

**PENERAPAN METODE K-MEANS UNTUK MENGELOMPOKKAN REKAM MEDIS
PASIE BERDASARKAN DIAGNOSA PENYAKIT GUNA MENENTUKAN DIAGNOSA
TERTINGGI PADA SUATU PERIODE (Study Kasus : Klinik Dokter Kita)**

Pendi Supratman¹, Verawati², Sukatmi³

^{1,2,3}Institut Teknologi Bisnis dan Bahasa Dian Cipta Cendikia

Email: ¹pendisupratman130302@gmail.com, ²Verawati@dcc.ac.id, ³sukatmi@gmail.com

ABSTRACT

The development of information technology in the digital era has driven major transformation in the health sector, especially in data management. One technique that plays an important role is data mining, which allows the discovery of hidden patterns and complex relationships in large amounts of data. This technique is very relevant for patient data analysis and disease diagnosis, especially in grouping patients based on disease type in order to understand distribution patterns and carry out more appropriate interventions. The K-Means method is an algorithm that is often used in the data grouping process, which allows identifying groups of patients with similar characteristics and helps determine the dominant disease in a certain period. At Our Doctor's Clinic, patient data continues to increase, but management is still manual so it is not optimal for in-depth analysis. The current grouping process based on disease diagnosis is general and makes it difficult for management to identify diseases that frequently appear in a certain period, which has an impact on difficulties in making strategic decisions. Therefore, This research applies the K-Means clustering method and RapidMiner Studio software version 10.2. with the aim of automating patient grouping based on diagnosis. The data used is patient data for 2023 which consists of 966 patients with a total of 1609 controls with 20 types of disease. The results of the research show that there are three groups of disease diagnoses that occur frequently (highly dominant) and therefore require careful attention, namely the diagnosis of Grastitis, the diagnosis of ISPA and the diagnosis of Myalgia. Through this analysis, it is hoped that clinics can identify dominant diseases, understand distribution patterns, and increase the effectiveness of drug procurement planning and resource allocation. The results of this clustering are also expected to provide a basis for predicting future disease trends, allowing clinics to take preventive measures more proactively. Thus, it is hoped that the K-Means method can improve the quality of health services at Our Doctor's Clinic, make data-based decision making easier, and provide faster and more precise treatment for patients.

Keyword : Data Mining, Diagnosis, K-Means, Rapidminer, Medical Recods

INTISARI

Perkembangan teknologi informasi dalam era digital telah mendorong transformasi besar di sektor kesehatan, khususnya dalam pengelolaan data. Salah satu teknik yang berperan penting adalah data mining, yang memungkinkan penemuan pola tersembunyi dan hubungan kompleks dalam data berukuran besar. Teknik ini sangat relevan untuk analisis data pasien dan diagnosa penyakit, khususnya dalam mengelompokkan pasien berdasarkan jenis penyakit guna memahami pola penyebaran serta melakukan intervensi yang lebih tepat. Metode K-Means merupakan salah satu algoritma yang sering digunakan dalam proses pengelompokan data, yang memungkinkan identifikasi kelompok pasien dengan karakteristik serupa dan membantu menentukan penyakit dominan dalam periode tertentu. Pada Klinik Dokter Kita, data pasien terus bertambah, namun pengelolaannya masih manual sehingga tidak optimal untuk analisis yang mendalam. Proses pengelompokan berdasarkan diagnosa penyakit saat ini bersifat umum dan menyulitkan manajemen dalam mengidentifikasi penyakit yang sering muncul dalam periode tertentu, yang berdampak pada kesulitan dalam pengambilan keputusan strategis. Penelitian ini menerapkan metode K-Means clustering dan perangkat lunak RapidMiner Studio versi 10.2. dengan tujuan untuk mengotomatisasi pengelompokan pasien berdasarkan diagnosa. Data yang digunakan adalah data pasien tahun 2023 yang terdiri dari 966 pasien dengan jumlah melakukan kontrol sebanyak 1609 kali dengan 20 macam penyakit. Hasil penelitian menunjukkan terdapat tiga kelompok diagnose penyakit yang sering terjadi (dominan tinggi) sehingga membutuhkan perhatian yang seksama, yaitu diagnosa Grastitis, diagnosa ISPa dan diagnosa Myalgia, Melalui analisis ini, diharapkan klinik dapat mengidentifikasi

penyakit yang dominan, memahami pola penyebaran, dan meningkatkan efektivitas perencanaan pengadaan obat serta alokasi sumber daya. Hasil pengelompokan ini juga diharapkan menjadi dasar untuk prediksi tren penyakit di masa depan, memungkinkan klinik untuk melakukan langkah pencegahan secara lebih proaktif. Dengan demikian, metode K-Means diharapkan dapat meningkatkan kualitas pelayanan kesehatan di Klinik Dokter Kita, mempermudah pengambilan keputusan berbasis data, dan memberikan penanganan yang lebih cepat dan tepat bagi pasien.

Kata Kunci : Data Mining, Diagnosa, K-Means, RapidMiner, Rekam Medis

1. PENDAHULUAN

Dalam era digital, sektor kesehatan mengalami transformasi signifikan dengan pemanfaatan teknologi informasi, terutama dalam pengelolaan data yang kompleks. Data mining, khususnya metode K-Means clustering, menjadi solusi dalam mengelompokkan pasien berdasarkan diagnosa penyakit, sehingga pola penyebaran penyakit dapat teridentifikasi secara efektif. Klinik Dokter Kita, yang data pasiennya terus meningkat, menghadapi kendala dalam pengelompokan pasien yang masih manual, sehingga menyulitkan identifikasi penyakit yang dominan dalam suatu periode. Penelitian ini bertujuan mengimplementasikan metode K-Means untuk mengotomatisasi proses tersebut agar manajemen dapat mengambil keputusan strategis yang berbasis data. Melalui pengelompokan otomatis ini, klinik diharapkan mampu meningkatkan akurasi layanan, mempercepat pengadaan obat, dan mengoptimalkan alokasi sumber daya kesehatan. Selain itu, hasil analisis ini juga dapat mendukung prediksi tren penyakit di masa depan, memungkinkan langkah preventif yang lebih proaktif dan efektif (M. Faizal Rizqi et al., 2024).

