IMPLEMENTASI ALGORITMA K-MEDOIDS UNTUK KLASTERISASI DATA PENYAKIT PASIEN DI RSUD KOTA BANDUNG

Asri Dwi Andini[1], Toni Arifin[2]

[1]Universitas Adhirajasa Reswara Sanjaya
e-mail: asridwiandini98@gmail.com

[2]Universitas Adhirajasa Reswara Sanjaya
e-mail: toni.arifin@ars.ac.id

Abstrak

Penyakit adalah suatu keadaan abnormal dimana tubuh ataupun pikiran mengalami ketidaknyamanan atau disfungsi terhadap orang yang dipengaruhinya. Setiap harinya jumlah pasien di rumah sakit selalu bertambah dengan jenis penyakit yang berbeda. Tumpukan data penyakit pasien yang berada pada rumah sakit pun hanya sebatas memberikan laporan jumlah pasien yang berobat dengan penyakit yang diderita. Oleh sebab itu, perlu adanya pengelompokan untuk membantu pihak rumah sakit menemukan informasi mengenai penyakit yang paling banyak diderita oleh pasien. Pada penelitian ini dilakukan pengelompokkan data penyakit pasien menggunakan teknik data mining clustering dengan menggunakan algoritma K-Medoids. K-Medoids merupakan salah satu metode clustering yang berfungsi untuk memecah dataset menjadi kelompok-kelompok dan mampu mengatasi kelemahan dari metode K-Means yang sensitive terhadap outlier. Hasil dari penelitian ini menunjukkan bahwa penggunaan 3 cluster menjadi jumlah cluster terbaik dengan nilai Silhouette Coefficient sebesar 0,409373. Dari pengujian yang dilakukan didapatkan hasil clustering yaitu cluster 0 sebanyak 18 data yang didominasi oleh penyakit yang berada pada klinik gawat darurat, lalu pada cluster 1 sebanyak 2 data yang didominasi oleh penyakit pada klinik kemuning dan klinik gawat darurat, sedangkan pada cluster 2 sebanyak 197 data yang didominasi oleh penyakit pada klinik rehab medik sebanyak 20 data.

Kata Kunci: Clustering, K-Medoids, Silhouette Coefficient, Penyakit Pasien

Abstract

Disease is an abnormal condition which is the body or mind experiences discomfort or dysfunction of the person affected. Every day the number of patients in the hospital always increases with different types of diseases. Piles of patient disease data at the hospital were only providing reports of disease suffered by patients. Therefore, grouping is required to help the hospital find information about the most common disease suffered by patients. In this study grouping patient disease data using data mining clustering techniques with the K-Medoids algorithm. K-Medoids is a clustering method that functions to break the dataset into groups and able to overcome the weaknesses of the K-Means method which is sensitive to outliers. The results of this study indicate that the use of 3 clusters is the best number of clusters with a Silhouette Coefficient of 0.409373. From the tests conducted, the results of clustering are cluster 0, 18 data are dominated by diseases in emergency clinics, in cluster 1 there are 2 data that are dominated by diseases in kemuning clinics and emergency clinics, whereas in cluster 2 there are 197 data which is dominated by diseases in medical rehab clinic as much as 20 data.

Keywords**:** *Clustering*, *K-Medoids*, *Silhouette Coefficient, Patient Disease*

Naskah diterima 20 April 2020; direvisi 04 Mei 2020; diterbitkan 31 Agustus 2020

1. Pendahuluan

Penyakit adalah suatu keadaan abnormal dimana tubuh ataupun pikiran mengalami ketidaknyamanan atau disfungsi terhadap orang yang dipengaruhinya (Juninda et al., 2019). Lingkungan dan pola hidup sehat dapat menjadi pengaruh besar dalam penyakit yang diderita pasien. Setiap harinya jumlah pasien di rumah sakit selalu bertambah dengan jenis penyakit yang berbeda. Pihak rumah sakit pun selalu dituntut untuk memperbarui data pasien yang dikumpulkan dari data penyakit pasien yang telah selesai melakukan perawatan maupun masih dalam tahap perawatan. Hal tersebut menyebabkan bertumpuknya data penyakit pasien di rumah sakit.

Dari data penyakit pasien tersebut, pihak rumah sakit membutuhkan pengolahan data penyakit lebih lanjut untuk menemukan informasi mengenai penyakit yang paling banyak diderita oleh pasien. Dengan mengetahui kondisi ini, pihak rumah sakit dapat mengambil tindakan kebijakan dalam antisipasi pengobatan dan pencegahan penyakit salah satunya yaitu dengan melakukan penyuluhan atau sosialisasi sehingga kegiatan tersebut akan tepat sasaran terhadap penyakit yang sering diderita pada masyarakat (Sundari & Ariani, 2019). Selain itu, pengolahan data penyakit ini dapat dijadikan sebagai bahan evaluasi pihak rumah sakit dalam merencanakan pembangunan fasilitas baru, penambahan tenaga medis spesialis ataupun penambahan jenis obat. Oleh karena itu, dibutuhkan suatu teknik ataupun metode untuk mengolah data penyakit pasien tersebut.

