



Pengelompokan Data Rekam Medis pada Pasien Penyakit dalam Untuk Meningkatkan Manajemen Informasi Kesehatan Berdasarkan Wilayah Kota Binjai Menggunakan Algoritma *Clustering K- Means* (Studi Kasus: Artha Medika Binjai)

Desiska Natalia Br. Purba^{1*}, Marto Sihombing², Indah Ambarita³

^{1,2,3} STMIK Kaputama, Indonesia

desiskapurba05@gmail.com^{1*}, martosihombing45@gmail.com², yesnovada@yahoo.com³

Alamat: Jl. Veteran No.4A, Tangsi, Kec. Binjai Kota, Kota Binjai, Sumatera Utara 20714

Korespondensi penulis: desiskapurba05@gmail.com

Abstract. *The history of disease in patients is generally recorded in medical record data in every hospital as well as at Artha Medika Hospital which is a health institution that was established in 2012 in the city of Binjai also has a very large amount of medical record data. However, in using the information management system owned by Artha Medika Hospital, there are weaknesses and it is still limited in managing medical record data in the hospital which is used in making reports to the head of the leadership. Therefore, a system is needed that can assist the hospital in improving health information management to be faster in managing data by approaching using data mining techniques with the k-means method. So that in finding new information based on medical record data of internal medicine patients can be used in the decision-making process by hospital management to be right on target so that it can produce 3 groups of data consisting of Age, Type of disease and Region. From testing on cluster 3, it can be seen that the results of the age group (X), type of disease (Y), region (Z) the amount of data owned is 645 cluster 3 data centred on the centroid of the information of the number of patient medical records data, namely age is 44-52 years, with the type of disease is chronic kidney disease and the region is South Binjai.*

Keywords: *Data Mining, K-Means Algorithm, Medical Records.*

Abstrak. Riwayat penyakit pada pasien umumnya tercatat dalam data rekam medis di setiap rumah sakit begitu juga pada rumah sakit Artha Medika yang merupakan lembaga kesehatan yang berdiri sejak tahun 2012 di kota Binjai juga memiliki jumlah data rekam medis yang sangat banyak. Namun dalam penggunaan sistem manajemen informasi yang dimiliki oleh RS. Artha Medika terdapat kelemahan dan masih terbatas dalam melakukan pengelolaan data rekam medis di rumah sakit yang digunakan dalam pembuatan laporan kepala pimpinan. Maka itu diperlukan suatu sistem yang dapat membantu pihak rumah sakit dalam meningkatkan manajemen informasi kesehatan agar lebih cepat dalam mengelola data dengan melakukan pendekatan menggunakan teknik data mining dengan metode k-means. Sehingga dalam menemukan informasi baru berdasarkan data rekam medis pasien penyakit dalam dapat digunakan dalam proses pengambilan keputusan oleh pihak manajemen rumah sakit agar tepat sasaran sehingga dapat menghasilkan 3 kelompok data yang terdiri dari Usia, Jenis penyakit dan Wilayah. Dari pengujian pada cluster 3 dapat diketahui bahwa hasil dari grup usia (X), jenis penyakit (Y), wilayah (Z) jumlah data yang dimiliki sebanyak 645 data *cluster* 3 berpusat pada *centroid* keterangan banyaknya data rekam medis pasien yaitu usia adalah 44-52 Tahun, dengan jenis penyakit adalah Penyakit ginjal kronis dan wilayah adalah Binjai Selatan.

Kata kunci: Algoritma K-Means, Data Mining, Rekam Medis.

1. LATAR BELAKANG

Manajemen informasi kesehatan di rumah sakit memiliki peran penting dalam meningkatkan kualitas pelayanan kepada pasien. Seiring dengan meningkatnya jumlah pasien, data rekam medis menjadi semakin kompleks dan memerlukan pengelolaan yang lebih efektif. RS Artha Medika Binjai, sebuah lembaga kesehatan yang beroperasi sejak tahun 2012, menghadapi tantangan dalam mengelola data rekam medis pasien secara

efisien, terutama dalam menyediakan informasi yang akurat dan terbaru bagi manajemen rumah sakit.

Penelitian sebelumnya menunjukkan bahwa keterbatasan dalam sistem manajemen informasi yang ada dapat berdampak negatif pada kualitas pelayanan, khususnya ketika data riwayat medis pasien tidak dapat diakses secara cepat dan tepat waktu. Sebagai contoh, studi yang dilakukan oleh Wandana et al. (2020) menggunakan metode K-Means Clustering untuk mengelompokkan data pasien berdasarkan kelas layanan, dan menemukan pola penyebaran penyakit yang berguna untuk pengambilan keputusan manajemen.

Namun, penelitian tersebut masih memiliki keterbatasan dalam hal penerapan teknik clustering yang lebih spesifik pada wilayah tertentu dan jenis penyakit tertentu, terutama dalam konteks rumah sakit di kota Binjai. Kebaruan penelitian ini terletak pada upaya untuk mengisi celah tersebut dengan mengembangkan sistem pengelompokan data rekam medis pasien penyakit dalam berdasarkan wilayah, jenis penyakit, dan usia di RS Artha Medika Binjai. Dengan menggunakan algoritma K-Means, penelitian ini bertujuan untuk meningkatkan manajemen informasi kesehatan di rumah sakit tersebut, sehingga dapat memberikan layanan yang lebih tepat sasaran dan efisien.

