

KLASTERISASI PASIEN PADA RSUD CIAMIS MENGGUNAKAN METODE *K-MEANS*

Eriana Asep Herdianan^{1*}, Aso Sudiarjo², Missi Hikmatyar³

^{1,2}Teknik Informatika; Universitas Perjuangan; Jl. Peta No.177, Kahuripan, Kec.Tawang, Tasikmalaya, Jawa Barat

Received: 1 Agustus 2024
Accepted: 5 Oktober 2024
Published: 12 Oktober 2024

Keywords:
K-Means;
Clustering;
Data Mining;
Penyakit Pasien.

Correspondent Email:
erianaasep.h@gmail.com

Abstrak. Penelitian ini bertujuan mengidentifikasi kluster penyakit dari data pasien rawat inap di RSUD Ciamis pada Desember 2023 menggunakan algoritma *K-Means*. Topik ini penting untuk memahami distribusi dan karakteristik penyakit guna meningkatkan efisiensi layanan kesehatan. Data diproses melalui seleksi, pembersihan, dan transformasi sebelum penerapan algoritma *K-Means*. Dengan bahasa pemrograman *Python*, penelitian ini menghasilkan 3 kluster dari 2.420 data pasien. Kluster 1 terdiri dari 912 pasien dengan penyakit yang melibatkan organ dan sistem tubuh tertentu, kluster 2 terdiri dari 911 pasien dengan penyakit terkait reproduksi, kesehatan perinatal, dan faktor penyebab eksternal, serta kluster 3 terdiri dari 597 pasien dengan penyakit kronis. Evaluasi menunjukkan kluster 3 memiliki kualitas terbaik dengan *Davies-Bouldin Index* (DBI) sebesar 0,406. Hasil penelitian ini penting untuk memperbaiki strategi penanganan penyakit dan alokasi sumber daya di rumah sakit, serta memberikan kontribusi dalam pengembangan metode klusterisasi penyakit di bidang kesehatan.

Abstract. This study aims to identify disease clusters from inpatient data at RSUD Ciamis in December 2023 using the *K-Means* algorithm. This topic is important to understand the distribution and characteristics of diseases to improve the efficiency of health services. The data was processed through selection, cleaning, and transformation before the application of *K-Means* algorithm. Using *Python* programming language, this study generated 3 clusters from 2,420 patient data. Cluster 1 consists of 912 patients with diseases involving specific organs and body systems, cluster 2 consists of 911 patients with diseases related to reproduction, perinatal health, and external causes, and cluster 3 consists of 597 patients with chronic diseases. The evaluation showed that cluster 3 had the best quality with a *Davies-Bouldin Index* (DBI) of 0.406. The results of this study are important for improving disease management strategies and resource allocation in hospitals, as well as contributing to the development of disease Clustering methods in the health sector.

1. PENDAHULUAN

Rumah sakit adalah tempat pelayanan kesehatan yang memberikan layanan kesehatan secara menyeluruh, termasuk rawat jalan, rawat inap, dan gawat darurat. Salah satu tanggung jawab dalam proses pelayanan tersebut adalah

menyediakan rekam medis [1]. Menurut Peraturan Menteri Kesehatan Republik Indonesia Nomor 24 Tahun 2022 tentang rekam medis, rekam medis merupakan dasar dalam proses pelayanan medis, yang mencakup

identitas, pemeriksaan, pengobatan, dan layanan lain yang diberikan kepada pasien.

Di era informasi saat ini, data kesehatan menjadi salah satu komponen penting dalam peningkatan kualitas pelayanan medis. Rumah Sakit Umum Daerah (RSUD) Ciamis, sebagai salah satu institusi kesehatan utama di wilayahnya, mengelola sejumlah besar data pasien yang dapat digunakan untuk berbagai analisis dan pengambilan keputusan. Namun, data yang melimpah ini seringkali belum dimanfaatkan secara optimal, terutama dalam hal pengelompokan penyakit berdasarkan karakteristik pasien. Ketidakefisienan dalam pengelolaan dan analisis data dapat menyebabkan berbagai tantangan, seperti lambatnya respon terhadap tren penyakit dan kurang tepatnya perencanaan sumber daya.

Identifikasi pola dan kluster penyakit dari data pasien yang tersedia masih menjadi tantangan tersendiri. Analisis data pasien secara mendalam diperlukan untuk menemukan pola yang dapat membantu dalam pengambilan keputusan medis dan pengelolaan rumah sakit. Tanpa mengetahui kluster penyakit, penanganan bisa menjadi kurang optimal dan alokasi sumber daya tidak efisien, yang pada akhirnya dapat mempengaruhi pencegahan dan pengendalian penyebaran penyakit.

Penelitian ini penting untuk meningkatkan efisiensi dan efektivitas layanan di RSUD Ciamis. Klusterisasi data pasien akan membantu rumah sakit memahami pola penyakit, mengidentifikasi tren, dan merencanakan sumber daya dengan lebih baik. Teknik ini dapat mempercepat respons terhadap penyakit dan meningkatkan kualitas pelayanan medis. Berdasarkan data rekam medis pasien rawat inap di RSUD Ciamis untuk periode Desember 2023, terdapat berbagai informasi yang perlu dianalisis. Diperlukan keahlian data *scientist* untuk menganalisis data tersebut dan menemukan pola baru dengan menggunakan teknik data mining. Penggunaan bahasa pemrograman *Python* akan membantu dalam pengolahan dan analisis data ini. Algoritma *K-Means* adalah salah satu metode yang efektif dalam data mining untuk mengelompokkan data besar menjadi kelompok-kelompok yang lebih kecil.