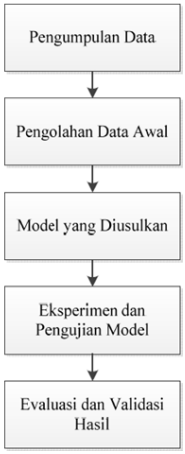
Data mining adalah metode dimana data yang telah didapatkan akan diolah untuk mengetahui pola dari data sehingga dapat di ambil informasi yang tersembunyi dari data tersebut (Triyanto, 2015). Salah satu teknik analisa *Data Mining* adalah *clustering* yang mana konsep dasar dari *clustering* adalah mengelompokkan sejumlah objek ke dalam *cluster* dimana *cluster* yang baik adalah *cluster* yang memiliki tingkat kesamaan yang tinggi antar objek di dalam suatu *cluster* dan tingkat ketidaksamaan yang tinggi dengan objek *cluster* yang lainnya (Silitonga & Sri, 2017).

Algoritma yang digunakan pada teknik *clustering* ini adalah Algoritma *K-Medoids* yang didasarkan pada penelitian terdahulu yang dilakukan oleh (Juninda et al., 2019) mengenai Penerapan Algoritma *K-Medoids* untuk Pengelompokan Penyakit di Pekanbaru Riau. Berdasarkan penelitiannya menunjukan bahwa algoritma *K-Medoids* dapat melakukan pengelompokan penyakit yang ada di Pekanbaru Riau dengan baik. Algoritma *K-Medoids* juga merupakan metode yang diciptakan untuk mengatasi kelemahan algoritma *K-Means* yang sensitif terhadap *outlier,* karena nilai yang sangat besar dapat secara substansial mendistorsi distribusi data (Asmiatun et al., 2020).

2. Metode Penelitian

2.1. Perancangan Penelitian

Menurut Dawson dalam (Amrin et al., 2019) terdapat empat metode penelitian yang umum digunakan diantaranya penelitian langsung, eksperimen, studi kasus dan *survey*. Pada tahapan penelitian ini, penelitian yang digunakan adalah penelitian eksperimen. Adapun tahapannya dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.2. Pengumpulan Data

Pada penelitian ini dataset yang digunakan adalah dataset publik yang mana data tersebut mudah diperoleh dan tersedia untuk umum yang dapat diteliti oleh siapapun. Dataset yang digunakan pada penelitian ini adalah data penyakit pasien di RSUD Kota bandung pada tahun 2019 yang diambil dari situs http://data.bandung.go.id.

2.3. Pengolahan Data Awal

Pada tahap ini dilakukan tahap pembersihan data *(cleaning data)* yang mana menghilangkan nilai-nilai yang salah atau kosong, memperbaiki dan memeriksa data yang tidak konsisten. Selanjutnya dilakukan transfromasi data pada *Microsoft Excel* yang

nantinya data tersebut dapat langsung diolah dengan model yang telah ditentukan.

A. *Data Cleaning*

Sebelum melakukan proses *data mining* perlu melakukan proses *cleaning data* terlebih dahulu. Proses *cleaning* meliputi membuang duplikasi data, memeriksa data yang inkonsisten dan memperbaiki kesalahan pada data seperti kesalahan penulisan. Atribut data yang tidak digunakan adalah tahun dan bulan yang nantinya akan dihilangkan sehingga atribut yang digunakan berjumlah 5 atribut yaitu Kode ICD, Nama Penyakit, Kamar, Laki-Laki, dan Perempuan.

B. Data Transformasi

Tahap transformasi adalah tahapan mengubah format data asli menjadi bentuk yang sesuai untuk mempermudah proses penambangan data. Data yang berjenis nominal seperti kamar harus dilakukan proses inisialisasi data terlebih dahulu ke dalam bentuk angka atau numerikal. Untuk melakukan inisialisasi dapat dilakukan dengan cara pengurutan angka berdasarkan frekuensinya. Pengubahan ini dilakukan pada *Micorosoft Excel*. Tabel atribut yang telah diinisialisasi dapat dilihat pada Tabel 1.

