformation Data*, data yang telah dibersihkan akan diubah menjadi format numerik agar dapat digunakan dalam analisis lebih lanjut. Fitur-fitur yang akan dikelola dalam transformasi ini meliputi 'Jenis Kelamin (JK)', 'Tajam Penglihatan', 'Tajam Pendengaran', dan 'Risiko Katarak'. Untuk fitur 'JK', data kategorikal akan diubah menjadi angka, misalnya 0 untuk laki-laki dan 1 untuk perempuan. Sedangkan untuk fitur 'Tajam Penglihatan', 'Tajam Pendengaran', dan 'Risiko Katarak', nilai-nilai yang ada akan dikategorikan dalam bentuk numerik berdasarkan kriteria tertentu. Proses ini penting untuk memastikan bahwa semua data dapat diolah dengan algoritma K-Means secara efektif.

3.3.3 *K-Means Clustering*

Setelah data melalui tahap pemrosesan, pembersihan, dan transformasi, langkah selanjutnya adalah melakukan analisis kluster menggunakan algoritma K-Means Clustering. Algoritma ini bertujuan untuk mengelompokkan data pasien hipertensi ke dalam beberapa cluster berdasarkan karakteristik tertentu. Dalam penelitian ini, peneliti telah menentukan jumlah cluster sebanyak 3 cluster. Proses K-Means Clustering dilakukan melalui beberapa tahapan berikut:.

1. Menentukan Titik Pusat cluster

Metode Elbow digunakan untuk menentukan jumlah cluster terbaik yang akan dibentuk dalam analisis K-Means. Dalam penelitian ini, metode ini diterapkan dengan menggunakan Google Collab. Peneliti dapat mengidentifikasi titik elbow pada grafik, titik ini menunjukkan jumlah cluster yang optimal.

2. Menentukan centroid awal

Centroid awal berfungsi sebagai titik referensi untuk menghitung jarak antara data dan titik pusat cluster. Dalam penelitian ini, centroid awal ditentukan secara acak dari data yang ada. Setelah centroid ditentukan, perhitungan jarak Euclidean dilakukan untuk setiap data ke centroid tersebut. Berikut centroid awal yang dapat dilihat pada tabel 3.1:

Tabel 3. 1 Centroid Data

	Usia	JK	Hasil - Tekanan Darah (Sistol)	Hasil - Tekanan Darah (Diastol)	Hasil - Gula Darah Sewaktu	Hasil - IMT	Hasil - Lingkar Perut	Tajam Penglihatan	Tajam Pendengaran	Risiko Katarak	Skor PUMA
C0	35	0	110	70	100	19,1	73	2	1	2	2
C1	42	1	140	90	109	26,6	85	2	1	2	5
C2	42	0	150	90	200	24,4	80	2	1	2	4

3. Menghitung jarak data ke titik pusat cluster

Pada tahap ini, perhitungan jarak dilakukan untuk setiap data ke centroid yang telah ditentukan pada iterasi pertama. Setiap data akan dihitung jaraknya menggunakan rumus Euclidean. Hasil dari perhitungan ini akan menunjukkan seberapa dekat setiap data dengan masing-masing centroid, dan berdasarkan jarak terkecil, data akan dialokasikan ke cluster terdekat. Berikut adalah contoh perhitungan data yang telah dilakukan iterasi pertama.

$$De = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2 + \dots + (x_{ki} - s_{ki})^2}$$

Keterangan:

De = *Euclidean Distance*

i = Banyaknya Objek

(x, y) = Koordinat objek

(s, t) = Koordinat *centroid*

Dari hasil perhitungan diatas maka hasil iterasi dapat dilihat pada Tabel 3.2:

Tabel 3. 2 Contoh perhitungan iterasi pertama

Usia	JK	Hasil - Tekanan Darah (Sistol)	Hasil - Tekanan Darah (Diastol)	Hasil - Gula Darah Sewaktu	Hasil - IMT	Hasil - Lingkar Perut	Tajam Penglihatan	Tajam Pendengaran	Risiko Katarak	Skor PUMA	C0	C1	C2	Jarak Terdekat	Cluster
52	1	180	100	100	23,8	65	2	1	2	4	78,61355	47,85227	106,4254	47,8522727	1
44	1	110	90	325	23,4	60	2	1	2	4	226,4917	219,5364	132,7818	132,781776	2
52	1	180	100	100	23,8	65	2	1	2	3	78,59447	47,88361	106,4301	47,8836089	1
36	1	110	90	99	23,4	60	2	1	2	3	24,32057	40,9297	110,6345	24,3205674	0
36	1	110	90	99	26,5	60	0	1	2	7	25,60781	40,85352	110,7042	25,6078113	0
47	1	110	70	110	26	80	2	1	2	4	18,59059	36,77445	100,6407	18,590589	0
56	0	140	80	250	26,7	88	2	1	2	7	155,7201	142,0951	54,53705	54,5370516	2
44	1	110	90	325	23,4	60	2	1	2	4	226,4917	219,5364	132,7818	132,781776	2
50	1	140	90	230	22,2	70	1	1	2	4	135,8293	122,2758	34,21754	34,2175394	2
47	1	160	100	100	23,5	73	2	1	2	3	59,71064	27,63349	101,3746	27,6334942	1

4. Menghitung titik pusat centroid baru

Setelah mengalokasikan data ke dalam cluster berdasarkan hasil iterasi pertama, langkah selanjutnya adalah menghitung titik pusat centroid baru. Pada penelitian ini, setelah melakukan perhitungan dan pengelompokan pada iterasi kedua, diperoleh jika posisi centroid sudah stabil, yang berarti tidak ada perubahan signifikan dalam alokasi data ke cluster. Ini menunjukkan bahwa proses clustering telah mencapai konvergensi dan hasil cluster sudah optimal.

3.3.4 Interpretation / Evaluation

Pada tahap ini merupakan proses untuk memahami dan menilai hasil klasterisasi pasien hipertensi di UPT Puskesmas Sumur Batu. Data diolah menggunakan Google Collab untuk menampilkan pola hasil clustering K-Means, yang kemudian divisualisasikan menggunakan softwaren tableau. Visualisasi ini bertujuan untuk menunjukkan pengelompokan antara hipertensi dan risiko penyakit lainnya, sehingga mempermudah analisis pola kesehatan pasien dan memberikan wawasan yang lebih jelas tentang hubungan antar faktor risiko.

3.4 Sampel Data

Penelitian ini menggunakan dataset penyakit hipertensi yang diperoleh dari Puskesmas Sumur Batu, yang terletak di kecamatan Teluk Betung Utara. Dataset yang digunakan berjumlah 21.241 data dalam format *excel*. Pada data tersebut terdapat 19 kolom fitur yang mencakup NIK, Nama peserta deteksi dini, kecamatan, Usia, Jenis Kelamin, Status Tekanan Darah, Hasil tekanan darah (sistol), Hasil tekanan darah (Diastol), Status Gula darah sewaktu, Hasil Gula darah sewaktu, Status IMT, Hasil IMT, Status Lingkar Perut, Hasil Lingkar Perut, Hasil Lemak darah kolesterol total, Tajam penglihatan, tajam pendengaran, risiko katarak, status PUMA, skor PUMA [31].

3.4.1 Dataset Pasien Hipertensi Puskesmas Sumur Batu

Berikut merupakan Sample Dataset Pasien Hipertensi Puskesmas Sumur Batu yang ada pada Gambar 3.3 :