Tujuan utama dari penelitian ini adalah untuk mengelompokkan data rekam medis pasien di RS Artha Medika Binjai dengan menggunakan algoritma K-Means Clustering. Dengan demikian, hasil dari pengelompokan ini diharapkan dapat membantu manajemen rumah sakit dalam membuat keputusan yang lebih baik terkait penyebaran penyakit dan alokasi sumber daya, serta meningkatkan kualitas layanan kesehatan di rumah sakit tersebut.

2. KAJIAN TEORITIS

Menurut (Pratama et al., 2023) menyatakan bahwa hasil penelitian ini menggunakan metode *clustering*, Penelitian ini melaksanakan prosedur administrasi klaim BPJS Kesehatan dalam pelayanan rawat inap di RSUD Dr. Soedirman Kebumen didapatkan bahwa prosedur administrasi klaim BPJS dalam pelayanan rawat Inap di RSUD Dr. Soedirman Kebumen berjalan sesuai Alur yang ada, tetapi belum memiliki SPO untuk klaim pembiayaan kesehatan. Perbedaan persepsi kode diagnosis dalam berkas klaim antara pihak verifikator internal rumah sakit dengan pihak verifikator BPJS Kesehatan. Selain itu kekurangan data pendukung sebagai penegakan diagnosis juga mempengaruhi keakuratan kode yang di klaim.

Menurut (Windania et al., 2023) menyatakan bahwa hasil Penelitian ini menyimpulkan hasil analisis bahwa data rekam medis dapat dikelompokkan menjadi 4 cluster dengan 2 variabel, yaitu: jenis kelamin dan jenis perawatan dan hasil analisis ini memperoleh hasil cluster 1 sebanyak 18217 pasien, cluster 2 sebanyak 1016 pasien, cluster 3 sebanyak 396 pasien dan cluster 4 sebanyak 307 pasien yang di rawat inap pada Rumah Sakit Royal Prima Medan.

Menurut (Surya et al., 2023) menyatakan bahwa hasil penelitian ini dengan menerapkan data mining Diketahui hasil dari proses 2 cluster menunjukkan pembagian data ke dalam dua kelompok berdasarkan jarak data terhadap centroid / pusat cluster. Dimana cluster 1 berjumlah 346 data dan cluster 2 berjumlah 547 data. Dari data tersebut dapat diketahui pada cluster 1 berjumlah 346 data dengan pusat diagnosa penyakit Myalgia (Nyeri Otot) dan jenis kelamin laki-laki pada alamat Kebun Lada dan pada cluster 2 berjumlah 547 data dengan pusat diagnosa penyakit Dengue Hemorrhagic Fever (Demam Berdarah) dan jenis kelamin perempuan pada alamat Pahlawan. Kemudian diketahui hasil dari proses 3 cluster menunjukkan pembagian data ke dalam tiga kelompok berdasarkan jarak data terhadap centroid / pusat cluster.

3. METODE PENELITIAN

Dalam permasalahan metode yang digunakan adalah *clustering*, proses awal yang dilakukan dalam pembentukan *cluster* adalah transformasi data kedalam bentuk numeric dengan kode-kode yang telah ditentukan, tentukan jumlah *group* (K), hitung *centroid*, hitung jarak objek ke *centroid* dan kemudian groupkan berdasarkan jarak terdekat, jika tidak ada objek yang pindah *group* maka iterasi selesai.

a. Inisialisasi

Dari data yang ada maka dapat dilakukan inisialisasi data sesuai dengan kebutuhan variabel sebagai berikut:

- 1) Inisialisasi Kriteria Usia
- 2) Inisialisasi Kriteria Jenis Penyakit
- 3) Inisialisasi Kriteria Wilayah

b. Transformasi Data

Agar data diatas dapat diolah dengan menggunakan metode *clustering* algoritma *k-means*, maka data yang berjenis data nominal dan non-nominal seperti Usia, Jenis kelamin, Jenis Penyakit harus diinisialisasikan terlebih dahulu dalam bentuk angka.

Pengelompokkan data penduduk pemilihan umum ini dapat dinyatakan dalam suatu variabel-variabel yang *independen* yaitu Usia (X), Jenis Penyakit (Y), Wilayah (Z).

c. Perhitungan Data Berdasarkan Algoritma *K-Means Clustering*

Jarak *Euclidean* antara dua titik atau objek atau X, Y dan Z yang didefinisikan sebagai berikut:

$$Euclidean (X, Y, Z) = \sqrt{\sum (X1 - X2)^2 + (Y1 - Y2)^2 + (Z1 - Z2)^2}$$

Lakukan *cluster* menjadi 3 kelompok (K=3) dan tentukan titik pusat *centroid*.

Adapun proses perhitungan *clustering* seperti dibawah ini:

Iterasi I

Centroid 1 = (4,2,5) diambil dari secara acak dari data 2

Centroid 2 = (1,3,4) diambil dari secara acak dari data 3

Centroid 3 = (6,4,3) diambil dari secara acak dari data 4

Keterangan :

Penentuan *centroid* awal dilakukan secara random.

