

Klasterisasi Rumah Sakit berdasarkan Kunjungan Pasien menggunakan Algoritma K-Means: Data 2019-2023

Liga Mayola^a, M. Hafizh^b, Hadi Syahputra^c

^aSistem Informasi, Fakultas Ilmu Komputer, Universitas Putra Indonesia YPTK Padang

^bTeknik Informatika, Fakultas Ilmu Komputer, Universitas Putra Indonesia YPTK Padang

Submitted: 2-12-2024, Reviewed: 15-12-2024, Accepted 26-12-2024

<https://doi.org/10.47233/jteksis.v7i1.1703>

Abstract

Hospitals play a vital role in public health services. Every day, patients visit hospitals to receive medical care. Over time, patient visit data will continue to accumulate, creating a massive volume of data that requires strategic management. This situation necessitates an in-depth analysis to understand patient visit patterns, aiding the effective management of hospital resources. This study aims to classify hospitals based on the number of patient visits over the past five years, providing strategic insights for resource management, facility planning, and improving the quality of healthcare services. The hospital classification process utilizes the K-Means Clustering Algorithm. The research data was obtained from the Satu Data Sumbar website. The patient visit data is categorized into three clusters: Cluster 1 (K1) represents hospitals with very high visit intensity, Cluster 2 (K2) represents hospitals with moderate visit levels, and Cluster 3 (K3) represents hospitals with low visit levels. These findings provide valuable insights for policymakers to optimize healthcare services.

Keywords: Data Mining, K-Means, Clustering, Patient Visits

Abstrak

Rumah sakit memiliki peran vital dalam layanan kesehatan masyarakat. Setiap hari selalu ada pasien yang mengunjungi rumah sakit untuk mendapatkan layanan kesehatan. Data kunjungan pasien seiring waktu akan terus bertumbuh dan akan menghasilkan gunung data dan memerlukan pengelolaan strategis. Hal ini menuntut analisis mendalam untuk memahami pola kunjungan pasien dalam rangka membantu pengelolaan sumber daya rumah sakit. Penelitian ini bertujuan untuk mengelompokkan rumah sakit berdasarkan jumlah kunjungan pasien selama 5 tahun terakhir guna memberikan wawasan strategis untuk pengelolaan sumber daya, perencanaan fasilitas, dan peningkatan kualitas pelayanan kesehatan. Pengelompokkan rumah sakit ini menggunakan Metode Clustering Algoritma K-Means. Sumber data penelitian diperoleh dari website Satu Data Sumbar. Data kunjungan pasien rumah sakit ini akan dikelompokkan menjadi 3 cluster. Hasil penelitian menunjukkan bahwa Cluster 1 (K1) merepresentasikan rumah sakit dengan intensitas kunjungan yang sangat tinggi. Cluster 2 (K2) menggambarkan rumah sakit dengan tingkat kunjungan menengah, sementara Cluster 3 (K3) memiliki kunjungan yang rendah. Temuan ini memberikan wawasan penting bagi pengambil kebijakan untuk optimalisasi pelayanan kesehatan.

Keywords: Data Mining, K-Means, Clustering, Kunjungan, Pasien

This work is licensed under Creative Commons Attribution License 4.0 CC-BY International license



PENDAHULUAN

Rumah sakit memiliki peran vital dalam layanan kesehatan masyarakat. Setiap hari selalu ada pasien yang mengunjungi rumah sakit untuk mendapatkan layanan kesehatan. Kunjungan pasien di setiap rumah sakit berbeda-beda. Hal ini dapat disebabkan oleh kapasitas, fasilitas, layanan dan sumber daya rumah sakit tersebut. Data kunjungan pasien seiring waktu akan terus bertumbuh dan akan menghasilkan gunung data.

Gunungan data jika dibiarkan begitu saja akan menjadi sampah data. Namun jika diolah dengan teknik data mining dari gunung data tersebut dapat ditemukan *knowledge* atau informasi yang berharga [1]. Data mining sebagai sebuah proses ekstraksi informasi baru dari sejumlah besar data yang dapat berguna dalam proses pengambilan keputusan [2].

Gunungan data kunjungan pasien dapat diklasterisasi dengan algoritma data mining sehingga dapat memberikan wawasan strategis untuk pengelolaan sumber daya, perencanaan fasilitas, dan peningkatan kualitas pelayanan kesehatan.

