

Data Mining dalam Pengelompokan Penyakit Pasien dengan Metode K-Medoids

Dwi Utari Iswavigra^{1✉}, Sarjon Defit², Gunadi Widi Nurcahyo³
^{1,2,3}Universitas Putra Indonesia YPTK Padang

dwi.utari.iswavigra1997@gmail.com

Abstract

Disease is a condition in which the mind and body experience a kind of disturbance, discomfort for those who experience it. Day by day, the number of patients at the Kuok Health Center is increasing with various types of different diseases. The increase number of patients requires the Kuok Health Center staff always update the patient's medical record data. The patient's medical record data is the form of a report containing the number of patients and their illnesses. Based on these data, the Puskesmas needs to find out information about the diseases that are most vulnerable and suffered by many patients. This study aims to classify patient disease data to find out the most common diseases suffered by patients at the Kuok Health Center, Kampar Regency. The grouping of patient disease data is carried out with the Data Mining Clustering and followed by the K-Medoids method. Next, cluster testing is carried out using the Silhouette Coefficient. The results of this study indicate that in cluster 1 the most common disease suffered by patients is non-insulin dependent diabetes mellitus (type II) with a total of 435 cases. In cluster 2, the most common disease suffered by patients was Essential Hypertension (Primary) with a total of 2785 cases. For cluster 3, the most common disease suffered by patients was Vulnus Laseratum, Punctum, with a total of 328 cases. From the cluster results obtained, the results of the Silhouette Coefficient test are 0.900033674.

Keywords: Clustering, Data Mining, K-Medoids, Disease, Silhouette Coefficient.

Abstrak

Penyakit merupakan suatu kondisi di mana pikiran dan tubuh mengalami semacam gangguan dan ketidaknyamanan bagi yang mengalaminya. Semakin hari, jumlah pasien di Puskesmas Kuok semakin meningkat dengan beragam jenis penyakit yang berbeda-beda. Peningkatan jumlah pasien mengharuskan *staff* Puskesmas Kuok harus selalu memperbaharui data rekam medis pasien. Data rekam medis pasien ini berupa laporan yang berisikan jumlah pasien dan penyakit yang diderita. Berdasarkan data tersebut, pihak Puskesmas perlu untuk mengetahui informasi tentang penyakit yang paling rentan dan banyak diderita pasien. Penelitian ini bertujuan untuk mengelompokkan data penyakit pasien untuk mengetahui penyakit yang paling banyak diderita oleh pasien di Puskesmas Kuok Kabupaten Kampar. Pengelompokan data penyakit pasien dilakukan dengan tahapan Data Mining *Clustering* dan dilanjutkan dengan tahapan metode K-Medoids. Selanjutnya, di lakukan pengujian *cluster* menggunakan *Silhouette Coefficient*. Hasil penelitian ini menunjukkan bahwa pada *cluster* 1 penyakit yang paling banyak diderita oleh pasien adalah penyakit Diabetes Melitus tidak bergantung Insulin (tipe II) dengan total kasus sebanyak 435 kasus. Pada *cluster* 2, penyakit yang paling banyak diderita oleh pasien adalah penyakit Hipertensi Esensial (Primer) dengan total kasus sebanyak 2785 kasus. Untuk *cluster* ke 3, penyakit yang paling banyak diderita oleh pasien adalah penyakit Vulnus Laseratum, Punctum, dengan total kasus sebanyak 328 kasus. Dari hasil *cluster* yang diperoleh, maka didapatkan hasil pengujian *Silhouette Coefficient* sebesar 0,900033674.

Kata kunci: *Clustering*, Data Mining, K-Medoids, Penyakit, *Silhouette Coefficient*.

© 2021 JiDT

1. Pendahuluan

Penyakit merupakan suatu kondisi di mana pikiran dan tubuh mengalami semacam gangguan dan ketidaknyamanan bagi yang mengalaminya. Semakin hari, jumlah pasien di Puskesmas Kuok semakin meningkat dengan beragam jenis penyakit yang berbeda-beda. Peningkatan jumlah pasien dapat dilihat pada laporan yang disebut dengan data rekam medis.

Data rekam medis harus selalu diperbaharui sesuai dengan jumlah pasien yang semakin hari semakin meningkat. Pembaharuan data rekam medis pasien mengakibatkan terjadinya penumpukan data. Penumpukan data rekam medis di Puskesmas Kuok hanya sebatas laporan yang berisikan informasi jumlah pasien dan penyakit yang diderita. Berdasarkan data tersebut, pihak Puskesmas memerlukan jenis informasi baru yaitu berupa informasi mengenai penyakit yang

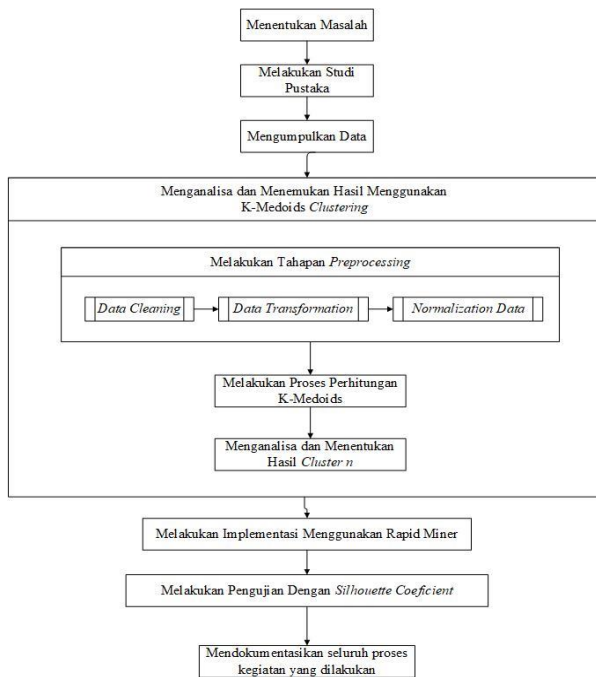
paling rentan dan banyak dialami oleh pasien. Informasi ini dapat membantu dalam mengambil tindakan untuk mengantisipasi bagaimana pencegahan dan pengobatan yang akan dilakukan. Informasi ini juga dapat menjadi bahan pertimbangan dalam perencanaan pembangunan fasilitas, penambahan obat dan penambahan *staff* [1].