Penerapan metode *K-Means Clustering* telah banyak dilakukan oleh peneliti sebelumnya, diantaranya tentang klusterisasi

data rekam medis pasien dengan menggunakan *K-Means Clustering* di rumah sakit Anwar Medika Bolong Bendo Sidoarjo. Hasilnya menunjukkan bahwa terdapat 4 kluster yang terbentuk berdasarkan kecamatan, diagnosa penyakit, usia, dan jenis kelamin. Kluster 1 terdiri dari 79 (15%) pasien perempuan, kluster 2 terdiri dari 214 (50%) pasien laki-laki, kluster 3 terdiri dari 89 (17%) pasien perempuan, dan kluster 4 terdiri dari 152 (28%) pasien laki-laki. Dengan menggunakan 534 data rekam medis pasien dari Rumah Sakit Anwar Medika, pengelompokan ini memberikan informasi baru mengenai pola penyebaran penyakit di setiap kecamatan dengan waktu penyelesaian sebanyak 0,06 detik [2].

Berdasarkan latar belakang tersebut, penelitian berjudul "Klusterisasi Pasien Pada RSUD Ciamis Menggunakan Metode *K-Means*" akan dilakukan. Penelitian ini bertujuan untuk mengidentifikasi kluster penyakit dari data pasien rawat inap periode Desember 2023 di RSUD Ciamis menggunakan algoritma *K-Means*. Diharapkan, pola atau kelompok penyakit yang ditemukan dapat meningkatkan efisiensi dan kualitas pelayanan rumah sakit. Selain itu, penelitian ini juga diharapkan memberikan kontribusi akademis dalam bidang ilmu komputer dan manajemen kesehatan, terutama dalam penerapan teknik data mining untuk pengelolaan rumah sakit.

2. TINJAUAN PUSTAKA

2.1 Rekam Medis

Rekam medis, atau *ICD (International Classification of Diseases)*, adalah dokumen yang mencatat riwayat pengobatan pasien di rumah sakit atau klinik. Dokter mendiagnosis penyakit dan memberikan tindakan yang dicatat dan dikodekan sebagai kode *ICD* oleh ahli rekam medis. Kode *ICD* ini adalah bahasa standar yang memudahkan dokter dari berbagai spesialisasi untuk membaca informasi pasien sesuai aturan. Perkembangan teknologi dalam rekam medis juga membantu dokter dan petugas kesehatan mengakses informasi pasien, mendukung pengambilan keputusan klinis [3].

2.2 Klusterisasi

Pengklusteran atau *Clustering* adalah teknik data mining yang digunakan untuk mengelompokkan data menjadi beberapa subset atau kluster. Teknik ini bertujuan

mendistribusikan objek, orang, atau peristiwa ke dalam kelompok di mana keterhubungan antar anggota dalam satu kluster sangat tinggi, sedangkan keterhubungan antar kluster berbeda lebih rendah. Terdapat dua metode pengelompokan dalam teknik *clustering*, yaitu *hierarchical clustering* dan *non-hierarchical clustering* [4].

2.3 Algoritma K-MEANS

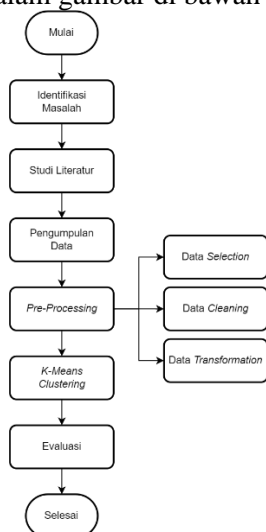
K-Means adalah algoritma klasterisasi *non-hierarchical* yang mengelompokkan data berdasarkan titik pusat atau *centroid*. Metode ini efektif untuk data besar dan hanya cocok untuk atribut numerik. Data dikelompokkan ke dalam kluster berdasarkan kriteria tertentu, dengan setiap kluster memiliki *centroid*. Algoritma ini memulai dengan memilih *k centroid* secara acak, kemudian menghasilkan titik *centroid* akhir berdasarkan jumlah kluster yang ditentukan [5].

2.4 Python

Python adalah bahasa pemrograman yang menekankan pada keterbacaan kode, menggabungkan kemampuan dan kapabilitas yang tinggi. *Python* memiliki sintaks yang jelas dan mudah dipahami serta dilengkapi dengan berbagai fungsi *library* yang lengkap dan banyak. Selain itu, *Python* juga mendukung berbagai paradigma pemrograman seperti fungsional, berbasis objek, dan imperative [6].

3. METODE PENELITIAN

Penelitian ini mencakup serangkaian langkah yang dimulai dari tahap awal hingga penyelesaian akhir. Proses-proses tersebut akan diuraikan dalam gambar di bawah ini.



Gambar 3. 1 Metodologi Penelitian

3.1 Identifikasi Masalah

Langkah awal yang dilakukan oleh peneliti adalah mengidentifikasi permasalahan yang terkait dengan isu yang akan diselidiki melalui pencarian topik. Setelah menentukan permasalahan yang relevan untuk diteliti, penulis merumuskan masalah mengenai Klasterisasi Pasien Pada RSUD Ciamis Menggunakan Metode *K-MEANS*.

3.2 Studi Literatur

Pada langkah ini, dilakukan untuk memperdalam pemahaman tentang teori-teori yang relevan dengan topik penelitian yang sedang dilakukan. Selain itu, studi literatur juga merupakan metode yang digunakan untuk mengidentifikasi sumber-sumber dan referensi yang dapat mendukung penelitian yang dilakukan.

3.3 Pengumpulan Data

Pada penelitian ini data yang digunakan adalah data rawat inap pasien di Rumah Sakit Umum Daerah Ciamis pada bulan Desember 2023 yang terdiri dari 2420 data dengan 8 atribut data yaitu alamat, usia, jenis kelamin, pendidikan, ruang perawatan, jenis pembayaran, Dx Utama, dan Deskripsi.