Gunungan data ini dapat diolah dengan teknik data mining Algoritma K-Means. Dimana K-Means dapat digunakan untuk mengelompokkan atau mengklasterisasi rumah sakit berdasarkan intensitas kunjungan. Rumah sakit dapat dikelompokkan sebagai rumah sakit dengan kunjungan yang tinggi, sedang dan rendah. Hasil klasterisasi dapat digunakan sebagai dasar pengambilan kebijakan. Seperti rumah sakit dengan kunjungan tinggi yang mungkin memerlukan pengembangan kapasitas, atau rumah sakit dengan kunjungan rendah yang mungkin menghadapi tantangan daya tarik layanan.

Selain itu juga dapat memberikan wawasan kepada manajemen rumah sakit tentang distribusi kunjungan. Hasil klasterisasi rumah sakit dapat menjadi landasan untuk penelitian lebih lanjut atau pengembangan kebijakan. Klasterisasi ini dapat dihubungkan dengan data lain seperti ketersediaan fasilitas, kepuasan pasien, atau demografi wilayah, untuk menghasilkan kebijakan yang lebih terarah.

Penelitian terdahulu yang menggunakan Algoritma K-Means untuk klasterisasi adalah [3] hasil penelitiannya dengan Algoritma K-Means dapat mengelompokkan rumah sakit berdasarkan kunjungan pasien dengan nilai kelompok optimal adalah tiga.

Penelitian [4] berhasil mengimplementasikan Algoritma K-Means Clustering untuk mengelompokkan jenis penyakit pasien di Puskesmas Pulo Brayon berdasarkan atribut seperti usia, jenis kelamin, dan diagnosis penyakit.

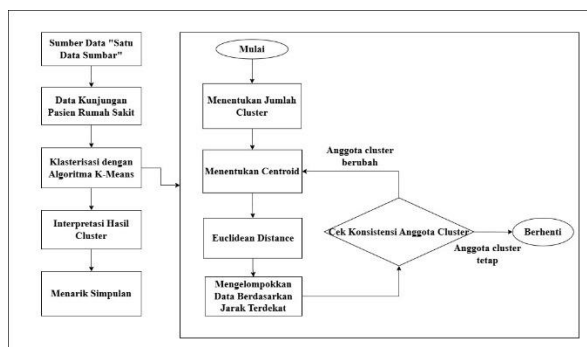
Selanjutnya penelitian [5] berhasil menggunakan Algoritma K-Means Clustering untuk mengelompokkan daerah di Kabupaten Asahan berdasarkan penggunaan air bersih menjadi tiga kategori: tinggi, sedang, dan rendah.

Berdasarkan uraian di atas maka penelitian ini bertujuan untuk menggunakan Algoritma K-Means untuk mengklasterisasi kunjungan pasien rumah sakit.

METODE PENELITIAN

2.1. Metode

Metode penelitian dapat dilihat pada Gambar 2.1. Penelitian ini menggunakan metode clustering. Clustering adalah suatu metode pengelompokan berdasarkan ukuran kedekatan atau kemiripan [6][7]. Clustering atau klasterisasi data kunjungan pasien diekstraksi dengan menggunakan Algoritma K-Means clustering sehingga rumah sakit dapat dikelompokkan berdasarkan kunjungan pasien. Algoritma K-Means umumnya merupakan metode klasterisasi yang paling dikenal dan paling sering digunakan [8].



Gambar 2.1 Kerangka Kerja Penelitian

Gambar 2.1 menggambarkan kerangka tahapan penelitian yang dilakukan. Uraian kerangka kerja penelitian secara detail dijelaskan sebagai berikut:

1. Sumber Data “Satu Data Sumbar”

Data yang digunakan bersumber dari website Satu Data Sumbar <https://data.sumbarprov.go.id/dataset>.

2. Data Kunjungan Pasien Rumah Sakit Data ini terdiri dari 24 rumah sakit di Kota Padang dengan jumlah kunjungan selama 5 tahun dari tahun 2019 hingga tahun 2023.