Tabel 1. Data Transformasi

Kamar	Frekuensi	Inisialisasi
Klinik Rehab Medik	22	1
Klinik Kandungan	21	2
Klinik Dalam	20	3
Gawat Darurat	19	4
Klinik Anak	16	5
Klinik Syaraf	15	6
Klinik Bedah	12	7
Klinik Gigi	12	8
Klinik Kulit Kelamin	12	9
Klinik Mata	12	10
Klinik Orthopedic	11	11
Klinik THT	11	12
Klinik Bedah Anak	6	13
Klinik Psikiatri	5	14
Klinik Bedah Mulut	4	15
Klinik Kemuning	4	16
Klinik Urologi	4	17
Klinik Dahlia	3	18
Klinik Akupuntur	2	19
Klinik Bedah Syaraf	2	20
Klinik Konservasi Gigi	2	21
Klinik vct	2	22

C. Populasi dan *Sample* Penelitian

Populasi *dataset* pada penelitian ini sebanyak 473 data dengan jumlah 7 atribut. Sebelum dilakukan pengambilan *sample,* terlebih dahulu dilakukan proses pembersihan data. Besarnya *sample* pada penelitian ini berdasarkan pada rumus slovin yaitu:

$$n = \frac{N}{N(d)^2 + 1}$$

dimana :
n = *Sample*
N = Populasi
d = Nilai presisi 95% atau sig. = 0,05 (tingkat kesalahan yang dikehendaki adalah 5%)
Berdasarkan rumus diatas, maka perhitungan *sample* sebagai berikut:

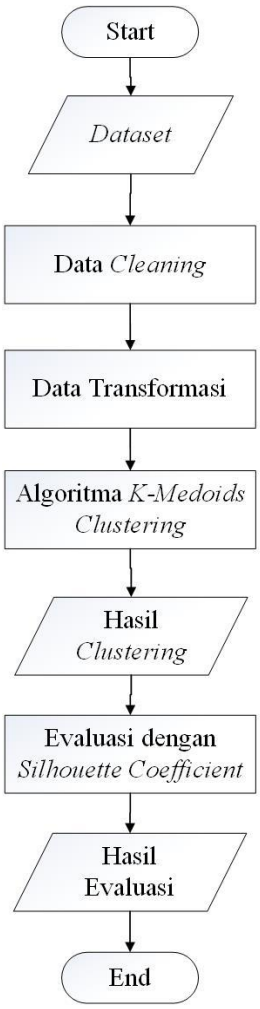
$$n = \frac{473}{473(0.05)^2 + 1}$$

$$n = \frac{473}{2.1825} \rightarrow n = 216{,}72394$$
$$= 217 \ (dibulatkan)$$

Teknik pengambilan *sample* penelitian ini menggunakan *Simple Random Sampling* dimana penarikan *sample* menggunakan prosedur yang memungkinkan setiap elemen dalam populasi akan memiliki peluang sama untuk dijadikan *sample*. *Sample* diambil secara acak tanpa memperhatikan tingkatan yang ada dalam populasi (Muningsih & Kiswati, 2015).

2.4. Model yang diusulkan

Pada tahap ini ditentukan model yang akan digunakan dalam penelitian. Model atau metode yang diusulkan dalam penelitian ini adalah algoritma *K- Medoids*. Model yang diusulkan ditunjukkan dengan *flowchart* pada Gambar 2.



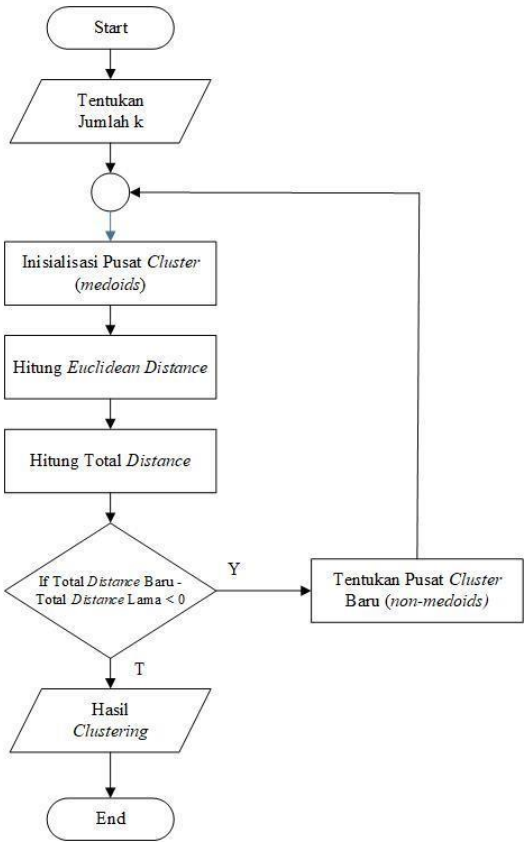
Gambar 2. Model Yang Diusulkan

2.5. Eksperimen dan pengujian model

Pada tahap ini dilakukan pengujian terhadap model yang diusulkan untuk mendapatkan hasil kinerja dari model yang diusulkan. Dataset yang telah siap diolah akan diuji dengan menggunakan algoritma *K-medoids* seperti pada Gambar 3.