No	Alamat	Usia	Jenis Kelamin	Pendidikan	Ruang Perawatan	Jenis Pembayaran	Dx Utama	Deskripsi
1	Ciamis	32 (3) Thn (Bln)	Perempuan	SMA/SMK Sederajat	DHL1_KM5	BPJS	K21.9	Gastro-Oesophageal Reflux Disease Without Oesophagitis
2	Ciamis	66 (7) Thn (Bln)	Perempuan	SD Sederajat	DHL2_KM4	BPJS	I20.9	Angina Pectoris. Unspecified
3	Ciamis	61 (1) Thn (Bln)	Perempuan	SD Sederajat	DHL2_KM4	BPJS	E87.1	Hypo-Osmolality And Hyponatremia
4	Ciamis	0 (0) Thn (Bln)	Perempuan	-	RECOVERY BED	BPJS	P59.9	Neonatal Jaundice. Unspecified
5	Ciamis	25 (1) Thn (Bln)	Perempuan	SMP Sederajat	VK OBSERVASI BED-2	VKLAIM SEP BPJS	O60	Preterm Delivery
...
2420	Garut	73 (1) Thn (Bln)	Laki-laki	SMA/SMK Sederajat	HS2_KM10-2	VKLAIM SEP BPJS	A09	Other Gastroenteritis And Colitis Of Infectious And Unspecified Origin

Gambar 3. 2 Dataset Mentah

3.4 Pre-processing Data

Langkah berikutnya adalah pra-pemrosesan data, di mana data mentah yang telah dikumpulkan dipersiapkan untuk kebutuhan model [7]. Sebelum menerapkan K-Means Clustering, data harus melalui proses pre-processing, termasuk seleksi, pembersihan, dan

transformasi data. Berikut adalah tahapan-tahapannya:

3.4.1 Data Selection

Sebelum menerapkan algoritma *K-Means* dalam data mining, data melewati tahap seleksi yang merupakan langkah pertama dalam *KDD (Knowledge Discovery Database)* [8]. Pada tahap ini, dataset awal dikumpulkan dan variabel dipilih. Beberapa atribut tambahan ditambahkan, seperti atribut provinsi untuk pengelompokan alamat pasien, atribut keterangan untuk klasifikasi diagnosa berdasarkan kode *ICD-10*, dan atribut karakteristik untuk membedakan penyakit menular dari tidak menular.

3.4.2 Data Cleaning

Setelah tahap seleksi data, langkah berikutnya adalah pembersihan data, yang melibatkan penghapusan duplikat, perbaikan data yang tidak konsisten, dan perbaikan kesalahan penulisan. Pada tahap ini, dilakukan imputasi untuk menggantikan nilai kosong dengan nilai estimasi seperti rerata, median, atau metode prediksi berbasis statistik atau *machine learning* [2]. Dalam penelitian ini, nilai yang hilang akan digantikan dengan nilai rata-rata variabel terkait. Berikut data yang telah dibersihkan, ditambahkan, digunakan, dan tidak digunakan.

No	Alamat	Provinsi	Usia	Jenis Kelamin	Pendidikan	Jenis Pembayaran	Deskripsi Diagnosa	Keterangan	Karakteristik
1	Ciamis	Jawa Barat	32	Perempuan	SMA/SMK Sederajat	BPJS	Gastro-Oesophageal Reflux Disease Without Oesophagitis	Penyakit pada sistem pencernaan	Tidak Menular
2	Ciamis	Jawa Barat	66	Perempuan	SD Sederajat	BPJS	Angina Pectoris, Unspecified	Penyakit pada sistem sirkulasi	Tidak Menular
3	Ciamis	Jawa Barat	61	Perempuan	SD Sederajat	BPJS	Hypo-Osmolality And Hyponatremia	Gangguan endokrin, nutrisi, dan metabolisme	Tidak Menular
4	Ciamis	Jawa Barat	0	Perempuan	Tidak Sekolah	BPJS	Neonatal Jaundice, Unspecified	Keadaan yang berasal dari periode perinatal	Tidak Menular
5	Ciamis	Jawa Barat	25	Perempuan	SMP Sederajat	VKLAIM SEP BPJS	Preterm Delivery	Kehamilan dan kelahiran	Tidak Menular
...
2420	Garut	Jawa Barat	73	Laki-laki	SMA/SMK Sederajat	VKLAIM SEP BPJS	Other Gastroenteritis And Colitis Of Infectious And Unspecified Origin	Penyakit infeksius dan parasitik	Menular

Gambar 3. 3 Dataset Setelah Dilakukan Cleaning

3.4.3 Data Transformation

Proses transformasi adalah langkah penting dalam mengubah data, setelah proses transformasi selesai, data akan diinisialisasi dan

diubah dari bentuk kata menjadi angka untuk memudahkan perhitungan nantinya.