3. Klasterisasi dengan Algoritma K-Means

Algoritma K-Means adalah salah satu algoritma data mining merupakan metode non-hierarki [9] yang dapat digunakan untuk mengelompokkan data [10]. K-Means clustering adalah metode pengelompokan data [11], yang memisahkan data menjadi beberapa kelompok (cluster) berdasarkan kemiripan tertentu [12]. Algoritma ini memberikan hasil yang efektif dalam menyelesaikan masalah klasterisasi [13]. Langkah-langkah Algoritma K-Means diuraikan sebagai berikut [14] :

a. Menentukan Jumlah Cluster

Menentukan jumlah cluster adalah langkah awal untuk proses klasterisasi. Jumlah cluster yang ditentukan harus disesuaikan dengan tujuan analisis. Pada penelitian ini jumlah clusternya adalah 3 cluster.

b. Menentukan Centroid

Centroid adalah titik pusat cluster. Oleh karena itu, jumlah centroid harus mengikuti jumlah cluster yang ditentukan. Tahap awal atau iterasi ke-nol penetapan centroid dipilih secara acak dari data yang ada. Perhitungan centroid baru untuk iterasi ke-satu setiap cluster dengan menghitung rata-rata nilai dari data yang terkelompok.

c. Euclidean Distance

Euclidean Distance pada Algoritma K-Means digunakan untuk menghitung jarak antara setiap titik data dengan centroid cluster. Jarak ini menentukan data akan dimasukkan ke cluster yang mana. Formula Euclidean Distance [15] pada Algoritma K-Means dapat dilihat pada Formula (1).

$$E_{ij} = \sqrt{(x_i - s_i)^2 + (x_j - s_j)^2} \quad (1)$$

Dimana E_{ij} merupakan jarak objek antara objek i dan j . s_{ij} juga merupakan pusat cluster. s_{ij} merupakan objek i atau j .

d. Mengelompokkan Data Berdasarkan Jarak Terdekat

Mengelompokkan data berdasarkan jarak terdekat adalah inti dari Algoritma K-Means Clustering. Proses ini menentukan cluster untuk setiap data dengan menghitung jaraknya ke

centroid, lalu menempatkan data ke cluster dengan jarak terkecil.

e. Cek Konsistensi Anggota Cluster

Mengecek konsistensi anggota cluster merupakan langkah penilaian stabilitas anggota kelompok. Jika ada anggota cluster yang mengalami perubahan maka dilakukan perulangan. Jika anggota cluster tetap maka proses dihentikan.

4. Interpretasi Hasil Cluster

Interpretasi hasil cluster adalah proses menganalisis dan memberikan makna pada kelompok-kelompok (cluster) yang terbentuk setelah data dianalisis menggunakan Algoritma K-Means.

5. Menarik Simpulan

Menarik simpulan adalah proses menyusun pernyataan yang padat berdasarkan hasil penelitian untuk menjawab tujuan dan pertanyaan penelitian.

HASIL DAN PEMBAHASAN

Metode yang digunakan untuk mengelompokkan rumah sakit berdasarkan kunjungan pasien adalah Algoritma K-Means. Analisis algoritma ini dimulai dengan menentukan jumlah kluster atau kelompok. Kemudian menentukan centroid (titik pusat) masing-masing kelompok, jumlah centroid sesuai dengan jumlah kluster yang ditentukan. Selanjutnya menghitung jarak antara titik pusat dengan setiap data menggunakan formula Euclidean Distance (1). Berikutnya membandingkan hasil perhitungan jarak antara setiap data dengan setiap titik pusat dan dipilih jarak yang terdekat dengan titik pusat. Proses ini merupakan cara untuk mengelompokkan data ke dalam kelompok tertentu. Kemudian untuk menguji apakah anggota kelompok bersifat tetap dan tidak ada perubahan maka iterasi diulangi. Jika anggota kelompok tetap maka proses dihentikan, dan sebaliknya ulangi iterasi hingga anggota kelompok bersifat tetap.