Beberapa penelitian mengenai metode K-Medoids dan topik penelitian telah dilakukan sebelumnya, diantaranya yaitu penelitian Sundari dkk (2019), penelitian ini menggunakan metode K-Medoids dalam mengelompokkan data imunisasi campak balita [2]. Penelitian Irwansyah dkk (2020), penelitian ini mengangkat topik tentang pengelompokan pasien penyakit cardiovascular, penelitian ini dilakukan menggunakan metode K-Medoids [3]. Penelitian Lee dkk (2020), penelitian ini mengangkat topik tentang pengelompokan laktasi kurva sapi perah menggunakan metode K-Medoids [4]. Penelitian Samudi dkk (2020), penelitian ini mengangkat topik tentang pengelompokan aplikasi pembelajaran pada masa pandemi Covid-19 dengan menggunakan metode K-Medoids [5]. Penelitian Rifa dkk (2020), topik yang di bahas pada penelitian ini adalah tentang pengelolaan resiko gempa di indonesia menggunakan K-Medoids *Clustering* [6].

Pada penelitian ini terdapat 3 cluster yang akan digunakan yaitu: golongan sebab sakit, jumlah pasien laki-laki dan jumlah pasien perempuan. Berdasarkan masalah di atas, maka perlu dilakukan penelitian dengan judul “Data Mining dalam Pengelompokan Penyakit Pasien dengan Metode K-Medoids”.

2. Metodologi Penelitian

Dalam suatu penelitian, terlebih dahulu dibentuk kerangka kerja penelitian. Kerangka kerja pada penelitian ini meliputi: menentukan masalah, melakukan studi pustaka, mengumpulkan data, melakukan tahapan *pre-processing*, menganalisa dan menemukan hasil menggunakan K-Medoids *Clustering*, melakukan implementasi menggunakan *Rapid Miner*, melakukan pengujian dengan *Silhouette Coefficient*, dan mendokumentasikan seluruh proses kegiatan yang dilakukan. Kerangka kerja penelitian ini dapat dilihat seperti pada Gambar 1.



Gambar 1. Kerangka Kerja Penelitian

Knowledge Discovery in Database (KDD) adalah proses *nontrivial* dalam mengidentifikasi pola yang valid, baru, berpotensi, berguna, dan bermakna dari volume data yang besar [7]. Salah satu metode yang terdapat pada Data Mining adalah *Clustering*.

Clustering digunakan untuk menemukan struktur *cluster* dalam kumpulan data yang dikarakterisasi dengan kesamaan terbesar dalam *cluster* yang sama dan perbedaan terbesar antara *cluster* yang berbeda [8]. Pada *clustering* terdapat beberapa metode salah satu diantaranya yaitu metode K-Medoids. Metode ini muncul untuk mengatasi kekurangan dari metode K-Means yaitu sensitif terhadap *outlier* data [9]. Metode K-Medoids memiliki tujuan yang sama dengan K-Means, yaitu untuk meminimalkan jarak antara titik yang ditentukan sebagai pusat *cluster* dan titik data *cluster* [10]. Metode K-Medoids bekerja dengan cara memisahkan dataset menjadi beberapa kelompok. [11].

Pada metode K-Medoids, ada beberapa langkah perhitungan yang harus diselesaikan yaitu [3][12]:

- Menentukan nilai k, yang akan menjadi jumlah *cluster*
- Pilih secara acak medoid awal sebanyak k dari n data
- Hitung jarak masing-masing objek ke medoids sementara menggunakan persamaan *Euclidian Distance*

$$JED = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} ; 1,2,3, \dots \dots n \quad (1)$$

Dimana *JED* merupakan Jarak *Euclidian Distance*, *x_i* merupakan Nilai medoid awal pada dimensi 1, *y_i*

merupakan Nilai objek pada data pertama dan n merupakan jumlah objek.

- d. Tandai jarak minimum objek ke medoids & hitung total kedekatannya.
- e. Tentukan anggota *cluster* pada medoids awal.
- f. Lakukan iterasi medoids baru dengan memilih kembali secara acak medoids baru sebanyak k dari n data (lakukan tahapan yang sama seperti sebelumnya).
- g. Langkah terakhir yaitu menghitung total simpangan (S) dengan cara mencari selisih antara total kedekatan medoids baru dengan total kedekatan medoids lama. Jika total simpangan $S > 0$, maka proses *clustering* berhenti. Tetapi jika $S < 0$, maka cari medoids baru lagi sampai mendapatkan total simpangan $S > 0$

Berikut ini merupakan Pseudocode K-Medoids *Clustering* yang disajikan pada Algoritma 1.

Algoritma 1: K-Medoids *Clustering*

```

kMedoids (k , dataset):
    int prevLeastDistances
    while true:
        int k medoids as m1, m2, ... mk
        int k clusters as c1, c2, ... ck
        int leastDistances = 0
        for node n in datasets && n not in medoids:
            for m in medoids:
                d = distance(m, n)
                if d = minimum:
                    cm.include(n)
                    leastDistances = leastDistances + d
        if prevLeastDistances > leastDistances:
            return clusters
        else : prevLeastDistances = leastDistances
    
```

3. Hasil dan Pembahasan

Selanjutnya, dilakukan proses di mana seluruh kegiatan mulai dari analisa data sampai penyelesaian dan pembahasan perhitungan metode. Pada tahapan ini seluruh proses akan terlihat. Ada 3 hal yang harus dilakukan, yaitu: Analisa Data, *Preprocessing* Data dan Perhitungan K-Medoids *Clustering*.

3.1 Analisa Data

Pada penelitian ini data yang digunakan merupakan laporan data penyakit Puskesmas Kuok pada tahun 2020. Laporan ini berisikan golongan sebab sakit, jumlah pasien laki-laki & perempuan. Pada laporan data penyakit Puskesmas Kuok pada tahun 2020, terdapat 17 golongan sebab sakit di mana ke-17 golongan sebab sakit tersebut masing-masing memiliki daftar nama-nama penyakit yang berbeda-beda. Total keseluruhan penyakit yang terdapat pada data tersebut berjumlah 155 penyakit.

3.2 Preprocessing Data

Sebelum memulai proses perhitungan K-Medoids, terlebih dahulu dilakukan tahapan *Preprocessing*. Pada tahapan *Preprocessing*, terdapat 3 proses yang dilakukan yaitu: *Data Cleaning*, *Tranformasi Data*, & *Normalisasi Data*.