1.1. Data Mining

Data mining merupakan proses mengekstraksi informasi berharga secara manual dari database yang tidak diketahui. Teknik ini muncul pada tahun 1990-an sebagai metode yang efektif untuk mengidentifikasi pola dan informasi, serta menemukan hubungan antar data. Proses ini memungkinkan pengelompokan data ke dalam satu atau lebih cluster, di mana objek-objek dalam satu cluster memiliki kesamaan yang tinggi satu sama lain (Zahra et al., 2024).

1.2. K-Means

K-Means merupakan salah satu metode dalam data mining yang digunakan untuk mengelompokkan data atau melakukan clustering, dimana data dikelompokkan ke dalam satu atau lebih cluster. Metode ini memungkinkan data dengan karakteristik yang serupa untuk berada dalam satu cluster yang sama. Sementara data dengan karakteristik yang berbeda akan dikelompokkan ke dalam cluster yang berbeda (Dinata et al., 2020).

1.3. Rekam Medis Pasien

Menurut UU Praktik Kedokteran Pasal 46 ayat (1), menerangkan bahwa rekam medis pasien adalah dokumen yang berisi catatan pasien berupa identitas, hasil pemeriksaan, pengobatan, tindakan, serta pelayanan lainnya yang diberikan kepada pasien. Rekam medis yang memuat data dan tindakan dokter memiliki peran yang sangat penting dalam bidang kesehatan, baik untuk pasien, dokter, maupun dalam pengembangan ilmu pengetahuan (Abduh, 2021). Rekam medis pasien biasanya juga berisi tentang diagnosa penyakit setiap pasien. Diagnosa penyakit merujuk pada istilah yang digunakan untuk proses menegakkan atau mengidentifikasi suatu penyakit atau masalah kesehatan yang dialami oleh pasien (Rohma et al., 2022). Pasien adalah setiap orang yang melakukan konsultasi terkait masalah kesehatan kepada dokter untuk memperoleh pelayanan kesehatan yang diperlukan (Prasetyo et al., 2023).

2. METODE PENELITIAN

2.1. Objek Penelitian

Penelitian ini akan menyelidiki data rawat jalan pada Klinik Dokter Kita. Untuk pengumpulan data ini, penelitian ini menggunakan metode wawancara dengan datang langsung ke tempat penelitian untuk mendapatkan keterangan-keterangan yang diperlukan sebagai bahan penelitian. Selain itu, peneliti juga melakukan pengamatan langsung untuk mendapatkan informasi mengenai rekam medis pasien dan diagnosa penyakit atas persetujuan pihak yang bersangkutan. Pada tahap ini diperoleh data yang akan diolah yaitu data pasien tahun 2023 sebanyak 966 pasien, jumlah control sebanyak 1609 dengan jumlah penyakit sebanyak 20 diagnosa.

2.2. Metodologi Penelitian

Metode penelitian yang digunakan adalah *Knowledge Discovery in Database (KDD)*. *Knowledge Discovery in Database (KDD)* di definisikan sebagai metode atau proses untuk memperoleh informasi dari basis data yang sudah ada (Alghifari & Juardi, 2021). Tahap penyelesaiannya adalah sebagai berikut:

2.2.1. Data Selection

Proses pemilihan data dari database yang akan digunakan untuk analisis dan pengolahan selanjutnya dikenal dengan *data selection*. Pada tahap ini, data pasien dan jumlah kontrol dikelompokkan, dengan total 966 pasien dan jumlah kontrol sebanyak 1609.

2.2.2. Data Preprocessing

Tahap awal dalam pengolahan data bertujuan untuk mempersiapkan data sebelum analisis lebih lanjut disebut juga dengan *data preprocessing*. Pada tahap ini, data yang akan diolah disaring untuk menghindari gangguan dari data yang tidak relevan (*noise*) atau data yang tidak konsisten. Pada tahap ini, data jumlah control di kelompokkan berdasarkan jumlah diagnosa penyakit. Dari 6 variabel, pada selection data dan ada variabel yang perlu di cleaning seperti umur, alamat, dan nama kecamatan. Sehingga hasil dari cleaning data tersisa 3 variabel seperti nomor rekam medis, nama pasien dan diagnosa.

2.2.3. Transformation

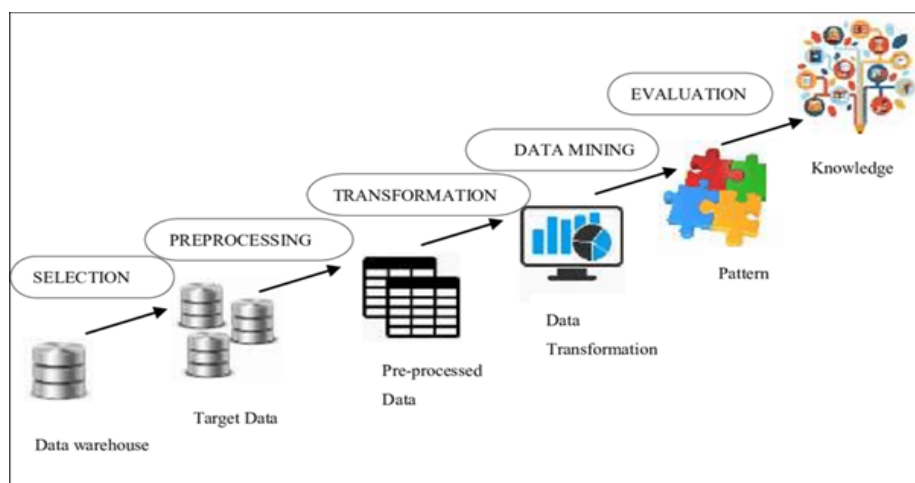
Transformation adalah tahap dimana data yang sedang diproses diubah agar sesuai dengan model atau algoritma yang digunakan dalam pengolahan data. Pada tahap ini data dikelompokkan berdasarkan diagnosa penyakit, jumlah kontrol dan jumlah pasien, yaitu jumlah kunjungan pasien dalam satu periode.