Tabel 1. Inisialisasi Alamat

Alamat	R
Bandung	1
Bandung Barat	2
Banjar	3
Bekasi	4
Bogor	5
...	...
Yogyakarta	25

Tabel 2. Inisialisasi Provinsi

Provinsi	S
DIY	1
DKI Jakarta	2
Jawa Barat	3
Jawa Tengah	4
Sumatra Barat	5

Tabel 3. Inisialisasi Usia

Usia	T
< 16 Tahun	1
17-25 Tahun	2
26-35 Tahun	3
36-45 Tahun	4
46-55 Tahun	5
>55 Tahun	6

Tabel 4. Inisialisasi Jenis Kelamin

Jenis Kelamin	U
Laki-laki	1
Perempuan	2

Tabel 5. Inisialisasi Pendidikan

Pendidikan	V
Tidak Sekolah	1
SD Sederajat	2
SMP Sederajat	3
SMA/SMK Sederajat	4
Diploma Sederajat	5
S1 Sederajat	6

Tabel 6. Inisialisasi Pembayaran

Jenis Pembayaran	W
ASURANSI	1
BPJS	2
UMUM	3
VKLAIM SEP BPJS	4

Tabel 7. Inisialisasi Diagnosa

Diagnosa	X
Typhoid Fever	1

Diagnosa	X
Bacterial Intestinal Infection. Unspecified	2
Amoebiasis	3
Acute Amoebic Dysentery	4
Other Gastroenteritis And Colitis Of Infectious And Unspecified Origin	5
...	...
Colostomy Status	294

Tabel 8. Inisialisasi Keterangan

Keterangan	Y
Penyakit infeksius dan parasitik	1
Neoplasma	2
Penyakit darah dan organ pembentuk darah, termasuk gangguan sistem imun	3
Gangguan endokrin, nutrisi, dan metabolik	4
Gangguan jiwa dan perilaku	5
...	...
Faktor-faktor yang memengaruhi status kesehatan dan hubungannya dengan jasa kesehatan	19

Tabel 9. Inisialisasi Karakteristik

Karakteristik	Z
Menular	1
Tidak Menular	2

3.5 K-Means Clustering

Dalam kegiatan data mining, dataset yang diperoleh dari RSUD Kota Ciamis dianalisis dengan menggunakan algoritma *K-Means*. Algoritma *K-Means* berfungsi untuk mengelompokkan data ke dalam beberapa kluster. [9]. *K-Means* sering digunakan untuk data numerik dalam kumpulan data besar. Algoritma ini dimulai dengan memilih titik pusat secara acak dan menentukan jumlah kluster yang diinginkan. Data kemudian dikelompokkan ke dalam kluster-kluster berdasarkan kriteria tertentu, dengan setiap kluster memiliki titik pusat yang disebut *centroid*. Proses ini diulangi hingga ditemukan titik pusat yang optimal. Berikut adalah langkah-langkah untuk menerapkan algoritma *K-Means* pada pengelompokan data pasien di RSUD Ciamis.

1. Inisialisasi dimulai dengan menentukan jumlah *cluster* dari dataset yang diberikan. Kemudian, data yang berbeda dipilih secara acak untuk mewakili *centroid* setiap kelompok awal. *Centroid* ini akan disesuaikan setiap iterasi sebelum *cluster*

diperbaiki, mengurangi waktu yang dibutuhkan dalam pengambilan keputusan memilih *centroid*.

2. Mengukur jarak antara data dengan *centroid* dan menetapkan data tersebut ke dalam kelompok yang memiliki *centroid* terdekat dilakukan menggunakan rumus *Euclidean distance* [10]. Rumus yang digunakan adalah:

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

Keterangan:

d_{ij} : Nilai *Euclidean distance*

x_i : Data pertama dari atribut pertama

x_j : *Centroid* untuk atribut pertama

y_i : Data kedua dari atribut kedua

y_j : *Centroid* untuk atribut kedua

3. Ketika semua data telah ditetapkan, hitung kembali posisi *centroid* [11]. Rumus yang digunakan untuk menentukan *centroid* baru adalah:

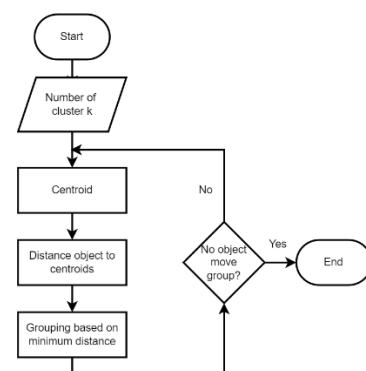
Keterangan:

C_k : *Centroid* baru

n_k : Jumlah data dalam *cluster* k

d_1 : Data dalam *cluster* k

4. Mengulangi langkah 2 dan langkah 3 hingga *centroid* tidak bergerak lagi. Langkah-langkah tersebut dapat digambarkan kedalam bentuk *flowchart* seperti pada Gambar dibawah ini:



Gambar 3. 4 Flowcart K-Means Clustering

3.6 Evaluasi

Pengujian dilakukan dengan metode *Davies-Bouldin Index* (DBI) yang digunakan untuk menentukan kluster terbaik berdasarkan kohesi dan separasi [12]. Kohesi mengukur kedekatan data dengan titik pusat dan kluster yang diikuti, sedangkan separasi mengukur jarak antara titik pusat kluster. Semakin rendah nilai DBI yang

dihasilkan, semakin baik kualitas kluster tersebut.

4. HASIL DAN PEMBAHASAN

4.1 Perhitungan K-Means Clustering

Setelah menginisialisasi semua data dalam bentuk angka, data tersebut dapat dikelompokkan menggunakan algoritma K-Means Clustering. Untuk mengelompokkan data ke dalam beberapa cluster, beberapa langkah perlu dilakukan.