3.2.1. Data Cleaning

Pada tahap *Data Cleaning* dilakukan proses penghapusan terhadap data-data yang rusak, kosong, duplikat, tidak lengkap dan bentuk kesalahan-kesalahan data lainnya. Pada penelitian ini, data yang digunakan sebanyak 155 data penyakit pasien. Setelah dilakukan tahap *cleaning*, didapatkan sebanyak 66 data yang sempurna, lengkap dan tidak rusak. Selanjutnya 66 data tersebut akan masuk ke tahap selanjutnya, yaitu *Data Transformation* yang disajikan pada Tabel 1.

Tabel 1. Hasil *Cleaning* Sampel Data

Kode	Nama Penyakit	Laki-Laki	Perempuan
A 03.9, A 06.0	Disentri basiler, disentri amuba	3	2
A 09	Gastroenteritis (termasuk kolera, giardiadis)	134	211
A 15.0	Tuberkulosis paru tanpa komplikasi	47	64
B 00.9	Herpes simpleks tanpa komplikasi	5	12
B 01.9	Varisela/cacar air tanpa komplikasi	7	6
B 02.9	Herpes zoster tanpa komplikasi	1	2
B 26	Mumps/parotitis (gondong)	4	3
B 35	Dermatofitosis (penyakit jamur superfisial)	35	77
B 76	Penyakit cacing tambang	1	1
B 76.9	<i>Cutaneous larva migran</i>	1	1
B 86	Skabies	21	21
E 10	Diabetes Melitus bergantung Insulin (tipe I)	1	3
E 11	Diabetes Melitus tidak bergantung Insulin (tipe II)	146	289
W 57	Reaksi gigitan serangga	7	12

3.2.2. Tranformasi Data

Pada tahap *Data Transformation*, dilakukan proses merubah dan mentransformasi format data yang menjadi format data dalam bentuk numerik / membuat inisialisasi. Seperti pada penelitian ini, setiap data golongan sebab sakit diinisialisasi kedalam bentuk angka yang disajikan pada Tabel 2.

Tabel23. Tranformasi Golongan Sebab Sakit

Golongan Sebab Sakit	Inisialisasi
Penyakit Infeksi Dan Parasit Tertentu	1
Neoplasma	2
Penyakit Darah Dan Alat Pembentuk Darah	3
Penyakit Endokrin, Gizi Dan Metabolik	4
Gangguan Mental Dan Perilaku	5
Penyakit Susunan Saraf Pusat	6
Penyakit Mata Dan Adneksa	7
Penyakit Telinga Dan Prosesus Mastoideus	8
Penyakit Sistem Peredaran Darah	9
Penyakit Sistem Pernafasan	10
Penyakit Sistem Pencernaan	11
Penyakit Kulit Dan Jaringan Subkutan	12
Penyakit Sistem Muskuloskeletal	13
Peny. Sistem Saluran Kemih Dan Kelamin	14
Kehamilan, Persalinan Dan Nifas	15
Tanda, Gejala&Penemuan Lab Klinis	16
Kecelakaan Dan Keracunan Dan Akibat Eksternal	17

3.2.3. Normalisasi Data

Selanjutnya dilakukan normalisasi data dengan cara mengurangi nilai bobot per cluster dengan nilai Min pada cluster tersebut kemudian membaginya dengan hasil pengurangan antara nilai Max dan Min cluster, sehingga didapatkan hasil normalisasi data seperti pada Tabel 3.

Tabel 3. Hasil Normalisasi Sampel Data

Kode	Nama Penyakit	Golongan sebab sakit	Laki-Laki	Perempuan
A 03.9,	Disentri basiler,	0	0,00189	0,0005
A 06.0	disentri amuba			
A 09	Gastroenteritis (termasuk kolera, giardiadisi)	0	0,12594	0,1192
A 15.0	Tuberkulosis paru tanpa komplikasi	0	0,04356	0,0357
B 00.9	Herpes simpleks tanpa komplikasi	0	0,00378	0,0062
B 01.9	Varisela/cacar air tanpa komplikasi	0	0,00568	0,0028
B 02.9	Herpes zoster tanpa komplikasi	0	0	0,0005
B 26	Mumps/parotitis (gondong)	0	0,00284	0,0011
B 35	Dermatofitosis (penyakit jamur superfisial)	0	0,03219	0,0431
B 76	Penyakit cacang tambang	0	0	0
B 76.9	Cutaneus larva migran	0	0	0
B 86	Skabies	0	0,01893	0,0113
E 10	Diabetes Melitus bergantung Insulin (tipe I)	0,1875	0	0,0011
E 11	Diabetes Melitus tidak bergantung Insulin (tipe II)	0,1875	0,13731	0,1635
W 57	Reaksi gigitan serangga	1	0,00568	0,0062

Selanjutnya, data yang telah di normalisasi dapat digunakan untuk memulai tahapan perhitungan K-Medoids Clustering.

3.3 Perhitungan K-Medoids Clustering

Berikut ini merupakan tahapan perhitungan K-Medoids Clustering.

3.3.1. Menentukan nilai k, jumlah cluster

Pada penelitian ini, data penyakit dibagi menjadi 3 cluster, yaitu: cluster 1 golongan sebab sakit, cluster 2 jumlah pasien laki-laki, dan cluster 3 jumlah pasien perempuan.

3.3.2. Pilih medoid awal secara acak

Selanjutnya, memilih medoid awal dari setiap cluster secara acak dari data yang telah di normalisasi yang disajikan pada Tabel 4.

Tabel 4. Medoids Awal 1

Golongan sebab sakit	Laki-laki	Perempuan
1	0,003788	0,003975
1	0,230114	0,047132
1	0,005682	0,006246

3.3.3. Hitung jarak masing-masing objek dengan menggunakan persamaan Euclidean Distance seperti pada Persamaan 1. Berikut ini adalah contoh perhitungan pada data pertama.

a. Data pertama ke pusat cluster 1.

$$JED(X_1C_1)=\sqrt{(0-1)^2+(0,001894-0,003788)^2+(0,000568-0,003975)^2}=1,000008$$

b. Data pertama ke pusat cluster 2.

$$JED(X_1C_2)=\sqrt{(0-1)^2+(0,001894-0,230114)^2+(0,000568-0,047132)^2}=1,026768$$

c. Data pertama ke pusat cluster 3.

$$JED(X_1C_3)=\sqrt{(0-1)^2+(0,001894-0,005682)^2+(0,000568-0,006246)^2}=1,000023$$

Lakukan perhitungan yang sama seperti di atas untuk data kedua sampai data terakhir, sehingga akan dihasilkan jarak masing-masing objek ke medoids sementara seperti pada Tabel 5.