2.2.4. Data Mining

Proses ekstraksi dan pengolahan pengetahuan menghasilkan model yang dapat digunakan untuk menghasilkan informasi yang penting dan bermanfaat. Penerapan algoritma K-Means untuk mengelompokkan pasien ke dalam beberapa cluster berdasarkan kesamaan diagnosa. Algoritma ini akan menentukan jumlah cluster (*k*) yang optimal untuk menghasilkan pengelompokan yang paling informatif. Tahap pertama yang kita lakukan adalah dengan menentukan titik centroid awal. Pada tahap ini data diolah untuk mendapatkan informasi diagnosa penyakit yang muncul dengan kecenderungan tinggi, sedang dan rendah.

2.2.5. Evaluasi

Evaluasi adalah proses mempresentasikan hasil dari model yang telah dikembangkan serta menguji akurasi dan kesesuaian terhadap data yang relevan. Tahapan knowledge discovery in database (KDD) dapat dilihat pada Gambar 1.



Gambar 1. Tahapan *Knowledge Discovery in Database (KDD)*

2.3. Aplikasi Yang Digunakan

2.3.1. RapidMiner

RapidMiner adalah aplikasi perangkat lunak canggih yang digunakan untuk analisis data dan pembelajaran mesin. Platform ini menawarkan berbagai alat untuk persiapan data, pemodelan, evaluasi, dan implementasi. RapidMiner dirancang agar mudah digunakan, memungkinkan pengguna untuk dengan cepat membangun dan menguji berbagai model, bahkan bagi mereka yang tidak memiliki latar belakang pemrograman (Rafi Nahjan et al., 2023). Implementasi adalah proses penerapan ide, konsep, kebijakan atau inovasi dalam suatu tindakan praktis sehingga memberikan dampak baik berupa perubahan pengetahuan, keterampilan maupun nilai dan sikap. Secara sederhana, implementasi dapat diartikan sebagai pelaksanaan atau penerapan aktivitas yang saling menyesuaikan (Magdalena et al., 2021).

2.3.2. Microsoft Office Excel

Microsoft Office Excel adalah aplikasi yang digunakan untuk mengelola data secara otomatis, termasuk perhitungan dasar, penerapan fungsi dan rumus, pembuatan grafik, serta pengelolaan data (Putri, 2021).

3. HASIL DAN PEMBAHASAN

3.1. Proses Tahapan Data Mining

3.1.1. Selection Data

Pada selection data pengumpulan data rekam medis pasien meliputi variabel-variabel seperti nomor rekam medis, nama pasien, usia, nama desa, nama kecamatan, dan diagnosa penyakit. Data rekam medis pasien dapat dilihat pada Tabel 1.

Tabel 1. Sampel Data Rekam Medis Pasien Pada Klinik Dokter Kita

No. RM	Nama	Umur	Nama Desa	Nama Kecamatan	Diagnosa Penyakit
10645	Made Jani	20 Thn	Rejosari	Negeri Agung	Dermatitis
865	Saniah	64 Thn	Sukabumi	Pakuan Ratu	Dermatitis
512	Bibit	47 Thn	Sukabumi	Pakuan Ratu	Myalgia
2632	Kunarti	49 Thn	Rejosari	Negeri Agung	Gastritis
9371	Lamijan	80 Thn	Tanjung Serupa	Pakuan Ratu	Gastritis
583	Suharno	46 Thn	Gedung Meneng	Negeri Agung	ISPA
11872	Ketut Lusiana	29 Thn	Kota Jawa	Pakuan Ratu	Gastritis
219	Sukatno	49 Thn	Sukabumi	Pakuan Ratu	Asthma
6926	Dini Puspita Sari	35 Thn	Way Tawar	Pakuan Ratu	Gingivitis
1046	Sumarmi	75 Thn	Sukabumi	Pakuan Ratu	Myalgia
11915	Jumiran	93 Thn	Sukabumi	Pakuan Ratu	Konjungtivitis
11914	Susanto	47 Thn	Way Tawar	Pakuan Ratu	Myalgia
6592	Maryana	55 Thn	Way Tawar	Pakuan Ratu	Hipertensi
2798	Wayan Jafa	18 Thn	Tanjung Rejo	Negeri Agung	Dermatitis
11909	Supriyadi	47 Thn	Sukabumi	Pakuan Ratu	Headache
5722	Agus Purwadi	48 Thn	Purwa Agung	Negara Batin	Vertigo
2864	Mukarom	53 Thn	Sunsang	Negeri Agung	Gastritis
7285	Sri Dati	56 Thn	Purwa Agung	Negara Batin	Diabetes Militus

3.1.2. Preprocessing Data

Preprocessing data adalah proses cleaning data, yaitu untuk menghilangkan data noise atau missing values pada data yang nilainya tidak sesuai. Proses cleaning data ini dapat dilihat pada Tabel 2.