1. Tentukan jumlah cluster yang diinginkan. Dalam penelitian ini, data yang akan dikelompokkan menjadi 3 cluster. Sampel data yang akan digunakan dapat dilihat pada tabel 10 berikut:

Tabel 10. Hasil Transformasi Data

No	R	S	T	U	V	W	X	Y	Z
1	8	3	3	2	4	2	136	11	2
2	8	3	6	2	2	2	86	9	2
3	8	3	6	2	2	2	59	4	2
4	8	3	1	2	1	2	241	16	2
5	8	3	2	2	3	4	219	15	2
...
2420	10	3	6	1	4	4	5	1	1

2. Tentukan titik pusat awal untuk setiap kluster. Dalam penelitian ini, titik pusat awal ditentukan secara acak dan didapatkan titik pusat dari setiap kluster pada tabel sebagai berikut:

Tabel 11. Titik Awal Pusat Cluster

Data ke	Centro id	R	S	T	U	V	W	X	Y	Z
1	C1	8	3	3	2	4	2	136	11	2
50	C2	8	3	1	1	1	3	238	16	2
1200	C3	8	3	5	1	3	2	7	1	1

3. Dalam penelitian ini, data akan ditempatkan ke dalam kluster menggunakan metode K-Means. Setiap data akan dialokasikan ke dalam satu kluster, sehingga data akan dimasukkan ke dalam kluster tersebut. Untuk menentukan kluster yang paling dekat dengan data, jarak setiap data dengan titik pusat setiap kluster perlu dihitung. Sebagai contoh, jarak dari data akan dihitung menggunakan rumus *Euclidean distance*.

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

Menghitung jarak terhadap data pasien dengan titik pusat (*centroid*) pada cluster 1:

- a. Perhitungan pada data pertama:

$$\begin{aligned} D(1,1) &= \sqrt{(8-8)^2 + (3-3)^2} \\ &\quad + (3-3)^2 + (2-2)^2 \\ &\quad + (4-4)^2 + (2-2)^2 \\ &\quad + (136-136)^2 \\ &\quad + (11-11)^2 + (2-2)^2 \\ &= 0 \end{aligned}$$

- b. Perhitungan pada data kedua:

$$\begin{aligned} D(2,1) &= \sqrt{(8-8)^2 + (3-3)^2} \\ &\quad + (6-3)^2 + (2-2)^2 \\ &\quad + (2-4)^2 + (2-2)^2 \\ &\quad + (86-136)^2 \\ &\quad + (9-11)^2 + (2-2)^2 \\ &= 50,16971 \end{aligned}$$

Menghitung jarak terhadap data pasien dengan titik pusat (*centroid*) pada cluster 2:

- a. Perhitungan pada data pertama:

$$\begin{aligned} D(1,2) &= \sqrt{(8-8)^2 + (3-3)^2} \\ &\quad + (3-1)^2 + (2-1)^2 \\ &\quad + (4-1)^2 + (2-3)^2 \\ &\quad + (136-238)^2 \\ &\quad + (11-16)^2 + (2-2)^2 \\ &= 102,1959 \end{aligned}$$

- b. Perhitungan pada data kedua:

$$\begin{aligned} D(2,2) &= \sqrt{(8-8)^2 + (3-3)^2} \\ &\quad + (6-1)^2 + (2-1)^2 \\ &\quad + (2-1)^2 + (2-3)^2 \\ &\quad + (86-238)^2 \\ &\quad + (9-16)^2 + (2-2)^2 \\ &= 152,2531 \end{aligned}$$

Menghitung jarak terhadap data pasien dengan titik pusat (*centroid*) pada cluster 3:

- a. Perhitungan pada data pertama:

$$\begin{aligned} D(1,3) &= \sqrt{(8-8)^2 + (3-3)^2} \\ &\quad + (3-5)^2 + (2-1)^2 \\ &\quad + (4-3)^2 + (2-2)^2 \\ &\quad + (136-7)^2 + (11-1)^2 \\ &\quad + (2-1)^2 \\ &= 129,4141 \end{aligned}$$

- b. Perhitungan pada data kedua:

$$\begin{aligned} D(2,3) &= \sqrt{(8-8)^2 + (3-3)^2} \\ &\quad + (6-5)^2 + (2-1)^2 \\ &\quad + (2-3)^2 + (2-2)^2 \\ &\quad + (86-7)^2 + (9-1)^2 \\ &\quad + (2-1)^2 \\ &= 79,42921 \end{aligned}$$

Berikut hasil perhitungan data pasien dengan titik pusat pada iterasi 1 menggunakan rumus *Euclidean distance*.

No	R	S	T	U	V	W	X	Y	Z	C1	C2	C3	Jarak terdekat	Cluster
1	8	3	3	2	4	2	13	1	2	0	102,195	129,414	0	1
2	8	3	6	2	2	2	86	9	2	50,1697	152,253	79,4292	50,1697119	1
6	8	3	6	1	2	2	10	5	9	31,2889	133,285	98,3412	31,2889756	1
11	8	3	1	2	1	2	11	1	0	19,3649	121,156	110,467	19,3649167	1
14	8	3	1	2	1	2	11	1	0	19,3649	121,156	110,467	19,3649167	1
...
242	1	0	3	6	1	4	4	5	1	131,453	233,568	3,74165	3,74165738	3

Gambar 4. 1 Hasil Perhitungan *Cluster* iterasi 1 dengan *Euclidean distance*

- Setelah semua data ditampilkan kedalam klaster yang terdekat, kemudian hitung kembali pusat klaster yang baru dengan menjumlahkan data masing-masing atribut berdasarkan hasil perhitungan klaster pada iterasi 1 kemudian dibagi dengan banyaknya data pada setiap klaster.

Centro id	R	S	T	U	V	W	X	Y	Z
C1	8,42727	2,99899	4,74444	1,49596	3,01313	2,24040	124,672	10,2545	2
C2	8,70623	2,99640	3,28297	1,67386	2,99760	2,43045	244,034	16,3525	2
C3	8,31375	2,99832	3,44798	1,52516	2,88255	2,28691	25,0822	2,15268	1,4177

Gambar 4. 2 *Centroid* Baru Hasil Iterasi 1

- Setelah menemukan titik pusat baru dari setiap klaster, langkah selanjutnya adalah mengulangi proses dari langkah ketiga dengan menggunakan rumus *Euclidean distance* sampai titik pusat dari setiap klaster tidak berubah dan tidak ada lagi data yang berpindah dari satu klaster ke klaster lainnya.