Tabel 1. Data Kunjungan Rumah Sakit

No	Nama Rumah Sakit	Tahun				
		2019	2020	2021	2022	2023
1	RSUP dr. M.D	28861	20861	22559	28513	32491
2	RSUD dr. R	5487	2548	2898	4677	4750
3	...					
24	RSJ HBS	2227	2114	2356	2538	2590

Proses pengolahan data Tabel 1 dengan Algoritma K-Means diuraikan sebagai berikut:

1. Menentukan jumlah cluster atau kelompok data. Pada penelitian ini data akan dikelompokkan menjadi 3 kelompok dengan rincian: rumah sakit dengan kunjungan rendah, sedang dan tinggi.
2. Menentukan titik pusat atau centroid masing-masing kluster. Jumlah centroid mengikuti jumlah kluster. Penentuan centroid awal dipilih secara acak. Centroid yang dipilih secara acak terlihat dalam Tabel 2.

Tabel 2. Centroid Cluster Iterasi 0

Centroid	Nama Rumah Sakit	Tahun				
		2019	2020	2021	2022	2023
K1	RSUP dr.M.D	28861	20861	22559	28513	32491
K2	RSI IS	6658	4986	5722	6577	6732
K3	RSIA RI	3133	5863	3436	2734	1777

K1, K2, K3 pada Tabel 2 menerangkan centroid Cluster 1 (K1), Cluster 2 (K2) dan Cluster 3 (K3). K1 dipilih jumlah kunjungan dari RSUP dr. M.D, K2 dipilih jumlah kunjungan RSI IS dan K3 dipilih kunjungan RSIA RI dengan angka kunjungan tertera pada tahun 2019 hingga 2023.

3. Menghitung jarak antara titik data dengan titik pusat dengan formula *Euclidean Distance*. Proses perhitungan jarak dihitung dari titik pusat ke semua data. Uraian perhitungan sebagai berikut:

$$\begin{aligned}
 &1. \text{ Perhitungan jarak antara K1, K2 dan K3 terhadap data ke-1} \\
 &K1 = \sqrt{(28861 - 28861)^2 + (20861 - 20861)^2 + (22559 - 22559)^2 + (28513 - 28513)^2 + (32491 - 32491)^2} \\
 &= 0
 \end{aligned}$$

$$\begin{aligned}
 &K2 = \sqrt{(6658 - 28861)^2 + (4986 - 20861)^2 + (5722 - 22559)^2 + (6577 - 28513)^2 + (6732 - 32491)^2} \\
 &= 46617
 \end{aligned}$$

$$\begin{aligned}
 &K3 = \sqrt{(3133 - 28861)^2 + (5863 - 20861)^2 + (3436 - 22559)^2 + (2734 - 28513)^2 + (1777 - 32491)^2} \\
 &= 53483
 \end{aligned}$$

2. Perhitungan jarak antara K1, K2 dan K3 terhadap data ke-2

$$K1 = \sqrt{(28861 - 5487)^2 + (20861 - 2548)^2 + (22559 - 2898)^2 + (28513 - 4677)^2 + (32491 - 4750)^2} = 51049$$

$$K2 = \sqrt{(6658 - 5487)^2 + (4986 - 2548)^2 + (5722 - 2898)^2 + (6577 - 4677)^2 + (6732 - 4750)^2} = 4778$$

$$K3 = \sqrt{(3133 - 5487)^2 + (5863 - 2548)^2 + (3436 - 2898)^2 + (2734 - 4677)^2 + (1777 - 4750)^2} = 5425$$

3. Perhitungan jarak antara K1, K2 dan K3 terhadap data ke-24

$$K1 = \sqrt{(28861 - 2227)^2 + (20861 - 2114)^2 + (22559 - 2356)^2 + (28513 - 2538)^2 + (32491 - 2590)^2} = 55116$$

$$K2 = \sqrt{(6658 - 2227)^2 + (4986 - 2114)^2 + (5722 - 2356)^2 + (6577 - 2538)^2 + (6732 - 2590)^2} = 8525$$

$$K3 = \sqrt{(3133 - 2227)^2 + (5863 - 2114)^2 + (3436 - 2356)^2 + (2734 - 2538)^2 + (1777 - 2590)^2} = 4091$$

4. Mengelompokkan objek dengan cara membandingkan nilai jarak antara objek terhadap masing-masing titik pusat cluster. Nilai jarak yang dipilih adalah nilai jarak terdekat atau dengan nilai jarak terkecil. Nilai perhitungan jarak antara data dengan semua titik pusat dapat dilihat pada Tabel 3.