2.6. Evaluasi dan validasi model

Pada tahap ini dilakukan evaluasi terhadap eksperimen dan pengujian pada model yang diusulkan sehingga dapat mengetahui hasil kinerja dalam penelitian ini. Evaluasi hasil *clustering* dilakukan dengan menggunakan *Silhouette Coefficient*.



Gambar 3. *Flowchart* Algoritma *K-Medoids*

3. Hasil dan Pembahasan

3.1. Perhitungan Manual Algoritma *K-Medoids*

Dalam penelitian ini dilakukan perhitungan manual algoritma *K-Medoids* dengan menggunakan *Microsoft Excel*. Adapun langkah-langkah perhitungan algoritma *K-Medoids* adalah sebagai berikut:

1. Inisialisasi pusat *cluster* sebanyak k (jumlah *cluster*)

Dalam penelitian ini, dibentuk 3 *cluster* berdasarkan kamar, jumlah pasien laki-laki dan jumlah pasien perempuan. Selanjutnya adalah menentukan titik pusat dari setiap *cluster* (*medoids*) yang dipilih secara *random* dari data yang telah di transformasi. Berikut merupakan *medoids* awal yang digunakan:

Tabel 2. *Medoids* Awal

Cluster	Kamar	Laki-laki	Perempuan
0	16	3449	661
1	4	1022	1263
2	3	756	1241

2. Hitung setiap objek ke *cluster* terdekat menggunakan persamaan ukuran jarak *Euclidean Distance*

Setelah menentukan *medoids* awal, hitung jarak setiap data ke masing-masing *medoids* menggunakan rumus *Euclidean Distance* sehingga diperoleh hasil penghitungan jarak setiap data untuk masing-masing *cluster*. Berikut merupakan contoh perhitungan pada data pertama:

Data pertama ke pusat *cluster* 0:

$$d(x_1c_0)$$
$$= \sqrt{(5-16)^2 + (89-3449)^2 + (73-661)^2}$$
$$= \sqrt{(-11)^2 + (-3360)^2 + (-588)^2}$$
$$= \sqrt{(121) + (11289600) + (345744)}$$
$$= \sqrt{11635465}$$
$$= 3411{,}08$$

Data pertama ke pusat *cluster* 1:

$$d(x_1c_1)$$
$$= \sqrt{(5-4)^2 + (89-1022)^2 + (73-1263)^2}$$
$$= \sqrt{(1)^2 + (-933)^2 + (-1190)^2}$$
$$= \sqrt{(1) + (870489) + (1416100)}$$
$$= \sqrt{2286590}$$
$$= 1512{,}15$$

Data pertama ke pusat *cluster* 2:

$$d(x_1c_2)$$
$$= \sqrt{(5-3)^2 + (89-756)^2 + (73-1241)^2}$$
$$= \sqrt{(2)^2 + (-667)^2 + (-1168)^2}$$
$$= \sqrt{(4) + (444889) + (1364224)}$$
$$= \sqrt{1809117}$$
$$= 1345{,}0342$$

Perhitungan tersebut dilakukan pada semua data yang mana terdapat 217 *dataset* penyakit pasien. Setelah dilakukan perhitungan pada semua data, maka proses selanjutnya adalah memilah *cluster* mana dari tiap data dimana nilai terkecil dari tiap perhitungan *cluster* akan menjadi penentu *cluster* yang akan diikuti dari data tersebut. Hasil perhitungan jarak pada iterasi-1 dapat dilihat pada Tabel 4.

3. Inisialisasikan pusat *cluster* baru secara acak pada masing – masing objek sebagai kandidat *non medoids*

Setelah didapatkan hasil jarak dari setiap objek pada iterasi ke-1, maka dilanjutkan ke perhitungan iterasi ke-2. Pada tahap ini, tentukan kembali *medoids* baru (*non medoids*) yang dapat dilihat pada tabel berikut:

Tabel 3. *Medoids* Kedua

Cluster	Kamar	Laki-laki	Perempuan
0	2	0	861
1	4	3619	3616
2	10	149	179

Tabel 4. Hasil Perhitungan Iterasi-1

Kode ICD	Nama Penyakit	C0	C1	C2	Jarak Minimum	*Cluster* Yang Diikuti
A01.0	Typhoid fever	3411,08	1512,147	1345,0342	1345,0342	2
A02.0	Salmonella enteritis	3500,929	1607,28	1435,3738	1435,3738	2
A09	Diarrhoea and gastroenteritis of	2500,575	0	266,9100	0	1
A09	Diarrhoea and gastroenteritis of	3189,438	1298,578	1150,4090	1150,4090	2
A15.9	Respiratory tuberculosis unspecified,	3347,726	1412,904	1244,6887	1244,6887	2
A16.2	Tuberculosis of lung, without mention of	3264,88	1339,201	1178,9185	1178,9185	2
A16.9	Respiratory tuberculosis unspecified, without	3371,243	1477,837	1314,2910	1314,2910	2
.....	.....	.....	.....	.....	.....	.....
Z96.1	Presence of intraocular lens	3335,02	1391,84	1223,25	1223,25	2