No	R	S	T	U	V	W	X	Y	Z	C1	C2	C3	Jarak terdekat	Cluster
1	8	3	3	2	4	2	13	1	2	16,4783	103,127	111,201	16,478342	1
2	8	3	6	2	2	2	86	9	2	33,7520	153,183	61,2879	33,7520728	1
6	8	3	6	1	2	2	10	5	9	14,8275	134,212	80,182	14,8275395	1
11	8	3	1	2	1	2	11	1	0	5,06672	122,187	92,2291	5,06672674	1
14	8	3	1	2	1	2	11	1	0	5,06672	122,187	92,2291	5,06672674	1
...
242	1	0	3	6	1	4	4	5	1	115,082	234,507	20,5370	20,5370054	3

Gambar 4. 3 Hasil Perhitungan Klaster Iterasi 5 dengan *Euclidean distance*

Pada iterasi kelima, titik pusat klaster (*centroid*) tidak mengalami perubahan lagi dan tidak ada data yang berpindah dari satu klaster ke klaster yang lain. Ini menandakan bahwa proses klasterisasi telah mencapai konvergensi, dimana setiap data telah diklasterkan ke dalam klaster yang paling sesuai berdasarkan karakteristiknya. Berikut adalah titik *centroid* setelah iterasi ke-5:

Centro id	R	S	T	U	V	W	X	Y	Z
C1	8,41995	2,99890	4,73245	1,49342	2,99780	2,24671	119,69	9,94298	2
C2	8,69045	2,99670	3,41712	1,66191	3,01646	2,40614	238,98	16,1536	2
C3	8,31323	2,99832	3,45226	1,52428	2,87939	2,28978	25,160	2,15912	1,418

Gambar 4. 4 *Centroid* Baru Hasil Iterasi Ke-5

4.2 Deskripsi Hasil *Clustering*

Berikut adalah deskripsi hasil pengelompokan data pasien pada RSUD Ciamis berdasarkan alamat, provinsi, usia, jenis kelamin, pendidikan, jenis pembayaran, diagnosa, keterangan, dan karakteristik. Pada proses ini menghasilkan 3 *cluster* yaitu *cluster* 1 berjumlah 912, *cluster* 2 berjumlah 911, dan *cluster* 3 berjumlah 597. Dengan rincian sebagai berikut:

- Cluster* 1 merupakan kelompok penyakit yang melibatkan organ dan sistem tubuh tertentu. Penyakit yang paling umum adalah *bronchopneumonia*, *unspecified* sebanyak 96 kasus, diikuti oleh *acute subendocardial myocardial infarction* sebanyak 46 kasus. Dari total pasien, terdapat 462 pasien laki-laki dan 450 pasien perempuan. Mayoritas pasien dalam kelompok ini adalah usia lanjut, dengan usia lebih dari 55 tahun.
- Cluster* 2 merupakan kelompok penyakit yang terkait dengan reproduksi, kesehatan perinatal, dan faktor penyebab eksternal. Penyakit yang paling umum adalah *abdominal and pelvic pain* sebanyak 95 kasus, diikuti oleh *fever*, *unspecified* sebanyak 80 kasus. Jumlah pasien laki-laki dalam *cluster* ini adalah 308, sedangkan jumlah pasien perempuan adalah 608. Mayoritas pasien dalam kelompok ini berusia dibawah 16 tahun.
- Cluster* 3 merupakan kelompok penyakit yang bersifat kronis dan kompleks. Penyakit yang paling umum adalah *other gastroenteritis and colitis of infectious and unspecified origin*, dengan jumlah kasus sebanyak 67, diikuti oleh *gastroenteritis and colitis of unspecified origin* sebanyak 53 kasus. Jumlah pasien laki-laki dalam *cluster* ini adalah 284, sedangkan jumlah pasien perempuan adalah 313. Mayoritas pasien dalam kelompok ini berusia dibawah 16 tahun.

4.3 Perhitungan *K-Means* pada *Python*

Implementasi algoritma *K-Means clustering* dilakukan dengan menggunakan bahasa pemrograman *Python* dalam aplikasi *Visual*

studio code. Data yang digunakan telah diinisialisasi dan diekspor dalam format *Microsoft excel* yang disimpan sebagai file *CSV*. Langkah pertama dalam proses ini adalah mengimpor *library* yang diperlukan. *Library* yang dibutuhkan mencakup *Scikit-learn*, *Pandas*, *Matplotlib*, dan *Numpy*. Setelah impor *library*, data dimuat dan ditampilkan menggunakan fungsi **read_csv** dari *library Pandas*, Sementara Fungsi **head()** digunakan untuk menampilkan beberapa baris pertama data

```
#Import library yang diperlukan
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

#Memuat data dari file csv
dataset = pd.read_csv('data_gamam.csv', sep=',', engine='python')
dataset.head()
```

Alamat(R)	Provinsi(S)	Usia(T)	Jenis Kelamin(U)	Pendidikan(V)	Jenis Pembayaran(W)	Diagnosa(X)	Keterangan(Y)	Karakteristik(Z)
0	8	3	3	2	4	2	136	11
1	8	3	6	2	2	2	86	9
2	8	3	6	2	2	2	59	4
3	8	3	1	2	1	2	241	16
4	8	3	2	2	3	4	719	15

Gambar 4. 5 Import Library dan Read Dataset

Pemeriksaan ulang dilakukan terhadap nilai yang hilang (*missing value*) dalam data yang telah diinisialisasi. Hal ini dilakukan dengan menggunakan kode **print(data.isnull().sum())**. Jika *outputnya* adalah nol, maka tidak ada nilai yang hilang dalam data.

```
#Cek Missing Value
print(dataset.isnull().sum())
```

```
0.0s
```