Tabel 2. Sample Data Diagnosa Setelah Dilakukan Preprocessing Data

Gejala	Diagnosa	Jumlah Kontrol
Rasa sakit atau tidak enak di kepala, leher, atau bagian atas bahu.	Headache	120
Kondisi yang menyebabkan penderitanya merasa pusing dan seolah-olah kepala berputar.	Vertigo	48
Penyakit yang memengaruhi saluran pernapasan atas dan bawah ini dikenal sebagai ISPA. Gejala umumnya yang	ISPA	164

terikat dengan ISPA meliputi batuk, pilek, nyeri otot, demam, dan lain-lain.		
Peradangan pada dinding lambung yang bisa menyebabkan nyeri dibagian uluhati.	Gastritis	229
Penyakit infeksi ini disebabkan oleh bakteri Salmonella typhi.		
Beberapa gejala yang muncul meliputi demam tinggi, gangguan pencernaan, diare, sakit kepala, dan nyeri otot.	Typhoid	87

3.1.3. Transformation

Transformasi data yaitu proses pembersihan data, normalisasi, dan seleksi fitur untuk memastikan bahwa data yang digunakan berkualitas baik dan siap untuk dianalisis. Sample data setelah di transformasikan dapat dilihat pada tabel 3.

Tabel 3. Sample Data Setelah Ditransformasikan

Gejala	Diagnosa	Jumlah Pasien	Jumlah Kontrol
Rasa sakit atau tidak enak di kepala, leher, atau bagian atas bahu.	Headache	66	120
Kondisi yang menyebabkan penderitanya merasa pusing dan seolah-olah kepala berputar-putar.	Vertigo	36	48
Penyakit yang memengaruhi saluran pernapasan atas dan bawah ini dikenal sebagai ISPA. Gejala umumnya yang terikat dengan ISPA meliputi batuk, pilek, nyeri otot, demam, dan lain-lain.	ISPA	100	164
Kondisi ketika suhu tubuh lebih tinggi dari normal, yaitu di atas 38°C.	Fever	30	42
Peradangan pada dinding lambung yang bisa menyebabkan nyeri dibagian uluhati.	Gastritis	178	229
Penyakit infeksi yang disebabkan oleh bakteri Salmonella typhi, Gejalanya seperti Demam tinggi, gangguan pencernaan, diare, sakit kepala, dan nyeri otot.	Typhoid	60	87
Gejala umum yang di rasakan penderita dengan diagnosa DHF adalah Demam tinggi, sakit kepala, menggigil, tidak nafsu makan, nyeri otot.	DHF	35	43
Penyakit yang terjadi ketika gusi mengalami peradangan akibat infeksi bakteri dan penumpukan plak dan karang gigi.	Gingivitis	20	26
Penderita Diabetes Militus biasanya mengalami perubahan berat badan, badan terasa lemas, sering mengantuk.	Diabetes Militus	18	59
Gejalanya biasanya berupa nyeri kepala bagian belakang, pusing, mudah lemas, pandangan kabur.	Hipertensi	21	41
Kondisi dimana tubuh mengalami kelainana berupa peradangan kronik saluran napas yang menyebabkan penyempitan saluran napas, sehingga menyebabkan gejala episodik berulang seperti mengi (napas yang berbunyi), sesak napas, dada terasa berat, hingga batuk yang terjadi terutama saat malam hari.	Asthma	20	50
Peradangan pada kulit yang menimbulkan gejala kemerahan, gatal, ruam	Dermatitis	80	106
Istilah medis untuk nyeri otot, yang merupakan	Myalgia	70	153

gejala dari suatu kondisi atau penyakit. Myalgia bukan penyakit, melainkan reaksi tubuh terhadap kondisi tertentu.			
Peradangan pada saluran pencernaan yang menyebabkan diare, mual, muntah, nyeri perut	GEA	28	49
Peradangan pada selaput bening mata (konjungtiva) yang menyebabkan mata merah, gatal, berair, dan bengkak.	Konjungtivitis	26	51
Gejalanya meliputi batuk yang tidak kunjung sembuh, sakit pada bagian dada, sakit kepala, kadang disertai pilek.	Bronchitis	51	103
Sensasi tidak nyaman yang dirasakan di area perut, mulai dari bawah tulang rusuk hingga lipat paha. Gejala abdominal pain bisa berupa kram, mulas, rasa tertusuk, atau kedutan di perut.	Abdominal Pain	60	109
Penyakit yang disebabkan oleh penumpukan asam urat dalam darah, sehingga membentuk kristal asam urat yang tajam dan mengendap di sendi, jaringan, dan organ lainnya	Asam Urat	35	69
Kondisi ketika kadar kolesterol dalam darah melebihi batas normal. Gejalanya seperti Nyeri di bagian tengkuk, Mudah merasa lelah dan nyeri di kaki, Nyeri di bagian dada, Kram di beberapa bagian tubuh pada malam hari saat terbangun dari tidur	Cholesterol	17	32
Kondisi dimana terdapat endapan kecil dan keras yang terbentuk di ginjal dan sering menyakitkan saat buang air kecil. Gejala yang sering muncul biasanya nyeri perut tembus ke pinggang, nyeri pada saat buang air kecil.	Batu Ginjal	15	28

3.1.4. Data Mining

Penerapan algoritme K-Means bertujuan untuk mengelompokkan pasien kedalam beberapa kluster berdasarkan kesamaan diagnosa. Algoritma ini akan menentukan jumlah kluster (k) yang optimal guna menghasilkan pengelompokan yang paling informatif. Tahap awal pada proses ini adalah dengan menentukan titik cetroid awal yang dipilih secara acak, yang hasilnya dapat dilihat di tabel 4.