Tabel 5. Jarak Masing-Masing Objek Ke Medoid Awal/Sementara

C1	C2	C3	Jarak Min	Cluster
1,000008	1,0267679	1,0000232	1,000007	1
1,0140075	1,0079939	1,0135253	1,007993	2
1,0012957	1,0173155	1,0011527	1,001152	3
1,0000025	1,0261067	1,0000017	1,000001	3
1,0000024	1,0258321	1,0000058	1,000002	1
1,0000129	1,0271906	1,0000322	1,000012	1
1,0000044	1,0265323	1,0000170	1,000004	1
1,0011704	1,0194051	1,0010321	1,001032	3
1,0000150	1,0272165	1,0000356	1,000015	1
1,0000150	1,0272165	1,0000356	1,000015	1
1,0001420	1,0226800	1,0001009	1,000100	3
0,8125137	0,8457093	0,8125359	0,812513	1
0,8387172	0,8260267	0,8379885	0,826026	2
0,0029574	0,2281256	0	0	3
Total Kedekatan Jarak Min			37,40430	
			755	

3.3.4. Tandai jarak minimum objek ke medoids & hitung total kedekatannya.

Berdasarkan jarak minimum objek ke medoids awal / sementara pada tabel 4, maka dapat dihitung total kedekatannya dengan cara menjumlahkan keseluruhan jarak minimum pada tabel 4, sehingga didapat total kedekatannya sebesar 37,40430755.

3.3.5. Lakukan iterasi medoids baru dengan memilih kembali secara acak medoids baru yang disajikan pada Tabel 6.

Tabel 6. Medoids Baru 1

Golongan sebab sakit	Laki-laki	Perempuan
0,8125	0,025568182	0,023282226
0,7500	0,462121212	0,554230551
1,0000	0,005681818	0,006246451

Ulangi kembali langkah c sampai d, sehingga didapatkan total kedekatannya sebesar 25,55033652.

3.3.6. Hitung total simpangan (S).

Selanjutnya, menghitung total simpangan (S) dengan cara mencari selisih antara total kedekatan medoids baru dengan total kedekatan medoids lama. Jika total simpangan $S > 0$, maka proses *clustering* berhenti. Tetapi jika $S < 0$, maka cari medoids baru lagi sampai mendapatkan total simpangan $S > 0$.

$$S = 25,55033652 - 37,40430755 = -11,854$$

Berdasarkan hasil total simpangan di atas, diketahui bahwa hasilnya negatif yang menandakan $S < 0$. Maka, tentukan kembali medoids baru secara acak dan lakukan kembali langkah c sampai d yang disajikan pada Tabel 7.

Tabel 7. Medoids Baru 2

Golongan sebab sakit	Laki-laki	Perempuan
0,9375	0,006628788	0,001703578
1	0,013257576	0,005110733
1	0,015151515	0,001703578

Ulangi kembali langkah c sampai d, sehingga didapatkan total kedekatannya sebesar 34,41620166.

Selanjutnya, hitung kembali total simpangan (S).

$$S = 34,41620166 - 25,55033652 = 8,865865139$$

Berdasarkan hasil total simpangan di atas, diketahui bahwa $S > 0$. Di karenakan $S > 0$, maka proses *clustering* dihentikan sehingga didapat hasil *cluster* yang disajikan pada Tabel 8.

Tabel 8. Sampel Hasil Cluster

Kode	Nama Penyakit	Golongan sebab sakit	Laki-Laki	Perempuan	Cluster
A					
03.9.	Disentri basiler,				
A	disentri amuba	1	3	2	1
06.0					
	Gastroenteritis				
A 09	(termasuk kolera,	1	134	211	1
	giardiadis)				
A	Tuberkulosis				
15.0	paru tanpa komplikasi	1	47	64	1
B	Herpes				
00.9	simpleks tanpa komplikasi	1	5	12	1
B	Varisela/cacar				
01.9	air tanpa komplikasi	1	7	6	1
B	Herpes zoster				
02.9	tanpa komplikasi	1	1	2	1
B 26	Mumps/parotitis (gondong)	1	4	3	1
B 35	Dermatofitosis (penyakit jamur superfisial)	1	35	77	1
B 76	Penyakit cacing tambang	1	1	1	1
B	Cutaneus larva migran	1	1	1	1
76.9					
B 86	Skabies	1	21	21	1
	Diabetes				
E 10	Melitus bergantung Insulin (tipe I) Diabetes	4	1	3	1
E 11	Melitus tidak bergantung Insulin (tipe II)	4	146	289	1
W	Reaksi gigitan serangga	17	7	12	3
57					

Berdasarkan hasil perhitungan metode K-Medoids yang telah dilakukan, didapat hasil *cluster* 1 sebanyak 53 data, *cluster* 2 sebanyak 5 data, dan *cluster* 3 sebanyak 8 data yang disajikan pada Tabel 9.

Tabel 9. Kesimpulan Hasil Cluster

Cluster	Kesimpulan
	Jumlah data = 53
1	a. Golongan sebab sakit yang paling banyak dialami pasien adalah penyakit infeksi & parasite tertentu. b. Penyakit yang paling banyak diderita oleh pasien adalah penyakit Diabetes Melitus tidak bergantung Insulin (tipe II) dengan total kasus sebanyak 435 kasus. c. Pasien di dominasi oleh perempuan d. Jumlah pasien laki-laki sebanyak 1955 orang e. Jumlah pasien perempuan sebanyak 2548 orang
2	Jumlah data = 5

- Pada cluster 2, didominasi sama rata masing-masing 1 data dari golongan sebab penyakit sistem peredaran darah, penyakit sistem pernafasan, penyakit sistem pencernaan, penyakit kulit dan jaringan subkutan dan penyakit sistem muskuloskeletal.
- Penyakit yang paling banyak diderita oleh pasien adalah penyakit Hipertensi Esensial (Primer) dengan total kasus sebanyak 2785 kasus.
- Pasien di dominasi oleh perempuan.
- Jumlah pasien laki-laki sebanyak 3454 orang.
- Jumlah pasien perempuan sebanyak 5712 orang.