Alamat(R)	0
Provinsi(S)	0
Usia(T)	0
Jenis Kelamin(U)	0
Pendidikan(V)	0
Jenis Pembayaran(W)	0
Diagnosa(X)	0
Keterangan(Y)	0
Karakteristik(Z)	0
dtype: int64	

Gambar 4. 6 Cek Missing Value Dataset Inisialisasi

Selanjutnya, modul **KMeans** dari *library scikit-learn* diimport, yang menyediakan implementasi algoritma *K-Means*. Kemudian tentukan jumlah *cluster* yang diinginkan.

```
# Menginisialisasi model K-Means dengan jumlah kluster yang diinginkan
from sklearn.cluster import KMeans
kmean = KMeans(n_clusters=3)
kmean
```

```
> KMeans
```

```
y_cluster = kmean.fit_predict(x_train)
y_cluster
```

```
array([0, 0, 2, ..., 0, 1, 2])
```

Gambar 4. 7 Import Sklearn Cluster

Kolom baru ditambahkan ke dalam data yang berisi nomor *cluster* untuk setiap entri. Beberapa baris pertama dari data beserta *cluster* yang ditetapkan ditampilkan.

```
# Menambahkan kolom cluster ke dalam data
dataset['cluster'] = y_cluster
```

```
# Menampilkan data dengan kelompok cluster
dataset
```

Alamat(R)	Provinsi(S)	Usia(T)	Jenis Kelamin(U)	Pendidikan(V)	Jenis Pembayaran(W)	Diagnosa(X)	Keterangan(Y)	Karakteristik(Z)	Cluster
0	8	3	3	2	4	2	136	11	2
1	8	3	6	2	2	2	86	9	2
2	8	3	6	2	2	2	59	4	2
3	8	3	1	2	1	2	241	16	2
4	8	3	2	2	3	4	719	15	2
...
2415	8	3	4	2	2	2	281	18	2
2416	8	3	4	1	2	2	134	11	2
2417	8	3	4	2	4	3	134	11	2
2418	8	3	1	2	1	2	258	17	2
2419	10	2	6	1	4	4	5	1	2

Gambar 4. 8 Dataset dengan Kelompok Cluster

Titik pusat (*centroid*) dari setiap *cluster* yang dihasilkan oleh algoritma *K-Means* ditampilkan menggunakan kode **print(kmeans.cluster_centers_)**.

```
# Menampilkan centroid dari setiap kluster
kmean.cluster_centers_
```

```
array([[ 8.41995614,  2.99890351,  4.73245614,  1.49342105,
  2.99780702,  2.24671053, 119.69298246,  9.94298246,
  2.          ],
 [ 8.69045005,  2.99670692,  3.41712404,  1.66190999,
  3.01646542,  2.40614709, 238.98902305, 16.15367728,
  2.          ],
 [ 8.31323283,  2.99832496,  3.45226131,  1.52428811,
  2.87939698,  2.28978224, 25.16080402,  2.15912898,
  1.41876047]])
```

Gambar 4. 9 Centroid Dari Setiap Cluster

Jumlah data per *cluster* ditampilkan. Ini membantu dalam mengetahui seberapa merata *cluster* yang dihasilkan oleh algoritma dengan menghitung jumlah data dalam setiap *cluster*.


```
# Menghitung jumlah data dalam setiap kluster
labels = kmean.labels
dataset['cluster'] = labels
jumlah_data_per_cluster = dataset['cluster'].value_counts()

print("Jumlah data per cluster:")
print(jumlah_data_per_cluster)

Jumlah data per cluster:
cluster
0    912
1    911
2    597
Name: count, dtype: int64
```

Gambar 4. 10 Jumlah Data Dari Setiap Cluster

4.4 Evaluasi

Evaluasi *cluster* yang digunakan pada penelitian ini adalah *Davies-Bouldin Index* (DBI), metode ini bertujuan untuk memaksimalkan pengukuran antara jarak satu *cluster* dengan *cluster* yang lainnya. Pengujian menggunakan *Davies-Bouldin Index* (DBI) dilakukan untuk menilai kualitas dan kecocokan *cluster* yang dihasilkan. Angka yang dihasilkan menunjukkan kualitas *cluster*, di mana nilai yang lebih rendah menandakan kualitas *cluster* yang lebih baik. Berikut adalah hasil evaluasi DBI yang diimplementasikan menggunakan *Python*:

```
# Evaluasi clustering menggunakan Davies-Bouldin Index
from sklearn.metrics import davies_bouldin_score
result = {}
for i in range(2,8):
    kmeans = KMeans(n_clusters=i, random_state=30)
    labels = kmeans.fit_predict(dataset)
    db_index = davies_bouldin_score(dataset, labels)
    result.update({i:db_index})

result

{2: 0.4722879579237086,
 3: 0.40638574958334267,
 4: 0.454048462809053,
 5: 0.50324296669535,
 6: 0.4993932703799295,
 7: 0.48744945852854815}
```

Gambar 4. 11 Jumlah Data Dari Setiap Cluster

Berikut merupakan tabel hasil pengujian menggunakan *Davies-Bouldin Index* (DBI) yang dijalankan di *Python*:

Tabel 12. Hasil Pengujian Davies-Bouldin Index

No	Jumlah Cluster	Davies-Bouldin Index
1	2	0.472
2	3	0.406
3	4	0.454
4	5	0.503
5	6	0.499
6	7	0.487

Berdasarkan hasil yang tercantum dalam tabel, didapatkan bahwa nilai *k* yang optimal adalah *k*=3 dengan *Davies-Bouldin Index* (DBI) sebesar 0.406. Hal ini menunjukkan bahwa *cluster* dengan jumlah 3 memiliki nilai DBI yang rendah dibandingkan dengan *cluster* lainnya.