	NIK	Nama Peserta Deteksi Dini	Kecamatan	Usia	JK	Status - Tekanan Darah	Hasil - Tekanan Darah (Sistol)	Hasil - Tekanan Darah (Diastol)	Status - Gula Darah Sewaktu	Hasil - Gula Darah Sewaktu	Status - IMT	Hasil - IMT	Status - Lingkar Perut	Hasil - Lingkar Perut	Tajam Penglihatan	Tajam Pendengaran	Risiko Katarak	Status - PUMA	Skor PUMA
1	1871095501720000	NIRMALA	TELUKBETUNG UTARA	52	female	HIPERTENSI	110.0	90	HIPERGLIKEMIK	325	NORMAL	25.3	NORMAL	60	GANGGUAN BERAT	NORMAL	NORMAL	RISIKO RENDAH PPOK	5
2	1871095304720000	TITA TRIASTUTI	TELUKBETUNG UTARA	50	female	HIPERTENSI	140.0	90	HIPERGLIKEMIK	230	NORMAL	26.9	NORMAL	70	GANGGUAN BERAT	NORMAL	NORMAL	RISIKO RENDAH PPOK	5
3	1871091901880000	CAHYO PRARIAN	TELUKBETUNG UTARA	47	male	HIPERTENSI	140.0	90	HIPERGLIKEMIK	250	NORMAL	25.3	NORMAL	90	GANGGUAN BERAT	NORMAL	NORMAL	RISIKO RENDAH PPOK	5
4	1871094106720000	RAHAYU WAHYUNI	TELUKBETUNG UTARA	47	female	HIPERTENSI	140.0	90	HIPERGLIKEMIK	209	NORMAL	25.3	OBESTITAS SENTRAL	92	GANGGUAN BERAT	NORMAL	NORMAL	RISIKO RENDAH PPOK	6
5	1871090701720000	AHMAD ASRORI	TELUKBETUNG UTARA	47	male	HIPERTENSI	140.0	80	HIPERGLIKEMIK	250	BB LEBIH	25.6	OBESTITAS SENTRAL	88	NORMAL	NORMAL	NORMAL	RISIKO RENDAH PPOK	6
6	1871095301720000	SULYAH	TELUKBETUNG UTARA	56	female	HIPERTENSI	140.0	80	HIPERGLIKEMIK	258	OBESTITAS	26.7	OBESTITAS SENTRAL	92	NORMAL	NORMAL	NORMAL	RISIKO RENDAH PPOK	7

Gambar 3. 2 Sample Data

3.4.2 Meta Data

Berikut merupakan Tabel Meta Data Pasien Hipertensi Puskesmas Sumur Batu yang ada pada Tabel 3.3 :

Tabel 3. 3 Meta Data

Attribute	Description	Value	Type Data
Usia	Usia Pasien dalam tahun	16 - 90	Numerical

JK	Jenis Kelamin Pasien	Male dan Female	Categorical
Status -Tekanan Darah	Kategori Tekanan Darah	Hipertensi, Normal, Pre-Hipertensi	Categorical
Hasil - Tekanan Darah (Sistol)	Nilai tekanan darah sistol dalam mmHg	90 – 210	Numerical
Hasil - Tekanan Darah (Distol)	Nilai tekanan darah diastol dalam mmHg	60 – 150	Numerical
Status - Gula Darah Sewaktu	Kategori kadar gula darah sewaktu	Hiperglikemik, Normal, Pre-diabetes	Categorical
Hasil - Gula Darah Sewaktu	Nilai gula darah sewaktu dalam mg/dL.	60 - 390	Numerical
Status - IMT	Kategori Indeks Massa Tubuh	BB Kurang, BB Lebih, Normal, Obesitas	Categorical
Hasil - IMT	Nilai Indeks Massa Tubuh (kg/m ²)	15 - 49	Numerical
Status - Lingkar Perut	Kategori lingkar perut	Normal, Obesitas Sentral	Categorical
Hasil - Lingkar Perut	Nilai lingkar perut dalam cm	60 – 120 cm	Numerical
Tajam Penglihatan	Kemampuan penglihatan pasien	Normal, Gangguan Berat	Categorical

Ketajaman Pendengaran	Kemampuan pendengaran pasien	Normal, Gangguan Pendengaran	Categorical
Risiko Katarak	Indikasi gejala katarak	Normal, Curiga Katarak	Categorical
Status - PUMA	Status risiko Penyakit Paru Obstruktif Kronis (PPOK) pasien	Normal, Risiko Rendah PPOK, Risiko Tinggi PPOK	Categorical
Skor PUMA	Skor risiko kesehatan terkait PPOK	0 - 10	Categorical