Jumlah data = 8

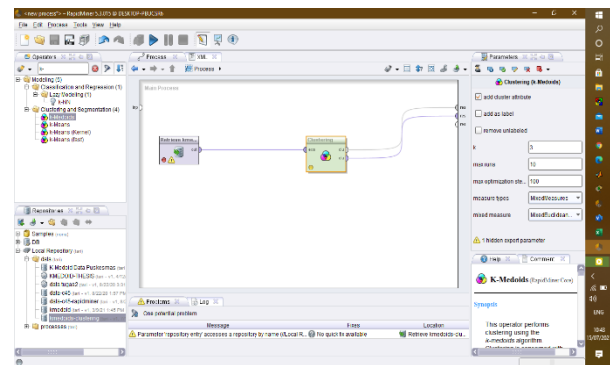
- Untuk cluster ke 3, didominasi oleh golongan sebab penyakit kecelakaan dan keracunan akibat eksternal dengan total 5 data.
 - Penyakit yang paling banyak diderita oleh pasien adalah penyakit Vulnus Laseratum, Punctum, dengan total kasus sebanyak 328 kasus.
 - Pasien di dominasi oleh laki-laki.
 - Jumlah pasien laki-laki sebanyak 325 orang.
 - Jumlah pasien perempuan sebanyak 188 orang.

Berdasarkan hasil kesimpulan *cluster* pada Tabel 10, menunjukkan bahwa pada *cluster* 1 penyakit yang paling banyak diderita oleh pasien adalah penyakit Diabetes Melitus tidak bergantung Insulin (tipe II) dengan total kasus sebanyak 435 kasus. Pada *cluster* 2, penyakit yang paling banyak diderita oleh pasien adalah penyakit Hipertensi Esensial (Primer) dengan total kasus sebanyak 2785 kasus. Untuk *cluster* ke 3, penyakit yang paling banyak diderita oleh pasien adalah penyakit Vulnus Laseratum, Punctum, dengan total kasus sebanyak 328 kasus.

3.4. Implementasi Rapid Miner

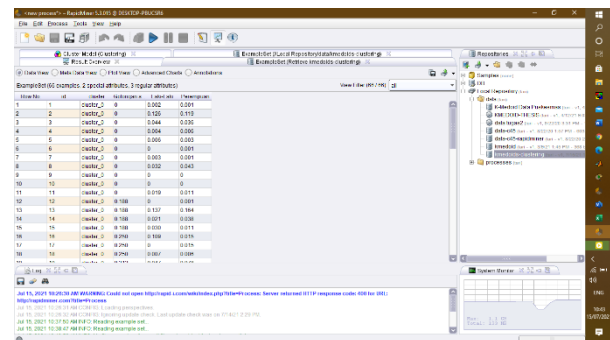
Pada penelitian ini, *software* Rapid Miner 5.3 digunakan untuk mengimplementasikan proses K-Medoid *clustering* dalam mengelompokkan penyakit pasien Puskesmas Kuok. Berikut ini proses K-Medoids yang dilakukan menggunakan Rapid Miner 5.3.

- Import* data yang ingin diolah dengan cara memilih menu *file* di sudut kiri atas kemudian pilih *import excel sheet* pada pilihan *import* data. Selanjutnya, pilih file yang akan di *import* dan klik *next* untuk lanjut ke tahap berikutnya.
- Setelah data di *import*, *drag & drop* data dan algoritma K-Medoid ke lembar kerja design untuk melakukan proses *clustering*.
- Hubungkan operator data dan algoritma K-Medoids ke *result*.

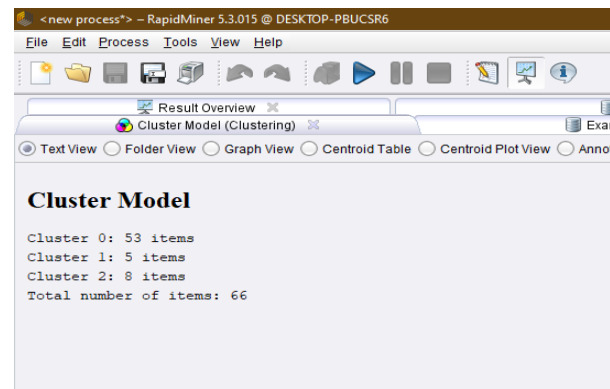


Gambar 2. Tampilan Hubungan Operator Data dan Algoritma K-Medoids Ke Result

Gambar 2 menyajikan hubungan ini terbentuk dengan cara menarik garis pada *retrieve* data dan dihubungkan ke algoritma K-Medoids. Kemudian, tarik kembali garis pada Algoritma K-Medoids *Clustering* ke *result*. Selanjutnya, atur parameter pada bagian samping kanan sesuai dengan kebutuhan. Pada penelitian ini nilai parameter *k* diisi dengan jumlah *cluster* yang digunakan yaitu 3. Terakhir, klik tombol *process* dan tunggu hingga proses *clustering* selesai. Setelah proses *clustering* K-Medoid selesai, maka akan muncul hasil *cluster* seperti pada Gambar 3. dan Gambar 4.



Gambar 3. Hasil Cluster Data View



Gambar 4. Hasil Cluster Text View

3.5. Pengujian Silhouette Coefficient

Pengujian *Silhouette Coefficient* merupakan pengujian yang dilakukan untuk menentukan apakah *cluster* yang diperoleh termasuk dalam kategori *cluster* baik atau tidak. Hasil *cluster* dikatakan baik apabila nilai *Silhouette Coefficient* yang diperoleh mendekati angka 1

(satu). Berikut ini merupakan ukuran nilai *Silhouette Coeficient* yang disajikan pada Tabel 10.