Tabel 3. Nilai perhitungan jarak antara data dan titik pusat cluster

No	NK1	NK2	NK3	NMIN
1	0	46617	53483	0
2	51049	4778	5425	4778
3	37096	9766	17062	9766
4	52255	11911	8217	8217
5	42636	4487	11507	4487
6	41278	5411	12690	5411
7	42799	6135	13287	6135
8	55584	8980	4099	4099
9	46617	0	7599	0
10	59733	13150	7506	7506
11	48063	5870	10215	5870
12	53980	7387	3336	3336

13	56558	10097	5456	5456
14	54443	7858	3300	3300
15	53483	7599	0	0
16	60014	13435	7825	7825
17	55505	9126	4570	4570
18	56790	10288	5580	5580
19	59924	13345	7771	7771
20	58365	11786	6374	6374
21	54727	8145	3687	3687
22	51435	5177	5158	5158
23	49674	7100	9565	7100
24	55116	8525	4092	4092

Nilai yang diwarnai pada Tabel 3 menunjukkan nilai jarak terkecil. NK1, NK2, NK3 merupakan kode untuk nilai jarak K1, K2, K3 terhadap data. NMIN merupakan kolom alat bantu untuk membandingkan nilai jarak terendah.

Tabel 3 menerangkan bahwa anggota Cluster 1 berjumlah 1 data. Anggota Cluster 2 berjumlah 8 data. Anggota Cluster 3 berjumlah 15 data. Hingga proses ini disebut dengan iterasi ke-nol.

5. Iterasi kedua harus dilakukan untuk menguji apakah anggota kelompok sudah konsisten atau masih ada perubahan. Iterasi kedua dimulai dari penentuan centroid cluster yang telah terbentuk pada iterasi pertama. Penentuan centroid baru dihitung berdasarkan nilai rata-rata anggota kelompok yang terbentuk.

$$K1 = (28861, 20861, 22559, 28513, 32491)$$

$$K2 = \frac{(5487+10554+10242+8791+6899+6658+9099+511)/8, (2548+7423+6247+6511+3402+4986+1447+2493)/8, (2898+8456+6596+8069+6607+5722+2140+4652)/8, (4677+11206+7624+9570+9449+6577+5464+7312)/8, (4750+13466+8692+9549+11835+6732+8111+8906)/8}{8} = (7280, 4382, 5643, 7735, 9005)$$

$$K3 = \frac{(3043+2688+829+3241+405+2785+3133+188+2700+1455+240+1302+2923+3556+2227)/15, (13749+1933+317+2789+1364+2844+5863+158+2216+1217+126+837+2508+2798+2114)/15, (1555+2456+232+2855+2190+2571+3436+158+2451+1467+240+712+2198+3657+2356)/15, ...}$$

$$\begin{aligned} & (2074+1925+111+2656+2310+2565+150+572+1 \\ & 217+191+795+2474+3909+7312+2538)/15, \\ & (2941+1873+10+2927+2368+2715+1777+170+3 \\ & 102+2521+218+861+2610+5727+2590)/15 \\ & = (2048,2726,1902,1748,2161) \end{aligned}$$

Centroid cluster yang baru dapat dilihat pada Tabel 4.

Tabel 4. Centroid Cluster Iterasi 1

Centroid	Tahun				
	2019	2020	2021	2022	2023
K1	28861	20861	22559	28513	32491
K2	7280	4382	5643	7735	9005
K3	2048	2726	1902	1748	2161

6. Melakukan perulangan langkah ketiga dengan titik pusat cluster baru Tabel 4. Uraian perhitungan sebagai berikut:

- Perhitungan jarak antara K1, K2 dan K3 terhadap data ke-1

$$\begin{aligned} K1 &= \sqrt{(28861 - 28861)^2 + (20861 - 20861)^2 + (22559 - 22559)^2 + (28513 - 28513)^2 + (32491 - 32491)^2} \\ &= 0 \end{aligned}$$

$$\begin{aligned} K2 &= \sqrt{(7280 - 28861)^2 + (4382 - 20861)^2 + (5643 - 22559)^2 + (7735 - 28513)^2 + (9005 - 32491)^2} \\ &= 44797 \end{aligned}$$

$$\begin{aligned} K3 &= \sqrt{(2048 - 28861)^2 + (2726 - 20861)^2 + (1902 - 22559)^2 + (1748 - 28513)^2 + (2161 - 32491)^2} \\ &= 55775 \end{aligned}$$