Tabel 4. Titik Centroid Awal

Titik Centroid Awal		
C1	178	229
C2	15	28

Setelah menentukan titik awal centroid, langkah berikutnya adalah menghitung jarak setiap data terhadap centroid yang paling dekat menggunakan metode Euclidean Distance. Euclidian distance akan dihitung dimulai dari data ke-1 iterasi 1 yaitu

$$[C1] = \sqrt{(66 - 178)^2 + (120 - 229)^2} = 156.2849961$$

$$[C2] = \sqrt{(66 - 15)^2 + (120 - 28)^2} = 105.1903037$$

Data berikutnya dihitung dengan cara dan metode yang sama.

Setelah diperoleh nilai untuk setiap cluster, langkah selanjutnya adalah menghitung jarak terdekat antara pusat data kluster dengan titik-titik dalam kluster C1 dan C2. Nilai yang didapat dari jarak terpendek ini akan dicatat

pada kolom kluster 1 dan kluster 2. Setiap nilai akan diberi tanda * sebagai penanda bahwa nilai tersebut termasuk dalam kluster dengan jarak terpendek. Hasil dari perhitungan ini dapat dilihat pada tabel 5 :

Tabel 5. Hasil Kluster Pada Iterasi 1

C1	C2	Jarak Terpendek	Cluster 1	Cluster 2
156,2849961	105,1903037	105,1903037		*
230,0543414	29	29		*
101,5332458	160,3776792	101,5332458	*	
238,4806072	20,51828453	20,51828453		*
0	258,7856256	0	*	
184,6293584	74,20242584	74,20242584		*
234,6167087	25	25		*
257,241132	5,385164807	5,385164807		*
233,4523506	31,144823	31,144823		*
244,9346852	14,31782106	14,31782106		*
238,7571988	22,56102835	22,56102835		*
157,2672884	101,5332458	101,5332458		*
132,0605922	136,5650028	132,0605922	*	
234,3074903	24,69817807	24,69817807		*
234,0683661	25,49509757	25,49509757		*
178,8994131	83,19254774	83,19254774		*
168,2973559	92,66067127	92,66067127		*
214,5903073	45,61797891	45,61797891		*
254,4209111	4,472135955	4,472135955		*
258,7856256	0	0		*

Setelah perhitungan pada Iterasi 1, langkah berikutnya adalah menentukan titik centroid yang baru. Centroid baru ini diperoleh dengan menghitung rata-rata dari nilai-nilai yang terdapat dalam masing-masing kluster. Tabel dibawah ini menunjukkan data yang termasuk dalam kluster 1 dan 2, yang akan digunakan dalam perhitungan untuk menetapkan pusat kluster terbaru, hasil kluster tersebut dapat dilihat di tabel 6, tabel 7 dan hasil iterasi 2 pada titik centroid di tabel 8 .

Tabel 6. Kluster 1 Iterasi 1

Cluster 1	
100	164
178	229
70	153

Titik centroid baru C1

$$C1 = \frac{100+178+70}{3} = 116$$

$$C2 = \frac{164+229+153}{3} = 182$$

Tabel 7. Kluster 2 Iterasi 2

Cluster 2	
66	120
36	48
30	42
60	87
35	43

Cluster 2	
20	26
18	59
21	41
20	50
80	106
28	49
26	51
51	103
60	109
35	69
17	32
15	28

Titik centroid baru C2

$$C1 = \frac{66+36+30+60+35+20+18+21+20+80+28+26+51+60+35+17+15}{17} = 36,35294118$$

$$C2 = \frac{120+48+42+87+43+26+59+41+50+106+49+51+103+109+69+32+28}{17} = 62,52941176$$

Tabel 8. Titik Centroid Baru Iterasi 2

Cluster Baru	
C1	C2
116	36,35294118
182	62,52941176

Perhitungan akan terus berlanjut hingga tidak terjadi perubahan pada anggota kluster. Langkah-langkah perhitungan dilakukan sesuai metode pada iterasi 1. Pada iterasi ke-3, perhitungan dilakukan dengan metode yang sama seperti pada iterasi 1 dan iterasi 2, sehingga diperoleh hasil pada tabel 9:

Tabel 9. Hasil Perhitungan Pada Akhir Iterasi

C1	C2	Jarak Terpendek	Cluster 1	Cluster 2
79,64923101	64,66696691	64,66696691		*
156,0640894	14,53369787	14,53369787		*
24,08318916	119,7799164	24,08318916	*	
164,3045952	21,48991878	21,48991878		*
77,80102827	218,578009	77,80102827	*	
110,2769242	34,02929737	34,02929737		*
160,8788364	19,5762196	19,5762196		*
183,1720503	40,02270117	40,02270117		*
157,2672884	18,68922677	18,68922677		*
170,0176461	26,44292672	26,44292672		*
163,2176461	20,60108842	20,60108842		*
84,09518417	61,60160538	61,60160538		*
54,3783045	96,52487712	54,3783045	*	
159,4772711	15,90020783	15,90020783		*
158,9370945	15,49550666	15,49550666		*
102,3034701	43,039573	43,039573		*
92,00543462	52,14114462	52,14114462		*
139,0323703	6,610519037	6,610519037		*
179,7247896	36,14666395	36,14666395		*
184,1656863	40,59837895	40,59837895		*