Penyelesaian :

Bagian 1 (2,1,2)

K=3, centroid 1 = (4, 2, 5), centroid 2 = (1, 3, 4), centroid 3 = (6, 4, 3)

Jarak dari C1 (X) = $\sqrt{(2 - 4)^2 + (1 - 2)^2 + (2 - 5)^2} = 3.74$

Jarak dari C2 (Y) = $\sqrt{(2 - 1)^2 + (1 - 3)^2 + (2 - 4)^2} = 3$

Jarak dari C3 (Z) = $\sqrt{(2 - 6)^2 + (1 - 4)^2 + (2 - 3)^2} = 5.10$

dst...

Tabel dibawah ini adalah tabel yang menunjukkan hasil penentuan *group* 1 sebagai berikut :

Tabel 1. Hasil Penentuan Group I

No	Usia (X)	Jenis Penyakit (Y)	Wilayah (Z)	Jarak Dari C1	Jarak Dari C2	Jarak Dari C3	Group
1	2	1	2	3.74	3.00	5.10	2
2	4	2	5	0	3.32	3.46	1
3	1	3	4	3.32	0	5.20	2
4	6	4	3	3.46	5.20	0	3

No	Usia (X)	Jenis Penyakit (Y)	Wilayah (Z)	Jarak Dari C1	Jarak Dari C2	Jarak Dari C3	Group
5	1	5	4	4.36	2	5.20	2
6	6	6	2	5.39	6.16	2.24	3
7	3	5	1	5.10	4.12	3.74	3
8	3	7	2	5.92	4.90	4.36	3
9	5	8	5	6.08	6.48	4.58	3
10	6	6	4	4.58	5.83	2.24	3
11	4	9	3	7.28	6.78	5.39	3
12	4	10	4	8.06	7.62	6.40	3
13	3	10	2	8.60	7.55	6.78	3
14	3	3	1	4.24	3.61	3.74	2
15	4	1	2	3.16	4.12	3.74	1
16	1	2	5	3	1.41	5.74	1
17	2	11	4	9.27	8.06	8.12	2
18	1	2	3	3.61	1.41	5.39	2
19	1	1	4	3.32	2	5.92	2
20	2	2	2	3.61	2.45	4.58	2

Group berdasarkan jarak minimal ke *centroid* terdekat yaitu :

Jika jarak terpendek berada di C1 maka data di masukan di kelompok 1

Jika jarak terpendek berada di C2 maka data di masukan di kelompok 2

Jika jarak terpendek berada di C3 maka data di masukan di kelompok 3.

Grup Lama = {0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0}

Group Baru = {2,1,2,3,2,3,3,3,3,3,3,3,3,2,1,1,2,2,2,2}

Terjadi perubahan *group*, maka dilanjutkan ke iterasi berikutnya. Tabel dibawah ini adalah tabel yang menunjukkan hasil penentuan *group* sebagai berikut:

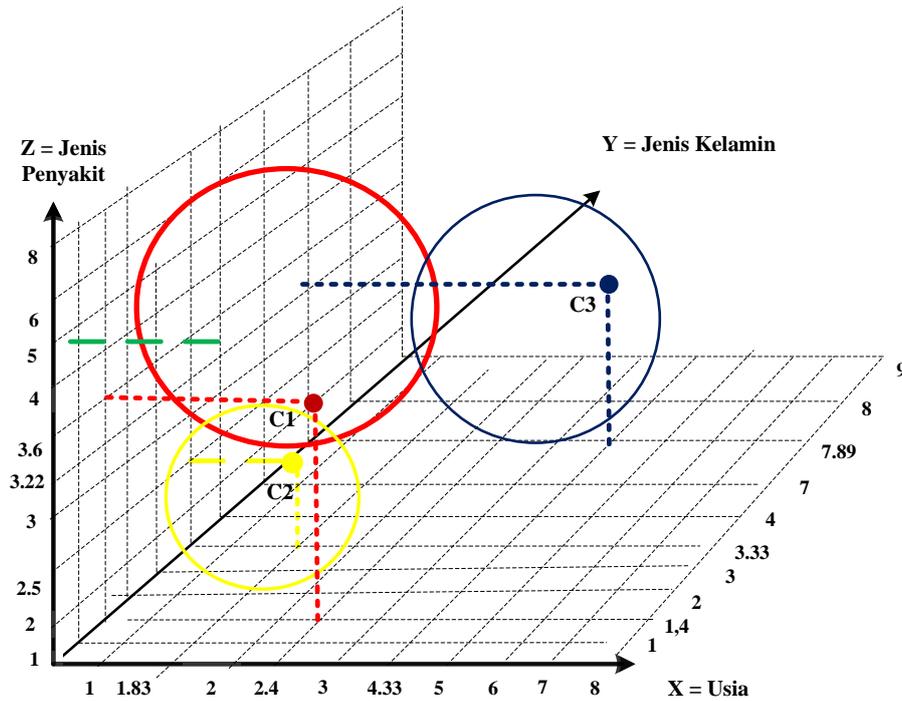
Tabel 2. Hasil Ietrasi III

No	Usia (X)	Jenis Penyakit (Y)	Wilayah (Z)	Jarak Dari C1	Jarak Dari C2	Jarak Dari C3	Group
1	2	1	2	1.70	2.39	7.37	1
2	4	2	5	2.21	3.57	6.16	1
3	1	3	4	2.16	1.75	5.97	2
4	6	4	3	4.48	4.25	4.24	3
5	1	5	4	3.88	2.39	4.48	2
6	6	6	2	6.06	4.98	2.80	3
7	3	5	1	4.48	2.53	3.88	2
8	3	7	2	5.85	3.88	2.01	3
9	5	8	5	7.23	6.17	1.91	3
10	6	6	4	5.85	5.17	2.64	3
11	4	9	3	7.79	6.09	1.18	3
12	4	10	4	8.76	7.17	2.27	3
13	3	10	2	8.77	6.79	2.78	3
14	3	3	1	3.11	1.93	5.53	2
15	4	1	2	2.30	3.22	7.00	1
16	1	2	5	2.07	2.95	7.00	1
17	2	11	4	9.62	7.82	3.96	3
18	1	2	3	1.64	1.65	6.77	2
19	1	1	4	1.51	2.89	7.69	1
20	2	2	2	1.75	1.43	6.45	2

Grup Lama = {1,1,2,3,2,3,2,3,3,3,3,3,3,2,1,1,3,2,1,2}

Group Baru = {1,1,2,3,2,3,2,3,3,3,3,3,3,2,1,1,3,2,1,2}

Dari hasil pembahasan Group Lama dari iterasi 1 didapat juga hasil dari pembahasan group baru seperti yang tertera di atas, Karena pada hasil iterasi ke-1 dan iterasi ke-2 tidak berubah atau terdapat persamaan, maka perhitungan iterasi dihentikan dan mendapatkan hasil sebagai berikut :



Gambar 1. Grafik *Pengclustering* Data Rekam Medis Pasien Penyakit Dalam

Keterangan:

	X	Y	Z
C1 :	2.4	1.4	3.6
C2 :	1.83	3.33	2.5
C3 :	4.33	7.89	3.22

Penjelasan dari hasil diatas ialah :

Dari 20 data terdapat 3 *group* yaitu *group* 1 terdapat 5 data, *group* 2 terdapat 6 data dan *group* 3 terdapat 9 data. Adapun penjelasan dari 3 *group* tersebut sebagai berikut:

1) Cluster 1 Terdapat 5 Data

2.4; 1.4; 3.6

Berdasarkan perhitungan diatas dapat diketahui bahwasannya pada *cluster* 1 data rekam medis pasien penyakit dalam RS Artha Medika Binjai pada *group* Usia (X) adalah 26- 34 Tahun, untuk kelompok Jenis penyakit (Y) ialah Gangguan sel darah merah selain krisis anemia sel sickle (sedang), dan pada *group* wilayah (Z) yaitu kecamatan Binjai Timur.

2) Cluster 2 Terdapat 6 Data

1.83; 3.33; 2.5

Berdasarkan perhitungan diatas dapat diketahui bahwasannya pada *cluster 2* data rekam medis pasien penyakit dalam RS Artha Medika Binjai pada *group* Usia (X) adalah 26- 34 Tahun, untuk kelompok Jenis penyakit (Y) ialah Infeksi viral & non-bakterial lain (sedang), dan pada *group* wilayah (Z) yaitu kecamatan Binjai Selatan.

3) Cluster 3 Terdapat 9 Data

4.33; 7.89; 3.22

Berdasarkan perhitungan diatas dapat diketahui bahwasannya pada *cluster 3* data rekam medis pasien penyakit dalam RS Artha Medika Binjai pada *group* Usia (X) adalah 44-52 Tahun, untuk kelompok Jenis penyakit (Y) ialah Kegagalan jantung (ringan), dan pada *group* wilayah (Z) yaitu kecamatan Binjai Selatan.

4. HASIL DAN PEMBAHASAN

Setelah data dikelompokkan kedalam sebuah *cluster*, maka peneliti akan melakukan pengujian hasil *cluster*, pengujian hasil *cluster* dilakukan untuk memastikan bahwa *cluster* yang tersebut memiliki kualitas yang baik dan representative terhadap data yang ada. Dalam pengujian hasil *cluster* menggunakan *cluster analysis*, yaitu mengukur *variance* atau penyebaran titik data dalam setiap *cluster* yang dimana perhitungannya terdapat $V_{minimum}$, $V_{maksimum}$ dan *cluster variance*. Menghitung *variance* untuk setiap *cluster* menggunakan rumus sebagai berikut:

$$Variance_k = \frac{1}{n_k} \sum_{i=1}^{n_k} (x_i - c_k)$$

Dimana :

n_k adalah jumlah titik data dalam cluster

x_i adalah titik data ke-i dalam cluster

c_k adalah centroid dari cluster

Pengujian Hasil 2 Cluster

a. Cluster 1 (4; 50; 3) dengan 149 data

$$Variance\ C1 = \frac{1}{149} (309.3096+274.7096+242.5896+214.3896+....+ 275.0296) = 212.137$$

$$V_{min} = \min(309.3096+274.7096+242.5896+214.3896+....+ 275.0296) = 1.229$$

$$V_{maks} = \max(309.3096+274.7096+242.5896+214.3896+....+ 275.0296) = 1205.63$$