Tabel 10. Ukuran Nilai *Silhouette Coeficient*

<i>Silhouette Coeficient</i>	Keterangan Interpretasi
$0.7 < SC \leq 1.0$	<i>Stronge Structure</i>
$0.5 < SC \leq 0.7$	<i>Medium Structure</i>
$0.25 < SC \leq 0.5$	<i>Weak Structure</i>
$SC \leq 0.25$	<i>No Structure</i>

Berdasarkan Tabel 10 dapat disimpulkan bahwa semakin nilai *silhouette coeficient* mendekati angka 1, maka semakin baik *cluster* tersebut dan semakin nilai *silhouette coeficient* mendekati angka -1, maka semakin buruk *cluster* tersebut. Berikut ini merupakan penjabaran dari perhitungan *silhouette coeficient* untuk menguji kualitas *cluster* pada penelitian ini yang disajikan pada Tabel 11.

Tabel 11. Sampel Nilai Min / Kedekatan Antar *Cluster*

Nama Penyakit	C1	C2	C3	Min	<i>Cluster</i>
Disentri basiler, disentri amuba	0,8131	1,0396	1,00002	0,813162138	1
Gastroenteritis (termasuk kolera, giardiadis)	0,8242	0,9299	1,013525	0,824282747	1
Tuberkulosis paru tanpa komplikasi	0,8127	1,0032	1,001153	0,812795208	1
Herpes simpleks tanpa komplikasi	0,8129	1,0357	1,000002	0,812970386	1
Varisela/cacar air tanpa komplikasi	0,813	1,0367	1,000006	0,813000388	1
Herpes zoster tanpa komplikasi	0,8132	1,0404	1,000032	0,813219481	1
Mumps/parotitis (gondong)	0,8131	1,0389	1,000017	0,813119454	1
Dermatofitosis (penyakit jamur superfisial)	0,8127	1,0042	1,001032	0,812770084	1
Penyakit cacing tambang	0,8132	1,0407	1,000036	0,81323554	1
<i>Cutaneus larva migran</i>	0,8132	1,0407	1,000036	0,81323554	1
Skabies	0,8126	1,0264	1,000101	0,812614544	1
Diabetes Melitus bergantung Insulin (tipe I)	0,625915	0,914263	0,812536	0,62591469	1
Diabetes Melitus tidak bergantung Insulin (tipe II)	0,650219	0,757987	0,837989	0,650218869	1
Reaksi gigitan serangga	0,18932	0,755727	0	0	3

a. Hitung jarak rata-rata antara objek ke-i dan semua objek di *cluster* 1,2, dan 3

Cluster 1 : 53 Items

$$a(i) = \frac{\sqrt{(0,8131 - 0,8131)^2 + (0,8242 - 0,8131)^2 + (0,8127 - 0,8131)^2 + (0,8129 - 0,8131)^2 + (0,8130 - 0,8131)^2 + \dots + (0 - 0,8131)^2}}{53}$$

$$a(i) = 3,23629 / 53 = 0,06106$$

Cluster 2 : 5 Items

$$a(i) = \frac{\sqrt{(0,7189 - 0,7189)^2 + (0,6121 - 0,7189)^2 + (0,1590 - 0,7189)^2 + (0,3217 - 0,7189)^2 + (0 - 0,7189)^2}}{5}$$

$$a(i) = 0,99983 / 5 = 0,19997$$

Cluster 2 : 8 Items

$$a(i) = \frac{\sqrt{(0,0628 - 0,0628)^2 + (0,0711 - 0,0628)^2 + (0,0626 - 0,0628)^2 + (0,0076 - 0,0628)^2 + (0,0105 - 0,0628)^2 + (0,0029 - 0,0628)^2 + (0,2281 - 0,0628)^2 + (0 - 0,0628)^2}}{8}$$

$$a(i) = 0,20176 / 8 = 0,02522$$

b. Hitung jarak rata-rata antara objek ke-i dan objek lain di *cluster* lain, kemudian ambil nilai terkecilnya dan tentukan perbedaan rata rata antara objek i untuk *cluster* yang paling dekat.

c. Jarak data pada *cluster* 1 ke *cluster* 2 dan *cluster* 3:

$$b(i)_{12} = \frac{\sqrt{(0,8131 - 0,7189)^2 + (0,8242 - 0,7189)^2 + (0,8127 - 0,7189)^2 + (0,8129 - 0,7189)^2 + (0,8130 - 0,7189)^2 + \dots + (0 - 0,7189)^2}}{12}$$

$$b(i)_{12} = 2,6824$$

Lakukan perhitungan yang sama untuk data kedua dan seterusnya pada *cluster* 1 ke *cluster* 2, sehingga didapatkan jarak data pada *cluster* 1 ke *cluster* 2 yang disajikan pada Tabel 12.

Tabel 12. Jarak Data Pada Cluster 1 Ke Cluster 2

Jarak Data Pada Cluster 1 Ke Cluster 2	Hasil
Jarak Data Cluster 1 Ke Data 1 Cluster 2	2,682405
Jarak Data Cluster 1 Ke Data 2 Cluster 2	2,138833
Jarak Data Cluster 1 Ke Data 3 Cluster 2	2,563076
Jarak Data Cluster 1 Ke Data 4 Cluster 2	1,839089
Jarak Data Cluster 1 Ke Data 5 Cluster 2	3,524778
Jumlah	12,7482
$b(i)_{12}$	12,7482 / 5 = 2,54964

$$b(i)_{13} = \frac{\sqrt{(0,8131 - 0,0628)^2 + (0,8242 - 0,0628)^2 + (0,8127 - 0,0628)^2 + (0,8129 - 0,0628)^2 + (0,8130 - 0,0628)^2 + \dots + (0 - 0,0628)^2}}{13}$$

$$b(i)_{13} = 3,127753$$

Lakukan perhitungan yang sama untuk data kedua dan seterusnya pada *cluster* 1 ke *cluster* 3, sehingga didapatkan jarak data pada *cluster* 1 ke *cluster* 3 yang disajikan pada Tabel 13.