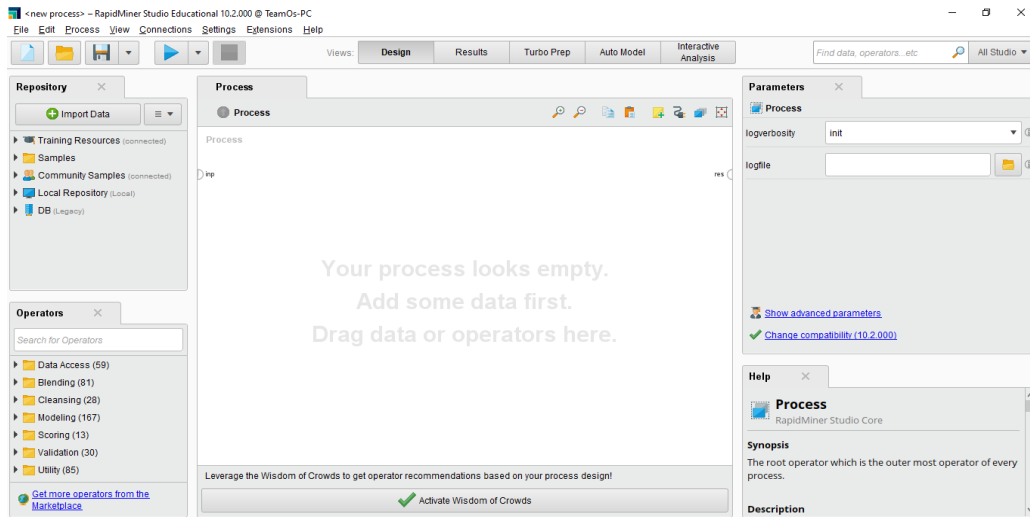
Sebagai penutup, hasil pengelompokkan kluster ditampilkan pada Tabel 10.

Tabel 10. Hasil Pengelompokan Data Kluster

No	Cluster	Keterangan	Hasil
1.	C1	Tinggi	3
2.	C2	Rendah	17

3.2. Implementasi Algoritma K-Means Pada RapidMiner

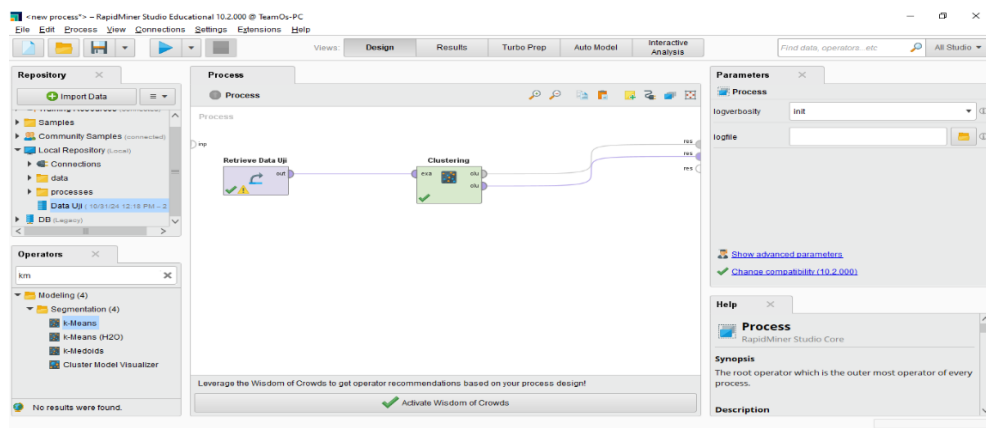
Pada implementasi dan pengujian ini, kami memanfaatkan perangkat lunak RapidMiner versi 10.2 dengan perangkat lunak ini, kami dapat membandingkan hasil pengolahan data secara manual dengan hasil yang diperoleh melalui bantuan perangkat lunak. Pada Gambar 2 dibawah ini menunjukkan tampilan utama RapidMiner Studio 10.2 saat aplikasi dibuka pertama kali.



Gambar 2. Tampilan Lembar Kerja RapidMiner

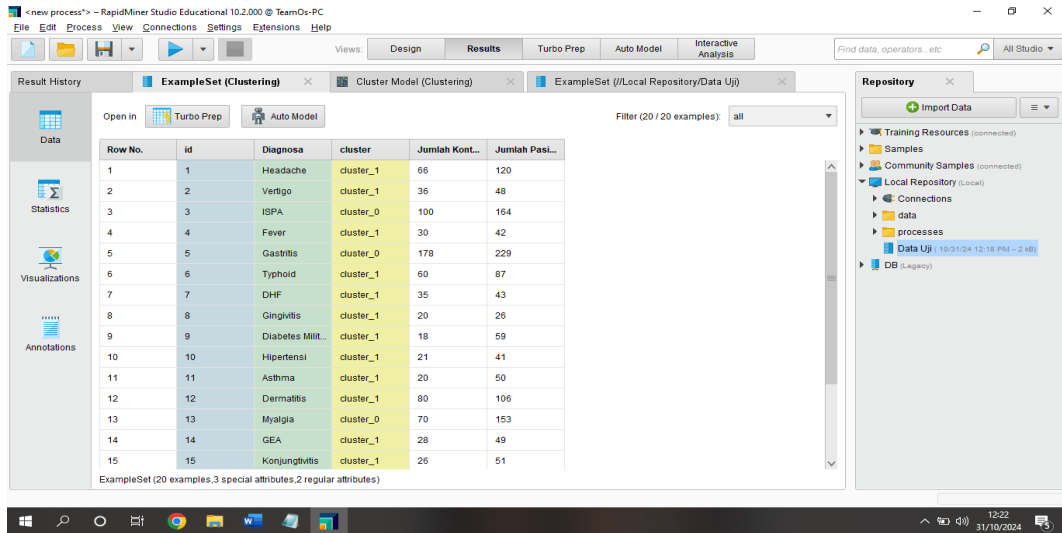
Saat membuka RapidMiner, akan ditampilkan halaman untuk memulai lembar kerja baru. Pilih opsi New Proses dan Blank untuk membuka tampilan desain lembar kerja di RapidMiner 10.2. Pada bagian ini, terdapat berbagai menu utama. Untuk memulai pengujian, data Excel, pilih menu Import Data. Setelah memilih Import Data, cari file excel yang akan digunakan. Setelah memilihnya, data excel akan ditampilkan dilayar. Temukan lembar kerja (sheet) yang berisi data yang ingin diolah, lalu blok data yang diperlukan. Setelah itu, pilih Next dan akhiri dengan menekan Finish.