- b. Cluster 2 (4; 14; 3) dengan 851 data

$$Variance\ C2 = \frac{1}{851} (172.6163+147.8163+128.9963+105.1363+\dots+100.916) = 59.893$$

$$V_{min} = \min(172.6163+147.8163+128.9963+105.1363+\dots+100.916) = 1.076$$

$$V_{maks} = \max(172.6163+147.8163+128.9963+105.1363+\dots+100.916) = 330.676$$

$$Cluster\ Variance = \frac{1}{2} (212.137+59.893) = 136.015$$

Pengujian Hasil 3 Cluster

- c. Cluster 1 (4; 29; 3) dengan 284 data

$$Variance\ C1 = \frac{1}{284} (311+346+385+425+\dots+3202) = 327.981$$

$$V_{min} = \min(311+346+385+425+\dots+3202) = 311$$

$$V_{maks} = \max(311+346+385+425+\dots+3202) = 3202$$

- d. Cluster 2 (5; 63; 3) dengan 71 data

$$Variance\ C2 = \frac{1}{71} (1849+1763+1681+\dots+3850) = 5039.645$$

$$V_{min} = \min(1849+1763+1681+\dots+3850) = 325$$

$$V_{maks} = \max(1849+1763+1681+\dots+3850) = 3850$$

- e. Cluster 3 (4; 11; 3) dengan 645 data

$$Variance\ C2 = \frac{1}{645} (79+66+50+39+28+\dots+45) = 27.880$$

$$V_{min} = \min(79+66+50+39+28+\dots+45) = 0$$

$$V_{maks} = \max(79+66+50+39+28+\dots+45) = 103$$

$$Cluster\ Variance = \frac{1}{3} (327.981+5039.645+27.880) = 1798.502$$

Pengujian Hasil 4 Cluster

- f. Cluster 1 (4; 67; 3) dengan 54 data

$$Variance\ C1 = \frac{1}{54} (176+146+128+149+\dots+336) = 90.521$$

$$V_{min} = \min(176+146+128+149+\dots+336) = 2$$

$$V_{maks} = \max(176+146+128+149+\dots+336) = 336$$

- g. Cluster 2 (4; 21; 3) dengan 359 data

$$Variance\ C2 = \frac{1}{359} (39+28+25+11+12+\dots+22) = 20.818$$

$$V_{min} = \min(39+28+25+11+12+\dots+22) = 0$$

$$V_{maks} = \max(39+28+25+11+12+\dots+22) = 82$$

h. Cluster 3 (4; 40; 3) dengan 103 data

$$Variance C2 = \frac{1}{103} (546+88+70+57+\dots+5) = 45.097$$

$$V_{min} = \min(546+88+70+57+\dots+5) = 1$$

$$V_{maks} = \max(546+88+70+57+\dots+5) = 546$$

i. Cluster 4 (3; 8; 3) dengan 484 data

$$Variance C2 = \frac{1}{484} (1004+655+58+46+\dots+24) = 19.770$$

$$V_{min} = \min(1004+655+58+46+\dots+24) = 1$$

$$V_{maks} = \max(1004+655+58+46+\dots+24) = 1004$$

$$Cluster Variance = \frac{1}{4} (90.521+20.818+45.097+19.770) = 44.052$$

Pengujian Hasil 5 Cluster

j. Cluster 1 (3; 7; 3) dengan 52 data

$$Variance C1 = \frac{1}{52} (2284+2381+2287+2482+\dots+6052) = 3723.125$$

$$V_{min} = \min(2284+2381+2287+2482+\dots+6052) = 2280$$

$$V_{maks} = \max(2284+2381+2287+2482+\dots+6052) = 6052$$

k. Cluster 2 (4; 68; 3) dengan 276 data

$$Variance C2 = \frac{1}{276} (3089+2978+2870+2761+\dots+2559) = 2692.264$$

$$V_{min} = \min(3089+2978+2870+2761+\dots+2559) = 2261$$

$$V_{maks} = \max(3089+2978+2870+2761+\dots+2559) = 3091$$

l. Cluster 3 (4; 42; 3) dengan 158 data

$$Variance C2 = \frac{1}{382} (1679+1598+1525+1442+\dots+1439) = 1245.950$$

$$V_{min} = \min(1679+1598+1525+1442+\dots+1439) = 957$$

$$V_{maks} = \max(1679+1598+1525+1442+\dots+1439) = 2482$$

m. Cluster 4 (4; 25; 3) dengan 147 data

$$Variance C2 = \frac{1}{147} (17+11+15+18+\dots+8) = 26.185$$

$$V_{min} = \min(17+11+15+18+\dots+8) = 1$$

$$V_{maks} = \max(17+11+15+18+\dots+8) = 2381$$

n. Cluster 5 (4; 16; 3) dengan 90 data

$$Variance C2 = \frac{1}{90} (339+376+414+\dots+336) = 745.847$$

$$V_{min} = \min(339+376+414+\dots+336) = 3$$

$$V_{maks} = \max(339+376+414+\dots+336) = 2287$$

$$Cluster Variance = \frac{1}{5} (3723.125+2692.264+1245.950+26.185+745.847) = 1686.674$$