Tabel 13. Jarak Data Pada Cluster 1 Ke Cluster 3

Jarak Data Pada Cluster 1 Ke Cluster 3	Hasil
Jarak Data Cluster 1 Ke Data 1 Cluster 3	3,127753
Jarak Data Cluster 1 Ke Data 2 Cluster 3	3,076259
Jarak Data Cluster 1 Ke Data 3 Cluster 3	3,128684
Jarak Data Cluster 1 Ke Data 4 Cluster 3	3,475466
Jarak Data Cluster 1 Ke Data 5 Cluster 3	3,457226
Jarak Data Cluster 1 Ke Data 6 Cluster 3	3,505714
Jarak Data Cluster 1 Ke Data 7 Cluster 3	2,207463
Jarak Data Cluster 1 Ke Data 8 Cluster 3	3,524778
Jumlah	25,50334
$b(i)_{13}$	3,187918

Berdasarkan nilai $b(i)_{12}$ dan $b(i)_{13}$, maka tentukan nilai $b(i)$ dengan cara mengambil nilai minimum antara kedua *cluster* tersebut.

$$b(i) = 2,54964$$

d. Jarak data pada *cluster 2* ke *cluster 1* dan *cluster 3* :

$$b(i)_{21} = \frac{\sqrt{(0,7189 - 0,8131)^2 + (0,6121 - 0,8131)^2 + (0,1590 - 0,8131)^2 + (0,3217 - 0,8131)^2 + (0 - 0,8131)^2}}{2}$$

$$b(i)_{21} = 1,174676$$

Lakukan perhitungan yang sama untuk data kedua dan seterusnya pada *cluster 2* ke *cluster 1*, sehingga didapatlah jarak data pada *cluster 2* ke *cluster 1* yang disajikan pada Tabel 14.

Tabel 14. Jarak Sampel Data Pada Cluster 2 Ke Cluster 1

Jarak Data Pada Cluster 2 Ke Cluster 1	Hasil
Jarak Data Cluster 2 Ke Data 1 Cluster 1	1,174676
Jarak Data Cluster 2 Ke Data 2 Cluster 1	1,196081
Jarak Data Cluster 2 Ke Data 3 Cluster 1	1,173972
Jarak Data Cluster 2 Ke Data 4 Cluster 1	1,174308
Jarak Data Cluster 2 Ke Data 5 Cluster 1	1,174365
Jarak Data Cluster 2 Ke Data 6 Cluster 1	1,174786
Jarak Data Cluster 2 Ke Data 7 Cluster 1	1,174594
Jarak Data Cluster 2 Ke Data 8 Cluster 1	1,173924
Jarak Data Cluster 2 Ke Data 53 Cluster 1	1,010216
Jumlah	42,39041
$b(i)_{21}$	0,799819

$$b(i)_{23} = \frac{\sqrt{(0,7189 - 0,0628)^2 + (0,6121 - 0,0628)^2 + (0,1590 - 0,0628)^2 + (0,3217 - 0,0628)^2 + (0 - 0,0628)^2}}{2}$$

$$b(i)_{23} = 0,901446$$

Lakukan perhitungan yang sama untuk data kedua dan seterusnya pada *cluster 2* ke *cluster 3*, sehingga didapatlah jarak data pada *cluster 2* ke *cluster 3* yang disajikan pada Tabel 15.

Tabel 15. Jarak Data Pada Cluster 2 Ke Cluster 3

Jarak Data Pada Cluster 2 Ke Cluster 3	Hasil
Jarak Data Cluster 2 Ke Data 1 Cluster 3	0,901446
Jarak Data Cluster 2 Ke Data 2 Cluster 3	0,887624
Jarak Data Cluster 2 Ke Data 3 Cluster 3	0,901697
Jarak Data Cluster 2 Ke Data 4 Cluster 3	0,996529
Jarak Data Cluster 2 Ke Data 5 Cluster 3	0,991477
Jarak Data Cluster 2 Ke Data 6 Cluster 3	1,00492
Jarak Data Cluster 2 Ke Data 7 Cluster 3	0,67382
Jarak Data Cluster 2 Ke Data 8 Cluster 3	1,010216
Jumlah	7,367729
$b(i)_{23}$	0,920966

Berdasarkan nilai $b(i)_{21}$ dan $b(i)_{23}$, maka tentukan nilai $b(i)$ dengan cara mengambil nilai minimum antara kedua *cluster* tersebut.

$$b(i) = 0,799819$$

e. Jarak data pada *cluster 3* ke *cluster 1* dan *cluster 2* :

$$b(i)_{31} = \frac{\sqrt{(0,0628 - 0,8131)^2 + (0,0711 - 0,8131)^2 + (0,0626 - 0,8131)^2 + (0,0076 - 0,8131)^2 + (0,0105 - 0,8131)^2 + (0,0029 - 0,8131)^2 + (0,2281 - 0,8131)^2 + (0 - 0,8131)^2}}{2}$$

$$b(i)_{32} = 2,15169$$

Lakukan perhitungan yang sama untuk data kedua dan seterusnya pada *cluster 3* ke *cluster 1*, sehingga didapatlah jarak data pada *cluster 3* ke *cluster 1* yang disajikan pada Tabel 16.

Tabel 16. Jarak Sampel Data Pada Cluster 3 Ke Cluster 1

Jarak Data Pada Cluster 3 Ke Cluster 1	Hasil
Jarak Data Cluster 3 Ke Data 1 Cluster 1	2,151694
Jarak Data Cluster 3 Ke Data 2 Cluster 1	2,183012
Jarak Data Cluster 3 Ke Data 3 Cluster 1	2,150661
Jarak Data Cluster 3 Ke Data 4 Cluster 1	2,151154
Jarak Data Cluster 3 Ke Data 5 Cluster 1	2,151238
Jarak Data Cluster 3 Ke Data 6 Cluster 1	2,151855
Jarak Data Cluster 3 Ke Data 7 Cluster 1	2,151574
Jarak Data Cluster 3 Ke Data 8 Cluster 1	2,15059
Jarak Data Cluster 3 Ke Data 53 Cluster 1	0,255269
Jumlah	59,38123
$b(i)_{31}$	1,1204

$$b(i)_{32} = \frac{\sqrt{(0,0628 - 0,7189)^2 + (0,0711 - 0,7189)^2 + (0,0626 - 0,7189)^2 + (0,0076 - 0,7189)^2 + (0,0105 - 0,7189)^2 + (0,0029 - 0,7189)^2 + (0,2281 - 0,7189)^2 + (0 - 0,7189)^2}}{2}$$

$$b(i)_{32} = 1,88662$$

Lakukan perhitungan yang sama untuk data kedua dan seterusnya pada *cluster 3* ke *cluster 2*, sehingga didapatlah jarak data pada *cluster 3* ke *cluster 2* yang disajikan pada Tabel 17.