- Perhitungan jarak antara K1, K2 dan K3 terhadap data ke-2

$$\begin{aligned} K1 &= \sqrt{(28861 - 5487)^2 + (20861 - 2548)^2 + (22559 - 2898)^2 + (28513 - 4677)^2 + (32491 - 4750)^2} \\ &= 51049 \end{aligned}$$

$$\begin{aligned} K2 &= \sqrt{(7280 - 5487)^2 + (4382 - 2548)^2 + (5643 - 2898)^2 + (7735 - 4677)^2 + (9005 - 4750)^2} \\ &= 6447 \end{aligned}$$

$$\begin{aligned} K3 &= \sqrt{(2048 - 5487)^2 + (2726 - 2548)^2 + (1902 - 2898)^2 + (1748 - 4677)^2 + (2161 - 4750)^2} \\ &= 5304 \end{aligned}$$

$$= 5304$$

....

- Perhitungan jarak antara K1, K2 dan K3 terhadap data ke-24

$$\begin{aligned} K1 &= \sqrt{(28861 - 2227)^2 + (20861 - 2114)^2 + (22559 - 2356)^2 + (28513 - 2538)^2 + (32491 - 2590)^2} \\ &= 55116 \end{aligned}$$

$$\begin{aligned} K2 &= \sqrt{(7280 - 2227)^2 + (4382 - 2114)^2 + (5643 - 2356)^2 + (7735 - 2538)^2 + (9005 - 2590)^2} \\ &= 10471 \end{aligned}$$

$$\begin{aligned} K3 &= \sqrt{(2048 - 2227)^2 + (2726 - 2114)^2 + (1902 - 2356)^2 + (1748 - 2538)^2 + (2161 - 2590)^2} \\ &= 1192 \end{aligned}$$

7. Mengelompokkan objek kembali berdasarkan nilai jarak terdekat seperti langkah 4. Nilai perhitungan jarak antara data dengan semua titik pusat iterasi 1 dapat dilihat pada Tabel 5.

Tabel 5. Nilai perhitungan jarak antara data dan centroid iterasi 1

No	NK1	NK2	NK3	NMIN
1	0	44797	55775	0
2	51049	6447	5304	5304
3	37096	7735	18832	7735
4	52255	13828	11105	11105
5	42636	3643	13370	3643
6	41278	4046	14615	4046
7	42799	3603	14108	3603
8	55584	11026	1170	1170
9	46617	2696	9229	2696
10	59733	15048	4169	4169
11	48063	5491	10031	5491
12	53980	9453	1936	1936
13	56558	11907	2235	2235
14	54443	9915	1407	1407
15	53483	10075	3806	3806
16	60014	15291	4429	4429
17	55505	11046	1804	1804
18	56790	12079	1797	1797
19	59924	15198	4358	4358
20	58365	13679	2853	2853
21	54727	10146	1277	1277
22	51435	6760	4769	4769
23	49674	7111	9297	7111
24	55116	10471	1192	1192

Tabel 5 menerangkan bahwa anggota Cluster 1 berjumlah 1 data, tetap. Anggota Cluster 2 berjumlah 7 data, berubah. Anggota Cluster 3 berjumlah 16 data, berubah. Hingga proses ini disebut dengan iterasi pertama. Karena ada perubahan pada anggota cluster, maka proses diulangi ke iterasi berikutnya.

8. Dimulai kembali dengan penentuan centroid iterasi kedua berdasarkan nilai titik tengah masing-masing cluster yang terbentuk pada iterasi pertama.

Tabel 6. Centroid Cluster Iterasi Kedua

Centroid	Tahun				
	2019	2020	2021	2022	2023
K1	28861	20861	22559	28513	32491
K2	7536	4644	6035	8172	9613
K3	2263	2715	1965	1931	2323

9. Melakukan perulangan langkah ketiga dengan titik pusat cluster baru Tabel 6. Hasil dari perhitungan jarak iterasi kedua dapat dilihat pada Tabel 7.