4. Hitung jarak setiap objek yang berada pada masing – masing *cluster* dengan kandidat *non medoids*

Hitung jarak setiap data ke masing-masing *non medoids* menggunakan rumus *Euclidean Distance* sehingga diperoleh hasil penghitungan jarak setiap data untuk masing-masing *cluster*. Berikut merupakan contoh perhitungan pada data pertama:

$$d(x_1c_0)$$
$$= \sqrt{(5-2)^2 + (89-0)^2 + (73-861)^2}$$
$$= \sqrt{(3)^2 + (89)^2 + (-788)^2}$$
$$= \sqrt{(9) + (7921) + (620944)}$$
$$= \sqrt{628874}$$
$$= 793{,}016$$

Data pertama ke pusat *cluster* 1:

$$d(x_1c_1)$$
$$= \sqrt{(5-4)^2 + (89-3619)^2 + (73-3616)^2}$$
$$= \sqrt{(1)^2 + (-3530)^2 + (-3543)^2}$$
$$= \sqrt{(1) + (12460900) + (12552849)}$$
$$= \sqrt{25013750}$$
$$= 5001{,}37$$

Data pertama ke pusat *cluster* 2:

$$d(x_1c_2)$$
$$= \sqrt{(5-10)^2 + (89-149)^2 + (73-179)^2}$$
$$= \sqrt{(-5)^2 + (-60)^2 + (-106)^2}$$
$$= \sqrt{(25) + (3600) + (11236)}$$
$$= \sqrt{14861}$$
$$= 121{,}906$$

Perhitungan tersebut dilakukan pada semua data yang mana terdapat 217 *dataset*

penyakit pasien. Setelah dilakukan perhitungan pada semua data, maka proses selanjutnya adalah memilah *cluster* mana dari tiap data dimana nilai terkecil dari tiap perhitungan *cluster* akan menjadi penentu *cluster* yang akan diikuti dari data tersebut. Hasil perhitungan jarak pada iterasi-1 dapat dilihat pada Tabel 5.

Tabel 5. Hasil Perhitungan Iterasi-2

Kode ICD	Nama Penyakit	C0	C1	C2	Jarak Minimum	*Cluster* Yang Diikuti
A01.0	Typhoid fever	793,016	5001,37	121,906	121,906	2
A02.0	Salmonella enteritis	845,1538	5098,967	215,6061	215,6061	2
A09	Diarrhoea and gastroenteritis of	1098,222	3504,428	1391,841	1098,222	0
A09	Diarrhoea and gastroenteritis of	735,7377	4775,669	146,3045	146,3045	2
A15.9	Respiratory tuberculosis unspecified,	714,6489	4904,522	22,58318	22,58318	2
A16.2	Tuberculosis of lung, without mention of	703,6292	4826,031	70,39176	70,39176	2
A16.9	Respiratory tuberculosis unspecified, without	784,3558	4964,666	94,72592	94,72592	2
.....	.....	.....	.....	.....	.....	.....
Z96.1	Presence of intraocular lens	698,133	4884,05	0	0	2

5. Hitung total simpangan (S) dengan menghitung total *cost* baru – total *cost* lama

Setelah didapatkan hasil nilai jarak pada iterasi ke-1 dan iterasi ke-2, hitung total simpangan (S) dengan mencari selisih dari total *cost* baru dan total *cost* lama.

S = total *cost* baru – total *cost* lama
= 1286918 – 1315466
= -28547,9

Dikarenakan total simpangan < 0, maka perhitungan dilanjutkan sampai dengan total simpangan > 0. Tahap selanjutnya tentukan kembali *medoids* baru sebagai titik pusat *cluster* pada iterasi ke-3, berikut merupakan *medoids* yang digunakan:

Tabel 6. *Medoids* Ketiga

Cluster	Kamar	Laki-laki	Perempuan
0	6	1578	1651
1	4	1203	2173
2	18	987	1549

Hitung kembali jarak setiap data ke masing-masing *non medoids* menggunakan rumus *Euclidean Distance* sehingga diperoleh hasil penghitungan jarak setiap data untuk masing-masing *cluster*. Berikut merupakan contoh perhitungan pada data pertama:

Data pertama ke pusat *cluster* 0:

$$d(x_1c_0)$$
$$= \sqrt{(5-6)^2 + (89-1578)^2 + (73-1651)^2}$$
$$= \sqrt{(-1)^2 + (1489)^2 + (-1578)^2}$$
$$= \sqrt{(1) + (2217121) + (2490084)}$$
$$= \sqrt{4707206}$$
$$= 2169{,}61$$