Tabel 3. Pengujian Hasil Cluster

Cluster	Centroid	Variance	V_{min}	V_{maks}	Cluster Variance
2	4; 50; 3	212.137	1	1206	136.015
	4; 14; 3	59.893	1	331	
3	4; 29; 3	327.981	311	3202	1798.502
	5; 63; 3	5039.645	325	3850	
	4; 11; 3	27.880	0	103	
4	4; 67; 3	90.521	2	336	44.052
	4; 21; 3	20.818	0	82	
	4; 40; 3	45.097	1	546	
	3; 8; 3	19.770	1	1004	
5	3; 7; 3	3723.125	2280	6052	1686.674
	4; 68; 3	2692.264	2261	3091	
	4; 42; 3	1245.950	957	2482	
	4; 25; 3	26.185	1	2381	
	4; 16; 3	745.847	3	2287	

Dari tabel tersebut dapat dijelaskan bawah:

a. Pada 2 cluster

- 1) Cluster 1 memiliki nilai *variance* 212.137 yang menunjukkan bahwa data dalam *cluster* ini jauh tersebar dibandingkan dengan cluster 2, dan memiliki nilai V_{min} 1 yang menunjukkan bahwa data dekat dengan pusat cluster, namun masih belum kompak, dan memiliki nilai V_{maks} 1206 menunjukkan bahwa data yang terbesar jauh lebih banyak dibandingkan dengan cluster 2.
- 2) Cluster 2 memiliki nilai *variance* 59.893 yang lebih sedikit tersebar dibandingkan dengan *cluster* 1 dan memiliki nilai V_{min} 1 yang menunjukkan bahwa data dekat dengan pusat cluster, serta memiliki nilai V_{maks} 331 menunjukkan data yang tersebar lebih sedikit dibandingkan *cluster* 1.

Pada pengujian hasil 2 *cluster* yang memiliki nilai *cluster variance* (rata-rata dari semua cluster) 136.015 menunjukkan bahwa secara keseluruhan pengelompokan dengan 2 *cluster* ini cukup kompak, dengan variasi yang terdistribusi diantara *cluster-cluster* tersebut.

b. Pada 3 cluster

- 1) Cluster 1 memiliki nilai *variance* 327.981 yang menunjukkan bahwa data jauh lebih sedikit tersebar dibanding dengan cluster 2, serta memiliki nilai V_{min} 311 yang menunjukkan jauh dengan pusat cluster, serta memiliki nilai V_{maks} 3220 menunjukkan adanya variasi yang signifikan.
- 2) Cluster 2 memiliki nilai *variance* 5039.645 yang menunjukkan bahwa data dalam *cluster* ini jauh tersebar dibanding cluster lainnya, dan memiliki nilai V_{min} 325

menunjukkan bahwa data sangat jauh dengan pusat cluster, serta memiliki nilai V_{maks} 3850 menunjukkan adanya variasi yang signifikan.

- 3) Cluster 3 memiliki nilai *variance* 27.790 yang lebih kompak dibandingkan dengan cluster lainnya dan memiliki nilai V_{min} 0 yang menunjukkan bahwa data sangat dekat dengan pusat cluster, serta memiliki nilai V_{maks} 103 menunjukkan data yang tersebar jauh lebih banyak dibandingkan *cluster* 1.

Pada pengujian hasil 3 *cluster* yang memiliki nilai *cluster variance* (rata-rata dari semua cluster) 1798.502 menunjukkan bahwa secara keseluruhan pengelompokan dengan 3 *cluster* ini cukup kompak dan signifikan.

c. Pada 4 cluster

- 1) Cluster 1 memiliki nilai *variance* 90.521 menunjukkan bahwa cluster tersebut mungkin tidak sangat kompak dan memiliki penyebaran yang cukup besar, dan memiliki nilai V_{min} 2 menunjukkan bahwa cluster tersebut memiliki variasi yang relatif kecil di dalamnya, dengan titik data yang cukup dekat dengan pusat cluster, serta memiliki nilai V_{maks} 336 mengandung data yang bervariasi secara signifikan.
- 2) Cluster 2 memiliki nilai *variance* 20.818 menunjukkan bahwa cluster ini cukup kompak, tetapi masih memiliki beberapa variasi di dalamnya, dengan nilai V_{min} 0 yang menunjukkan bahwa semua titik data dalam cluster berada pada posisi yang sangat dekat atau identik dengan pusat cluster, serta memiliki nilai V_{maks} 82 menunjukkan penyebaran yang relatif kecil.
- 3) Cluster 3 memiliki nilai *variance* 45.097 menunjukkan bahwa cluster ini mungkin mengandung variasi yang signifikan, dengan nilai V_{min} 1 yang menunjukkan bahwa titik data yang berada relatif dekat satu sama lain, serta memiliki nilai V_{maks} 546 yang menunjukkan data cluster ini memiliki penyebaran data yang besar relative.
- 4) Cluster 4 memiliki nilai *variance* 19.770 menunjukkan bahwa cluster ini memiliki keseimbangan yang baik antara kekompakan dan variasi, dengan nilai V_{min} 1 yang menunjukkan bahwa titik data yang berada relatif dekat satu sama lain, serta memiliki nilai V_{maks} 1004 yang menunjukkan bahwa data dalam cluster ini bervariasi secara signifikan