No	NK1	NK2	NK3	NMIN
1	0	43909	55476	0
2	51049	7368	4973	4973
3	37096	6837	18527	6837
4	52255	14313	11087	11087
5	42636	3369	13060	3369
6	41278	3340	14313	3340
7	42799	2974	13803	2974
8	55584	11928	1071	1071
9	46617	3439	8923	3439
10	59733	15955	4412	4412
11	48063	6117	9714	6117
12	53980	10350	1627	1627
13	56558	12777	2339	2339
14	54443	10809	1101	1101
15	53483	10890	3711	3711
16	60014	16192	4681	4681
17	55505	11928	1770	1770
18	56790	12971	1922	1922
19	59924	16099	4607	4607
20	58365	14584	3073	3073

21	54727	11049	954	954
22	51435	7627	4477	4477
23	49674	7559	9090	7559
24	55116	11366	978	978

Tabel 7 menerangkan bahwa anggota Cluster 1 berjumlah 1 data, tetap. Anggota Cluster 2 berjumlah 7 data, tetap. Anggota Cluster 3 berjumlah 16 data, tetap. Hingga proses ini disebut dengan iterasi kedua dan proses dihentikan karena anggota cluster sudah konsisten.

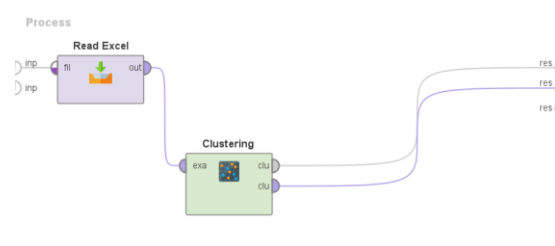
Tabel 8. Centroid Masing-Masing Cluster

Centroid	Tahun				
	2019	2020	2021	2022	2023
K1	28861	20861	22559	28513	32491
K2	7536	4644	6035	8172	9613
K3	2263	2715	1965	1931	2323

Dimana cluster 1 menunjukkan rumah sakit dengan kunjungan terbanyak setiap tahunnya. Cluster 2 menunjukkan rumah sakit dengan kunjungan sedang. Dan cluster 3 menunjukkan rumah sakit dengan kunjungan rendah.

PENGUJIAN DENGAN SOFTWARE RAPIDMINER STUDIO VERSI 10.3

Salah tools yang dapat digunakan untuk analisis data mining adalah RapidMiner. RapidMiner dikategorikan sebagai alat visual programming. Pengujian perhitungan yang dilakukan secara manual diuji dengan RapidMiner. Desain proses pengujian dapat dilihat pada Gambar 1.



Gambar 1. Desain Proses pada Rapidminer Studio

Operator Read Excel merupakan operator untuk membaca data yang akan diuji. Algoritma pada operator Clustering yang digunakan adalah

Algoritma K-Means. Hasil dari proses ini dapat dilihat pada Gambar 2.

Cluster Model

```
Cluster 0: 7 items
Cluster 1: 16 items
Cluster 2: 1 items
Total number of items: 24
```

Gambar 2. Hasil Cluster dari Rapidminer Studio

Anggota cluster model yang terbentuk pada Gambar 2 sama dengan hasil perhitungan manual. Hanya ada perbedaan dalam nama cluster. Hal ini dapat terjadi karena pemilihan centroid awal dalam RapidMiner bisa berbeda dengan perhitungan manual sehingga dapat menyebabkan perbedaan nama cluster.

Tabel 9. Jumlah Anggota Cluster Manual dan Cluster RapidMiner

Cluster Manual/ Cluster Rapid Miner	Jumlah Anggota
K1/ Cluster 2	1
K2/ Cluster 0	7
K3/ Cluster 1	16

SIMPULAN

Hasil clustering menunjukkan bahwa Algoritma K-Means berhasil mengelompokkan rumah sakit berdasarkan kunjungan pasien selama lima tahun terakhir. Cluster 1 (K1) merepresentasikan rumah sakit dengan intensitas kunjungan yang sangat tinggi. Cluster 2 (K2) menggambarkan rumah sakit dengan tingkat kunjungan menengah, sementara Cluster 3 (K3) memiliki kunjungan yang rendah.