Data pertama ke pusat *cluster* 1:

$$d(x_1c_1)$$
$$= \sqrt{(5-4)^2 + (89-1203)^2 + (73-2173)^2}$$
$$= \sqrt{(1)^2 + (-1114)^2 + (-2100)^2}$$
$$= \sqrt{(1) + (1240996) + (4410000)}$$
$$= \sqrt{5650997}$$
$$= 2377{,}18$$

Data pertama ke pusat *cluster* 2:

$$d(x_1c_2)$$
$$= \sqrt{(5-18)^2 + (89-987)^2 + (73-1549)^2}$$
$$= \sqrt{(-13)^2 + (-898)^2 + (-1476)^2}$$
$$= \sqrt{(169) + (806404) + (2178576)}$$
$$= \sqrt{2985149}$$
$$= 1727{,}76$$

Perhitungan tersebut dilakukan pada semua data yang mana terdapat 217 *dataset* penyakit pasien. Setelah dilakukan perhitungan pada semua data, maka proses selanjutnya adalah memilah *cluster* mana dari tiap data dimana nilai terkecil dari tiap perhitungan *cluster* akan menjadi penentu *cluster* yang akan diikuti dari data tersebut. Hasil perhitungan jarak pada iterasi-1 dapat dilihat pada Tabel 7.

Tabel 7. Hasil Perhitungan Iterasi-3

Kode ICD	Nama Penyakit	C0	C1	C2	Jarak Minimum	*Cluster* Yang Diikuti
A01.0	Typhoid fever	2169,61	2377,183	1727,758	1727,758374	2
A02.0	Salmonella enteritis	2266,765	2465,931	1818,938	1818,937602	2
A09	Diarrhoea and gastroenteritis of	678	927,826	288,4736	288,473569	2
A09	Diarrhoea and gastroenteritis of	1946,635	2183,726	1527,769	1527,768634	2
A15.9	Respiratory tuberculosis unspecified,	2072,151	2276,898	1627,467	1627,467358	2
A16.2	Tuberculosis of lung, without mention of	1994,563	2212,371	1559,801	1559,800949	2
A16.9	Respiratory tuberculosis unspecified, without	2133,424	2347,163	1696,235	1696,234948	2
A17.0	Tuberculous meningitis	2098,774	2313,963	1662,538	1662,538421	2
.....	.....	.....	.....	.....	.....	.....
Z96.1	Presence of intraocular lens	2051,546	2255,435	1605,991	1605,991283	2

Setelah dilakukan perhitungan jarak pada iterasi ke-3, hitung kembali total simpangan baru dengan mencari selisih total *cost* baru dengan total *cost* lama.

S = total *cost* baru – total *cost* lama

= 12195953 - 1286918
= 9034,99

Dikarenakan total simpangan > 0, maka perhitungan dihentikan sehingga didapatkan hasil *cluster* akhir sebagai berikut:

Tabel 8. Hasil *Cluster* Akhir

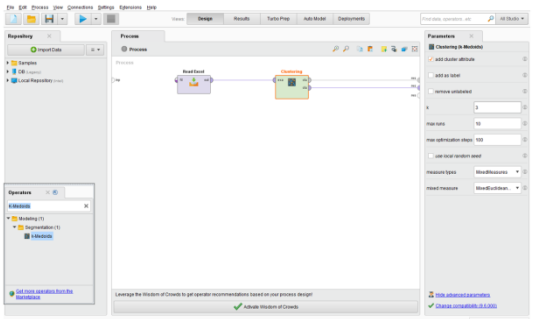
Kode ICD	Nama Penyakit	Kamar	Laki-laki	Perempuan	*Cluster*
A01.0	Typhoid fever	5	89	73	2
A02.0	Salmonella enteritis	16	8	16	2
A09	Diarrhoea and gastroenteritis of	4	1022	1263	0
A09	Diarrhoea and gastroenteritis of	5	295	187	2
A15.9	Respiratory tuberculosis unspecified,	3	139	160	2
A16.2	Tuberculosis of lung, without mention of	5	218	192	2
.....	.....	.....	.....	.....	.....
Z96.1	Presence of intraocular lens	10	149	179	2

Berdasarkan hasil perhitungan manual dengan algoritma *K-Medoids* yang telah dilakukan, didapatkan hasil *cluster* akhir dengan nilai k=3 yaitu *cluster* 0 sebanyak 18 *items*, *cluster* 1 sebanyak 2 *items*, dan *cluster* 2 sebanyak 197 items sehingga total keseluruhan adalah 217 *items*.