5. KESIMPULAN

- Berdasarkan analisis *clustering*, dapat disimpulkan bahwa pengelompokan data rekam medis pasien rawat inap RSUD Ciamis pada bulan Desember 2023 sukses dan menghasilkan hasil yang optimal. Dalam pengelompokan pasien menggunakan algoritma *K-Means*, terdapat 2420 data yang terbagi menjadi 3 *cluster*. *Cluster* 1 merupakan kelompok penyakit yang melibatkan organ dan sistem tubuh tertentu memiliki 912 data dengan mayoritas pasien lansia (di atas 55 tahun), terdiri dari 462 laki-laki dan 450 perempuan, dengan penyakit *bronchopneumonia, unspecified* yang paling umum. *Cluster* 2 merupakan kelompok penyakit yang terkait dengan reproduksi, kesehatan perinatal, dan faktor penyebab eksternal memiliki 911 data dengan mayoritas pasien anak-anak dan remaja (dibawah 16 tahun), terdiri dari 308 laki-laki dan 608 perempuan, dengan *abdominal and pelvic pain* sebagai penyakit paling umum. *Cluster* 3 merupakan kelompok penyakit yang bersifat kronis dan kompleks memiliki 597 data dengan mayoritas pasien juga anak-anak dan remaja, terdiri dari 284 laki-laki dan 313 perempuan, dengan penyakit *other gastroenteritis and colitis of infectious and unspecified origin* yang paling umum.
- Hasil klasterisasi data pasien menunjukkan bahwa algoritma *K-Means* K terbaik berhasil mengelompokkan data pasien ke dalam *cluster* 3 dengan nilai evaluasi DBI sebesar 0.406 yang mendekati 0.

UCAPAN TERIMA KASIH

Puji syukur penulis sampaikan kepada Allah Yang Maha Kuasa, atas rahmat dan karunia-Nya yang memberikan kesehatan dan kesempatan sehingga penulis dapat menyelesaikan karya ilmiah ini. Ucapan terima kasih terutama penulis tujuhan kepada orang tua

atas kesabaran, ketulusan hati, dorongan moral dan material, serta doa yang tiada hentinya. Penulis juga menyampaikan terima kasih kepada semua pihak yang terlibat dalam penulisan karya ilmiah ini.

DAFTAR PUSTAKA

- [1] Maryam, "KEBIJAKAN RUMAH SAKIT DALAM PENYERAHAN DOKUMEN RAWAT INAP DARI RUANGAN DAHLIA KE BAGIAN CASE-MIX TAHUN 2021-2022 DI RSUD CIAMIS," pp. 1–6, 2022.
- [2] A. Ali, "Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo," *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 19, no. 1, pp. 186–195, Nov. 2019, doi: 10.30812/matrik.v19i1.529.
- [3] W. S. Nanda, A. M. H. Pardede, and M. Simanjuntak, "ANALISIS DATA MINING UNTUK KLASTERISASI DATA REKAM MEDIS MENGGUNAKAN ALGORITMA K-MEANS PADA RUMAH SAKIT SYLVANI BINJAI," *Indonesian Journal of Education And Computer Science*, vol. 1, no. 3, pp. 82–88, 2023.
- [4] U. Ma'rifatin, "Implementasi Algoritma K-Means untuk Pengelompokan Penyakit Pasien Pada Puskesmas Warujayeng," 2020.
- [5] R. Anggraini, E. Haerani, J. Jasril, and I. Afrianty, "Pengelompokan Penyakit Pasien Menggunakan Algoritma K-Means," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 6, p. 1840, Dec. 2022, doi: 10.30865/jurikom.v9i6.5145.
- [6] M. Maesaroh, T. Nur Padilah, and J. Haerul Jaman, "PENERAPAN ALGORITMA K-MEANS CLUSTERING PADA PENGELOMPOKAN DAERAH PENYEBARARAN DIARE DI PROVINSI JAWA BARAT," 2023. doi: 10.36040/jati.v7i4.7208.
- [7] R. R. Burhanuddin, "KLASIFIKASI PENYAKIT PADI MELALUI CITRA DAUN MENGGUNAKAN METODE NAIVE BAYES," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 2, Apr. 2024, doi: 10.23960/jitet.v12i2.4012.
- [8] A. F. Khairani, A. Nazir, T. Darmizal, Y. Vitriani, and Y. Yusra, "Klasterisasi Peserta BPJS Berdasarkan Rekam Medis Menggunakan Algoritma K-Means," *Journal of Computer System and Informatics (JoSYC)*, vol. 4, no. 3, pp. 625–631, May 2023, doi: 10.47065/josyc.v4i3.3442.
- [9] E. H. J. I. A. Anggraini, "Pengelompokan Penyakit Pasien Menggunakan Algoritma K-Means," vol. 9, no. 6, pp. 1840–1849, 2022, doi: 10.30865/jurikom.v9i6.5145.
- [10] I. N. M. Adiputra, "CLUSTERING PENYAKIT DBD PADA RUMAH SAKIT DHARMA KERTI MENGGUNAKAN ALGORITMA K-MEANS," *INSERT: Information System and Emerging Technology Journal*, vol. 2, no. 2, p. 99, 2021.
- [11] Arif, "CLUSTERING DATA PASIEN PUSKESMAS PUHPELEM MENGGUNAKAN K-MEANS CLUSTERING," 2023.
- [12] M. Orisa, "Optimasi Cluster pada Algoritma K-Means," *Prosiding SENIATI*, vol. 6, no. 2, pp. 430–437, Jul. 2022, doi: 10.36040/seniati.v6i2.5034.