Berikutnya menampilkan read data excel dan metode K-Means yang dapat dilihat pada Gambar 3.



Gambar 3. Tampilan Read Data Excel dan Metode K-Means

Seret file Excel yang sudah di import ke lembar Design di RapidMiner. Setelah itu, cari operator K-Means pada daftar operators dan seret ke area kerja. Hubungkan data Excel dengan K-Means, seperti yang terlihat pada gambar di atas, untuk melanjutkan proses data mining, klik dua kali pada operator K-Means Clustering dan tentukan jumlah cluster yang diinginkan. Dalam penelitian ini, digunakan dua cluster, sehingga kolom k diisi dengan angka 2. Setelah data excel dan operator K-Means siap, hubungkan data dari operator Read Excel ke operator K-Means dengan mengklik “out” pada Read Excel dan mengarahkannya ke “Result” pada K-Means. Kemudian klik Run untuk menjalankan proses, dan data akan dikelompokkan sesuai cluster yang di tentukan, tampilan dari hasil clustering dapat dilihat pada gambar 4.

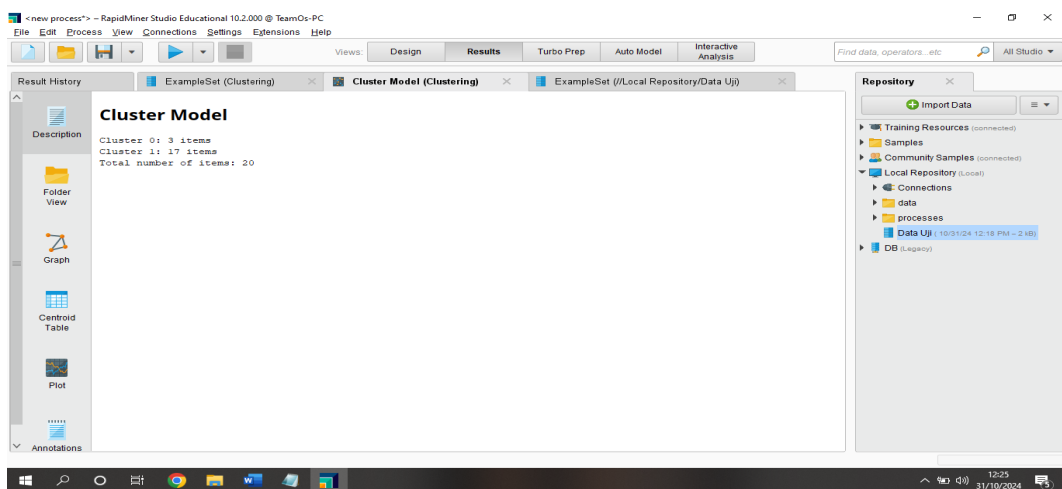


Row No.	id	Diagnosa	cluster	Jumlah Kont...	Jumlah Pasi...
1	1	Headache	cluster_1	66	120
2	2	Vertigo	cluster_1	36	48
3	3	ISPA	cluster_0	100	164
4	4	Fever	cluster_1	30	42
5	5	Gastritis	cluster_0	178	229
6	6	Typhoid	cluster_1	60	87
7	7	DHF	cluster_1	35	43
8	8	Gingivitis	cluster_1	20	26
9	9	Diabetes Millit...	cluster_1	18	59
10	10	Hipertensi	cluster_1	21	41
11	11	Asthma	cluster_1	20	50
12	12	Dermatitis	cluster_1	80	106
13	13	Myalgia	cluster_0	70	153
14	14	GEA	cluster_1	28	49
15	15	Konjungtivitis	cluster_1	26	51

Gambar 4. Tampilan Hasil Clustering Pada Exemple Set

Gambar di atas menunjukkan hasil clustering setelah proses running. Dari hasil yang diperoleh, proses clustering menggunakan Excel dan RapidMiner menghasilkan hasil yang serupa. Ini menunjukkan konsistensi antara pengolahan data manual dengan pengolahan data melalui perangkat lunak, memberikan validasi atas akurasi dan keandalan hasil clustering yang dihasilkan.

Tampilan Clustering Model memungkinkan kita melihat jumlah data yang dikelompokkan ke dalam setiap kluster. Berdasarkan gambar 5, terlihat bahwa data yang masuk ke dalam kluster 0 (kategori diagnosa tertinggi) berjumlah 3 item, sedangkan data yang masuk ke dalam kluster 1 (kategori diagnosa terendah) berjumlah 17 item. Dengan demikian, total keseluruhan item yang terkelompok adalah 20 item.



Cluster Model

Cluster 0: 3 items
Cluster 1: 17 items
Total number of items: 20

Gambar 5. Tampilan Cluster Model

3.3. Pembahasan Hasil Implementasi RapidMiner

Hasil perhitungan pada Microsoft Office Excel dan pengujian pada aplikasi RapidMiner menunjukkan adanya konsistensi yang signifikan. Pada cluster 1, terdapat tiga jenis diagnosa yang menunjukkan bahwa beberapa diagnosa lebih sering terjadi dan terkonsentrasi dalam cluster tersebut. Di sisi lain, cluster 2 memiliki 17 jenis diagnosis, yang mengindikasikan bahwa diagnosis tersebut lebih jarang terjadi. Di RapidMiner, hasil clustering ini dapat diamati melalui beragam visualisasi, seperti plot grafik, tampilan grafik, dan scatter chart, sehingga memudahkan kita untuk melihat data yang di kelompokkan dalam cluster yang berdekatan.