Pada pengujian hasil 4 *cluster* yang memiliki nilai *cluster variance* (rata-rata dari semua cluster) 44.052 menunjukkan bahwa bahwa titik data dalam cluster ini cukup tersebar dari pusat cluster dan secara keseluruhan pengelompokan dengan 4 *cluster* ini mengandung variasi yang signifikan dan tidak kompak.

d. Pada 5 cluster

- 1) Cluster 1 memiliki nilai *variance* 3723.125 menunjukkan bahwa ada penyebaran yang sangat luas dalam *cluster* ini, dan memiliki nilai V_{min} 2280 menunjukkan beberapa data jauh tersebar dari pusat cluster, namun masih belum kompak, serta memiliki nilai V_{maks} 6052 menunjukkan adanya variasi yang signifikan.
- 2) Cluster 2 memiliki nilai *variance* 2662.264 menunjukkan bahwa cluster ini mengalami penyebaran data yang cukup besar, dan memiliki nilai V_{min} 2261 yang menunjukkan bahwa data jauh tersebar dari pusat cluster, serta memiliki nilai V_{maks} 3091 menunjukkan adanya variasi yang signifikan.
- 3) Cluster 3 memiliki nilai *variance* 1245.950 yang menunjukkan bahwa data dalam *cluster* ini tersebar secara signifikan dan memiliki variasi yang cukup besar dalam data cluster, dengan nilai V_{min} 957 yang menunjukkan bahwa data sangat dekat dengan pusat cluster, serta memiliki nilai V_{maks} 2482 yang menunjukkan data yang tersebar sangat jauh dibandingkan *cluster* 1.
- 4) Cluster 4 memiliki nilai *variance* 26.185 yang menunjukkan bahwa data dalam *cluster* ini menunjukkan bahwa ada penyebaran data yang relatif kecil, dan memiliki nilai V_{min} 1 yang menunjukkan bahwa titik data yang berada relatif dekat satu sama lain, serta memiliki nilai V_{maks} 2381 yang menunjukkan data yang tersebar sangat jauh dari pusat *cluster* sehingga adanya variasi yang sangat besar.
- 5) Cluster 5 memiliki nilai *variance* 745.847 yang menunjukkan bahwa ada penyebaran data yang cukup besar di dalam cluster ini, memiliki nilai V_{min} 3 yang menunjukkan bahwa data dekat dari pusat cluster, serta memiliki nilai V_{maks} 2287 yang menunjukkan data yang tersebar sangat jauh dari pusat *cluster* sehingga adanya variasi yang sangat besar.

Pada pengujian hasil 5 *cluster* yang memiliki nilai *cluster variance* (rata-rata dari semua cluster) 1686.674 menunjukkan bahwa secara keseluruhan pengelompokan dengan 5 *cluster* ini menunjukkan bahwa data tidak kompak dan mengandung variasi yang signifikan.

KESIMPULAN DAN SARAN

Berdasarkan hasil analisis dan pengujian yang telah dilakukan, penelitian ini berhasil mengidentifikasi pengelompokan data rekam medis pasien dengan menggunakan metode *clustering*. Hasil penelitian menunjukkan bahwa pengelompokan data rekam medis ke dalam beberapa cluster menghasilkan variasi yang berbeda-beda. Pengujian pada cluster 3, 4, dan 5 menunjukkan bahwa data pasien memiliki distribusi yang signifikan dengan variasi yang cukup

besar, terutama pada cluster 4 dan 5 yang menunjukkan adanya penyebaran data yang lebih luas dibandingkan dengan cluster lainnya. Hal ini menunjukkan bahwa metode clustering yang digunakan mampu mengelompokkan data dengan cukup baik, meskipun terdapat perbedaan dalam kekompakan dan variasi antar cluster.

Penelitian ini memberikan gambaran bahwa pengelompokan data rekam medis pasien berdasarkan usia, jenis penyakit, dan wilayah dapat membantu dalam memahami pola distribusi penyakit di wilayah tertentu, serta memberikan informasi yang berguna bagi pengelolaan data rekam medis yang lebih efektif. Namun, perlu diingat bahwa generalisasi hasil penelitian ini harus dilakukan dengan hati-hati mengingat adanya variasi yang signifikan antar cluster.

Sebagai saran, penelitian ini dapat dikembangkan lebih lanjut dengan mengintegrasikan metode clustering ini ke dalam sebuah sistem yang memungkinkan perhitungan dan visualisasi data secara langsung dalam platform yang lebih user-friendly, seperti MATLAB. Selain itu, untuk mendapatkan hasil yang lebih komprehensif, disarankan untuk melakukan penelitian lanjutan dengan menggunakan metode clustering lain yang mungkin lebih efektif dalam mengelompokkan data rekam medis dengan tingkat variasi yang berbeda. Penelitian juga dapat diperluas dengan mempertimbangkan faktor-faktor lain seperti tingkat keparahan penyakit atau riwayat medis pasien, untuk meningkatkan akurasi dan relevansi pengelompokan data.