3.2. Perhitungan Algoritma *K-Medoids* dengan *Rapidminer*

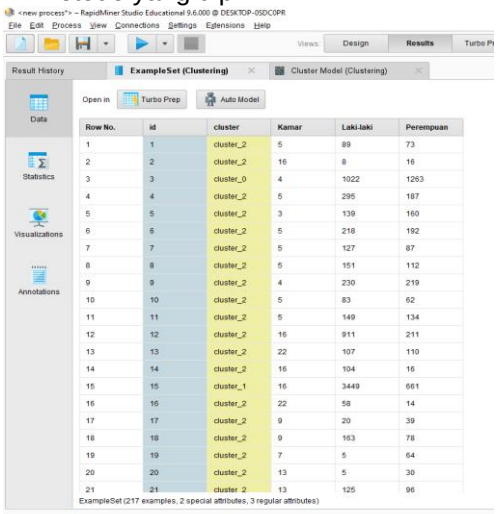
Pada tahap ini pemodelan dilakukan dengan menggunakan *Rapidminer* versi 9.6 dengan langkah sebagai berikut:

1. Pilih *read excel* lalu drag ke panel *process* kemudian *Import* Data *Excel* pada panel *parameters* dengan klik *excel file* kemudian pilih file *excel* yang akan diuji.
2. Pada panel *operators* ketik *K-Medoids* lalu *drag* ke panel *process* dan tentukan jumlah k yang akan digunakan.
3. Pada panel *process* hubungkan *connector* pada masing-masing *process* lalu klik tombol *Run* untuk melihat hasil *clustering.*

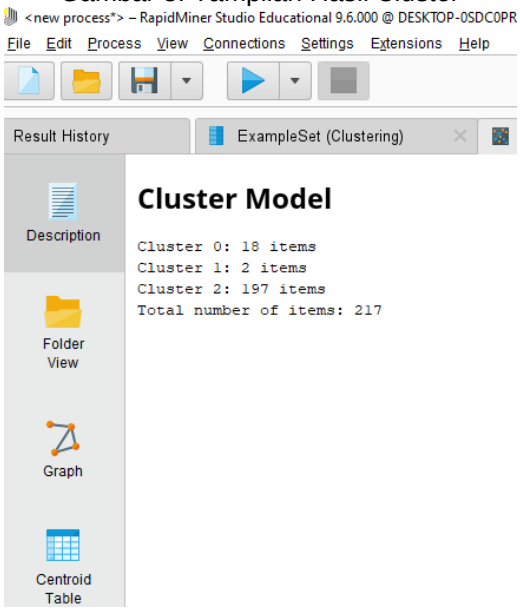


Gambar 4. Tampilan Proses *Clustering*

4. Setelah proses *running* selesai maka akan tampil hasil *clustering* dan *cluster model* yang dihasilkan dari *dataset* dan metode yang dipilih.



Gambar 5. Tampilan Hasil *Cluster*



Gambar 6. Tampilan *Cluster Model*

Berdasarkan hasil pengujian *dataset* sebanyak 217 data dengan menggunakan *software rapidminer* 9.6, pada *cluster model* yang dibentuk dengan 3 *cluster* diperoleh hasil *cluster* 0 sebanyak 18 *items*, *cluster* 1 sebanyak 2 *items* dan *cluster* 2 sebanyak 197 *items* dengan total keseluruhan 217 *items*.

3.3. Evaluasi Hasil *Cluster*

Dari hasil pengujian *dataset* yang telah dilakukan, maka tahap selanjutnya adalah mengevaluasi hasil *cluster* untuk menentukan kualitas hasil *cluster* tersebut. Pada tahap ini evaluasi hasil *clustering* yang digunakan adalah Metode *Silhouette Coefficient*. Adapun langkah-langkah perhitungan *Silhouette Coefficient* adalah sebagai berikut:

1. **Hitung jarak rata-rata dari suatu objek misalkan objek ke-i dengan semua objek lain yang berada di dalam satu *cluster***

Pada tahap ini hitung nilai *average dissimilarity* atau *a(i)*. Berikut merupakan contoh perhitungan jarak Data ke-1 terhadap semua data yang berada pada *cluster* yang sama yaitu *cluster* 2:

$$a(i) = \sqrt{\begin{array}{c}(121{,}9057 - 121{,}9057)^2 + (215{,}606 - 121{,}9057)^2 + (146{,}304 - 121{,}9057)^2 + \\ (22{,}58318 - 121{,}9057)^2 + (70{,}3917 - 121{,}9057)^2 + (94{,}72592 - 121{,}9057)^2 + \\ (67{,}21607 - 121{,}9057)^2 + (90{,}53729 - 121{,}9057)^2 + \cdots + (0 - 121{,}9057)^2\end{array}}$$

$$a(i) = \frac{306449{,}4}{197}$$

$$a(i) = 1555{,}581$$

2. **Hitung rata-rata jarak dari objek ke-i tersebut dengan semua objek pada *cluster* lainnya, kemudian ambillah nilai terkecilnya.**

Pada tahap ini hitung *lowest average dissimilarity* atau *b(i)*. Berikut merupakan contoh perhitungan jarak data pertama terhadap semua data yang berada pada *cluster* 0 dan 1 kemudian cari nilai rata-rata jarak yang paling kecil.