4. KESIMPULAN

Berdasarkan dari hasil penelitian yang sudah diuraikan sebelumnya, maka peneliti menyimpulkan dari penentuan titik centroid awal, kemudian berikutnya menghitung jarak setiap data terhadap centroid dengan nilai $C1 = 36.35294118$ serta $C2 = 62.52941176$, kemudian berikutnya menghitung nilai untuk setiap cluster dengan menghitung jarak terdekat antara pusat data kluster dengan titik-titik dalam cluster $C1$ dan $C2$ sampai ke pengelompokan hasil kluster dengan hasil akhir pengelompokan data nilai $C1$ tertinggi = 3 dan $C2$ terendah 17. Penelitian tersebut dari total 966 pasien dan jumlah kontrol sebanyak 1609 dalam satu periode terdapat tiga diagnosa utama yang memerlukan perhatian khusus, yaitu gastritis, ISPA (Infeksi Saluran Pernapasan Akut), dan myalgia. Gastritis adalah kondisi peradangan pada lapisan lambung yang dapat menyebabkan nyeri di area uluhati, sering kali akibat pola makan tidak teratur, infeksi, atau stres. ISPA menyerang saluran pernapasan bagian atas dan bawah dengan gejala seperti batuk, pilek, nyeri otot, dan demam, dan penyebaran dapat terjadi dengan cepat. Myalgia, di sisi lain, adalah kondisi nyeri otot sebagai reaksi tubuh terhadap situasi tertentu. Gastritis dan ISPA terlihat menonjol sebagai kondisi umum yang berdampak pada kesehatan masyarakat secara luas, sehingga peningkatan kesadaran tentang gejala, penyebab, dan pencegahan ketiga kondisi ini sangat penting untuk mengurangi resiko serta meningkatkan kesehatan dan kesejahteraan masyarakat. Peneliti menyarankan bahwa kepada penelitian berikutnya yang akan mengambil penelitian serupa, bahwasannya pada jumlah atribut agar bisa ditambah, karena untuk melengkapi tingkat kelengkapan informasi pada data contohnya, yaitu menambah atribut fasilitas pada rumah sakit, serta menambah data untuk tahun terbaru agar rasio kunjungan bisa dianalisis lebih baik. Dan untuk penentuan cluster agar bisa disesuaikan dengan tujuan dari penelitian.

UCAPAN TERIMA KASIH

Penulis ingin mengucapkan terimakasih kepada pihak-pihak yang sudah membantu dalam proses penelitian ini. Seluruh dukungan dan kerja sama dari berbagai pihak sangat berharga dan menjadi faktor penting yang memungkinkan penulis untuk menyelesaikan karya ini. Penulis berharap hasil penelitian ini dapat memberikan kontribusi yang berarti bagi ilmu pengetahuan serta membawa manfaat bagi berbagai pihak.

DAFTAR PUSTAKA

- Abduh, R. (2021). Kajian Hukum Rekam Medis Sebagai Alat Bukti Malapraktik Medis. *DE LEGA LATA: Jurnal Ilmu Hukum*, 6(1), 221–233.
- Alghifari, F., & Juardi, D. (2021). Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes. *Jurnal Ilmiah Informatika*, 9(02), 75–81. <https://doi.org/10.33884/jif.v9i02.3755>
- Dinata, R. K., Safwandi, S., Hasdyna, N., & Azizah, N. (2020). Analisis K-Means Clustering pada Data Sepeda Motor. *INFORMAL: Informatics Journal*, 5(1), 10. <https://doi.org/10.19184/isj.v5i1.17071>
- Magdalena, I., Salsabila, A., Krianasari, D. A., & Apsarini, S. F. (2021). Implementasi Model Pembelajaran Daring Pada Masa Pandemi Covid-19 Di Kelas Iii Sdn Sindangsari Iii. *Jurnal Pendidikan Dan Dakwah*, 3(1), 119–128. <https://ejournal.stitpn.ac.id/index.php/pandawa>
- Prasetyo, M. Z., Susanto, E., & Wantoro, A. (2023). SISTEM INFORMASI REKAM MEDIS PASIEN THALASSEMIA (STUDI KASUS : POPTI Cabang BANDAR LAMPUNG). *Jurnal Teknologi Dan Sistem Informasi*, 4(3), 349–355. <https://doi.org/10.33365/jtsi>.
- Putri, A. R. (2021). Optimalisasi Penggunaan Microsoft Excel untuk Pengolahan Nilai Raport di SMAN 1 Ngunut Tulungagung. *Jurnal Pengabdian Kepada Masyarakat*, 3(1), 1–5.
- Rafi Nahjan, M., Nono Heryana, & Apriade Voutama. (2023). Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 101–104. <https://doi.org/10.36040/jati.v7i1.6094>
- Rohma, M. R., Zamzami, A., Utami, H. P., Karsyam, H. A., & Widianingrum, D. C. (2022). Kasus penyakit mulut dan kuku di Indonesia: epidemiologi, diagnosis penyakit, angka kejadian, dampak penyakit, dan

- pengendalian. *Conference of Applied Animal Science Proceeding Series*, 3, 15–22. <https://doi.org/10.25047/animpro.2022.331>
- Zahra, F., Ridla, M. A., & Azise, N. (2024). Implementasi Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Barang (Studi Kasus : Toko Sinar Harahap). *JUSTIFY: Jurnal Sistem Informasi Ibrahimy*, 3(1), 55–65. <https://doi.org/10.35316/justify.v3i1.5335>