DAFTAR REFERENSI

- Ali, A. (2019). 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*, 19(1), 186–195. <https://doi.org/10.30812/matrik.v19i1.529>
- Atina, A. (2019). Aplikasi Matlab pada Teknologi Pencitraan Medis. *Jurnal Penelitian Fisika Dan Terapannya (JUPITER)*, 1(1), 28.
- Bella, M. W., Wahyuning, A. R., & Kahar, N. (2023). Analisis Data Rekam Medis Pasien Berdasarkan Usia Di Puskesmas Sekernan Ilir Jambi. *Seminar Nasional Informatika (SENATIKA)*, 4(1), 381–389.
- Benri, M., Metisen, H., & Latipa, S. (2015). Analisis Clustering Menggunakan Metode K-Means Dalam Pengelompokanpenjualan Produk Pada Swalayan FADHILA. *Jurnal Media Infotama*, 11(2), 110–118.
- Eko Prasetyo. 2012. *Data Mining Konsep dan Aplikasi menggunakan Matlab*. Yogyakarta : CV ANDI OFFSET.
- Emnita, G., Zarli, M., Salim., A. H. (2022). Analisis dan Perancangan Perangkat Lunak Pengelolaan Pasien Askes pada Rumah Sakit Artha Medica Binjai. [repositori.usu.ac.id,1-92.https://repositori.usu.ac.id/handle/123456789/73722](https://repositori.usu.ac.id/handle/123456789/73722)
- Haryati, S., Sudarsono, A., & Suryana, E. (2015). Implementasi Data Mining Untuk

- Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu). *Jurnal Media Infotama*, 11(2), 130–138.
- Khairani, A. F., Nazir, A., Darmizal, T., Vitriani, Y., & Yusra, Y. (2023). Klasterisasi Peserta BPJS Berdasarkan Rekam Medis Menggunakan Algoritma K-Means. *Journal of Computer System and Informatics (JoSYC)*, 4(3), 625–631. <https://doi.org/10.47065/josyc.v4i3.3442>
- Kristyawan, Y., & Al Hakim, L. (2020). Implementasi Case-Based Reasoning Pada Sistem Pakar Diagnosis Penyakit Dalam. *Jurnal Sistem Informasi Dan Bisnis Cerdas (SIBC)*, 13(2), 99–106.
- Mardia, dkk. 2021. Sistem Informasi Akuntansi dan Bisnis. Medan: Yayasan Kita Menulis.
- Ordila, R., Wahyuni, R., Irawan, Y., & Yulia, S. M. (2020). Penerapan Data Mining Untuk Pengelompokan Data Rekam Medis Pasien Berdasarkan Jenis Penyakit Dengan Algoritma Clustering (Studi Kasus : Poli Klinik PT.Inecda). *Jurnal Ilmu Komputer*, 9(2), 148–153. <https://doi.org/10.33060/jik/2020/vol9.iss2.181>
- Poernomo, B., & Dewi, R. I. sari. (2017). Penerapan Data Mining Untuk Prakiraan Cuaca Di Kota Malang Menggunakan Algoritma Iterative Dichotomiser Tree (ID3). *JOUTICLA*, 3(2), 101–108.
- Pratama, A., Fauzi, H., Indira, Z., & Purnama Adi, P. (2023). Analisis Faktor Penyebab Pending Klaim Rawat Inap Akibat Koding Rekam Medis Di Rumah Sakit Umum Daerah (RSUD) Dr. Soedirman Kebumen Corresponding Author. *Jurnal Ilmiah Perekam Dan Informasi Kesehatan Imelda*, 8(1), 124–134. <http://jurnal.uimedan.ac.id/index.php/JIPIKI>124Journalhomepage:<http://jurnal.uimedan.ac.id/index.php/JIPIKI>
- Prayoga, Y., Mahmudi, A., & Zulfia Zahro, H. (2021). Penerapan Metode K-Means Pada Sistem Informasi Akademik Untuk Pengelompokkan Siswa Berprestasi Di Upt Sma Negeri 3 Kota Pasuruan Berbasis WEB. *Jurnal Mahasiswa Teknik Informatika*, 5(2), 822–828.
- Roni Setiawan “ Penerapan Data Mining Menggunakan Algoritma k means Clustering untuk Menentukan Strategi Promosi Mahasiswa Baru “ *Jurnal Lentera ICT* , Vol.3 No.1, Mei 2016.
- Surya, N. W., Akim, P. M. H., & Magdalena, S. (2023). Analisis Data Mining Untuk Klasterisasi Data Rekam Medis Menggunakan Algoritma K-Means Pada Rumah Sakit Sylvani Binjai. *Indonesian Journal of Education And Computer Science*, 1(3), 82–88.
- Wandana, J., Defit, S., & Sumijan. (2020). Klasterisasi Data Rekam Medis Pasien Pengguna Layanan BPJS Kesehatan Menggunakan Metode K-Means. *Jurnal Informasi Dan Teknologi*, 2(4), 119–125. <https://doi.org/10.37034/jidt.v2i4.73>
- Windania, P., Armando, S., & Saputra, A. (2023). Penerapan Data Mining Untuk Pengelolaan Data Rekam Medis Menggunakan Metode K-Means Clustering Pada Rumah Sakit Royal Prima Medan. *Jurnal Tekinkom*, 6(1), 158–168. <https://doi.org/10.37600/tekinkom.v6i1.857>.