Tabel 17. Jarak Data Pada Cluster 3 Ke Cluster 2

Jarak Data Pada Cluster 3 Ke Cluster 2	Hasil
Jarak Data Cluster 3 Ke Data 1 Cluster 2	1,886617
Jarak Data Cluster 3 Ke Data 2 Cluster 2	1,586638
Jarak Data Cluster 3 Ke Data 3 Cluster 2	0,354413
Jarak Data Cluster 3 Ke Data 4 Cluster 2	0,778815
Jarak Data Cluster 3 Ke Data 5 Cluster 2	0,255269
Jumlah	4,861752
$b(i)_{32}$	0,97235

Berdasarkan nilai $b(i)_{31}$ dan $b(i)_{32}$, maka tentukan nilai $b(i)$ dengan cara mengambil nilai minimum antara kedua *cluster* tersebut.

$$b(i) = 0,97235$$

f. Hitung nilai *Silhouette Coefficient*.

Nilai *Silhouette Coefficient* pada *cluster 1* :

$$s(i) = 1 - \frac{0,06106}{2,54964}$$

$$s(i) = 0,976050664$$

Nilai *Silhouette Coefficient* pada *cluster 2* :

$$s(i) = 1 - \frac{0,19997}{0,799819}$$

$$s(i) = 0,749987149$$

Nilai *Silhouette Coefficient* pada *cluster 3* :

$$s(i) = 1 - \frac{0,02522}{0,97235}$$

$$s(i) = 0,97406321$$

- g. Hitung rata-rata *Silhouette Coefficient (SC)* semua cluster.

$$SC = \frac{0,976050664+0,749987149+0,97406321}{3}$$
$$SC = 0,900033674$$

Berdasarkan hasil *Silhouette Coefficient (SC)* di atas, maka dapat disimpulkan bahwa cluster yang digunakan termasuk dalam kategori baik dan kuat dengan nilai sebesar 0,900033674.

4. Kesimpulan

Berdasarkan hasil analisa dan pengujian pada pembahasan sebelumnya, maka diperoleh kesimpulan bahwa hasil *Cluster* menunjukkan bahwa pada *Cluster* 1, Diabetes Melitus tidak bergantung Insulin (tipe II) merupakan penyakit yang paling banyak diderita oleh pasien dengan total kasus sebanyak 435 kasus. Pada *cluster* 2, Hipertensi Esensial (Primer) merupakan penyakit yang paling banyak diderita oleh pasien dengan total kasus sebanyak 2785 kasus. Dan pada *cluster* 3, Vulnus Laseratum, Punctum merupakan penyakit yang paling banyak diderita oleh pasien dengan total kasus sebanyak 328 kasus.

Daftar Rujukan

- [1] Andini, A.D dan Arifin, T. (2020). Implementasi algoritma k-medoids untuk klasterisasi data penyakit pasien di rsud kota bandung. *Jurnal responsif*, Vol. 2 No.2, pp. 128-138. E-ISSN: 2685-6964 128.
- [2] Sundari, S., Damanik, I. S., Windarto, A. P., Tambunan, H. S., Jalaluddin, J., & Wanto, A. (2019). Analisis K-Medoids Clustering Dalam Pengelompokan Data Imunisasi Campak Balita di Indonesia. *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 1, 687. doi:10.30645/senaris.v1i0.75
- [3] Irwansyah, E., Salim Pratama, E., & Ohlyver, M. (2020). *Clustering of Cardiovascular Disease Patients Using Data Mining Techniques with Principal Component Analysis and K-Medoids*. doi:10.20944/preprints202008.0074.v
- [4] Lee, M., Lee, S., Park, J., & Seo, S. (2020). *Clustering and Characterization of the Lactation Curves of Dairy Cows Using K-Medoids Clustering Algorithm*. *Animals*, 10(8), 1348. doi:10.3390/ani10081348
- [5] Samudi, S., Widodo, S., & Brawijaya, H. (2020). *The K-Medoids Clustering Method for Learning Applications during the COVID-19 Pandemic*. *Sinkron*, 5(1), 116. doi:10.33395/sinkron.v5i1.10649
- [6] Rifa, I. H., Pratiwi, H., & Respatiwan, R. (2020). *Clustering Of Earthquake Risk in Indonesia Using K-Medoids and K-Means Algorithms*. *Media statistika*, 13(2), 194–205. doi:10.14710/medstat.13.2.194-205
- [7] Yan, H., Yang, N., Peng, Y., & Ren, Y. (2020). *Data mining in the construction industry: Present status, opportunities, and future trends*. *Automation in Construction*, 119, 103331. doi:10.1016/j.autcon.2020.103331
- [8] Sinaga, K. P., & Yang, M.-S. (2020). *Unsupervised K-Means Clustering Algorithm*. *IEEE Access*, 8, 80716–80727. doi:10.1109/access.2020.2988796
- [9] Rahman, F., Ridho, I. I., Muflih, M., Pratama, S., Raharjo, M. R., & Windarto, A. P. (2020). *Application of Data Mining Technique using K-Medoids in the case of Export of Crude Petroleum*

Materials to the Destination Country. *IOP Conference Series: Materials Science and Engineering*, 835, 012058. doi:10.1088/1757-899x/835/1/012058

- [10] Hooshyar, D., Yang, Y., Pedaste, M., & Huang, Y.-M. (2020). *Clustering Algorithms in an Educational Context: An Automatic Comparative Approach*. *IEEE Access*, 8, 146994–147014. doi:10.1109/access.2020.3014948
- [11] Muruganandham, S. K., Sobyta, D., Nallusamy, S., Mandal, D. K., & Chakraborty, P. S. (2018). *Study on Leaf Segmentation Using K-Means and K-Medoid Clustering Algorithm for Identification of Disease*. *Indian Journal of Public Health Research & Development*, 9(5), 289. doi:10.5958/0976-5506.2018.00456.4
- [12] Desi Asima Silitonga., Agus Perdana Windarto., & Dedy Hartama, Sumarno. (2019). Penerapan Metode K-Medoid pada Pengelompokan Rumah Tangga Dalam Perlakuan Memilah Sampah Menurut Provinsi. *Seminar Nasional Sains & Teknologi Informasi (SENSASI)*, ISBN: 978-602-52720-2-8 Hal: 313 – 318