DAFTAR PUSTAKA

- Di. P. Mulya, "Analisa Dan Implementasi Association Rule Dengan Algoritma Fp-Growth Dalam Seleksi Pembelian Tanah Liat (Studi Kasus Di Pt. Anveve Ismi Berjaya)," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 1, no. 1, hal. 47–57, 2019, doi: 10.47233/jteksis.v1i1.6.
- S. Maulia, B. S. Ginting, dan A. Sihombing, "Implementasi Data Mining Pengelompokan Jenis Penyakit Pasien Menggunakan Metode Clustering (Studi Kasus: Puskesmas Sambirejo)," *J. Inform. Kaputama*, vol. 5, no. 1, hal. 71–80, 2021, doi: 10.59697/jik.v5i1.304.
- M. Faizal Rizqi, M. Martanto, dan U. Hayati, "Clustering Kunjungan Pasien Menggunakan Algoritma K-Means Pada Rumah Sakit Di Wilayah Bekasi," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, hal. 230–237, 2024, doi: 10.36040/jati.v8i1.8327.
- & J. O. H. S.Lo, S.W. Hulu, H. Halawa, "Implementasi Algoritma K-Means Clustering Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Pulo Brayan," *Sains dan Teknol.*, vol. 5, no. 1, hal. 150–157, 2023.
- S. F. Munawar, A. Nasution, "Pengelompokan Wilayah Berdasarkan Kubik Air menggunakan Algoritma K-Means pada Perumda Air Minum Tirta Silaupiasa Kabupaten Asahan Regional Grouping Based on Cubic Water Using K - Means Algorithm at," vol. 13, hal. 789–801, 2024.
- I. Kanedi dan E. Suryana, "Penerapan Metode K-Means Clustering Dalam Pengelompokan Data Pasien Rawat Inap Peserta BPJS Di Rumah Sakit Umum Daerah Kabupaten Kaur," vol. 20, no. 2, hal. 493–500, 2024.
- R. Maulidina dan S. Y. Riska, "Application of the K-Means Algorithm for Clustering Plantation Crop Production in Indonesia," *Smatika J.*, vol. 13, no. 02, hal. 339–349, 2023, doi: 10.32664/smatika.v13i02.991.
- K. P. Sinaga dan M. S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, hal. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.
- A. Lingga, "Penerapan Algoritma K-Means Clustering Untuk Klasterisasi Penyakit Pasien," vol. 1, no. 3, hal. 93–102, 2024.
- Mohamad Fahmi Hafidz dan Sri Lestari, "Solution to Scalability and Sparsity Problems in Collaborative Filtering using K-Means Clustering and Weight Point Rank (WP-Rank)," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 7, no. 4, hal. 743–750, 2023, doi: 10.29207/resti.v7i4.4543.
- N. N. Hasanah dan A. S. Purnomo, "Implementasi Data Mining Untuk Pengelompokan Buku Menggunakan Algoritma K-Means Clustering (Studi Kasus: Perpustakaan Politeknik LPP Yogyakarta)," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 4, no. 2, hal. 300–311, 2022, doi: 10.47233/jteksis.v4i2.499.
- A. Naufal, M. Fadhil, dan . W., "Clustering and Financial Performance Analysis of Indonesian Coal Mining Industry Stock Prices," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 6, no. 1, hal. 20–28, 2024, doi: 10.47233/jteksis.v6i1.1081.
- Ajeng Arifa Chantika Rindu, Ria Astriratma, dan Ati Zaidiah, "K-Means Algorithm Implementation for Project Health Clustering," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 7, no. 5, hal. 1064–1076, 2023, doi: 10.29207/resti.v7i5.5181.
- Haris Kurniawan, Sarjon Defit, dan Sumijan, "Data Mining Menggunakan Metode K-Means Clustering Untuk Menentukan Besaran Uang Kuliah Tunggal," *J. Appl. Comput. Sci. Technol.*, vol. 1, no. 2, hal. 80–89, 2020, doi: 10.52158/jacost.v1i2.102.
- I. Rizky Mahartika dan A. Wibowo, "Data Mining Klasterisasi dengan Algoritme K-Means untuk Pengelompokan Provinsi Berdasarkan Konsumsi Bahan Bakar Minyak Nasional," *Semin. Nas. Sist. Inf. dan Teknol.*, vol. 3, no. 1, hal. 87–91, 2019.