- Jarak data pertama pada *cluster* 2 ke *cluster* 0:

$$b(i)_{20} = \sqrt{\begin{array}{c}(121{,}9057 - 1098{,}22)^2 + (215{,}606 - 1098{,}22)^2 + (146{,}304 - 1098{,}22)^2 + \\ (22{,}58318 - 1098{,}22)^2 + (70{,}3917 - 1098{,}22)^2 + (94{,}72592 - 1098{,}22)^2 + \\ (67{,}21607 - 1098{,}22)^2 + (90{,}53729 - 1098{,}22)^2 + \cdots + (0 - 1098{,}22)^2\end{array}}$$

$$b(i)_{20} = 12964{,}71359$$

$$b(i)_{20} = \sqrt{\begin{array}{c}(121{,}9057 - 448{,}1361)^2 + (215{,}606 - 448{,}1361)^2 + (146{,}304 - 448{,}1361)^2 + \\ (22{,}58318 - 448{,}1361)^2 + (70{,}3917 - 448{,}1361)^2 + (94{,}72592 - 448{,}1361)^2 + \\ (67{,}21607 - 448{,}1361)^2 + (90{,}53729 - 448{,}1361)^2 + \cdots + (0 - 448{,}1361)^2\end{array}}$$

$$b(i)_{20} = 4017{,}0703$$

Hitung perhitungan diatas pada semua data yang berada pada *cluster* 0 sehingga diperoleh:

$$b(i)_{20} = \frac{12964{,}71359 + 4017{,}0703 + 9465{,}56 + 5095{,}27 + 2247{,}98 + \cdots + 2713{,}34}{18}$$

$$b(i)_{20} = \frac{132448{,}3591}{18}$$

$$b(i)_{20} = 7358{,}242174$$

- Jarak data pertama pada *cluster* 2 ke *cluster* 1:

$$b(i)_{21} = \sqrt{\begin{array}{c}(121{,}9057 - 2959{,}91)^2 + (215{,}606 - 2959{,}91)^2 + (146{,}304 - 2959{,}91)^2 + \\ (22{,}58318 - 2959{,}91)^2 + (70{,}3917 - 2959{,}91)^2 + (94{,}72592 - 2959{,}91)^2 + \\ (67{,}21607 - 2959{,}91)^2 + (90{,}53729 - 2959{,}91)^2 + \cdots + (0 - 2959{,}91)^2\end{array}}$$

$$b(i)_{21} = 38923{,}90196$$

$$b(i)_{21} = \sqrt{\begin{array}{c}(121{,}9057 - 0)^2 + (215{,}606 - 0)^2 + (146{,}304 - 0)^2 + (22{,}58318 - 0)^2 + \\ (70{,}3917 - 1098{,}22)^2 + (94{,}72592 - 0)^2 + (67{,}21607 - 0)^2 + (90{,}53729 - 0)^2 \\ + (134{,}4247 - 0)^2 + (45{,}27693 - 0)^2 + (762{,}6952 - 0)^2 + \cdots + (0 - 0)^2\end{array}}$$

$$b(i)_{21} = 2713{,}343918$$

Hitung perhitungan diatas pada semua data yang berada pada *cluster* 1 sehingga diperoleh:

$$b(i)_{21} = \frac{38923{,}90196 + 2713{,}343918}{2}$$

$$b(i)_{21} = \frac{41637{,}24587}{2}$$

$$b(i)_{21} = 20818{,}62294$$

Dari perhitungan kedua *cluster* diatas, ambil nilai minimum diantara kedua *cluster* tersebut.

b(i) = (7358,242174 < 20818,62294)

b(i) = 7358,242174

3. **Hitung nilai *silhouette coefficient***

Setelah mengetahui nilai *a(i)* dan *b(i)*, hitung nilai *silhouette coefficient.* Berikut merupakan hasil perhitungan nilai *silhouette coefficient* pada *cluster* 2:

$$s(i) = 1 - \frac{a(i)}{b(i)}$$

$$s(i) = 1 - \frac{1555{,}581}{7358{,}242174}$$

$$s(i) = 0{,}78859$$

Lakukan perhitungan diatas pada semua *cluster*. Berikut adalah hasil perhitungan nilai *silhouette coefficient* pada semua *cluster*:

Tabel 9. Nilai *Silhouette Coefficient*

Cluster	**Nilai *Silhouette Coefficient***
0	0,034497
1	0,405027
2	0,788593

Setelah didapatkan nilai *silhouette coefficient*, hitung nilai *silhouette coefficient* global dengan menghitung rata-rata nilai

silhouette coefficient pada semua *cluster* sebagai berikut:

$$s(i) = \frac{0{,}034497 + 0{,}405027 + 0{,}788593}{3}$$

$$s(i) = \frac{1{,}228118}{3}$$

$$s(i) = 0{,}409373$$

Untuk mengetahui apakah hasil validasi *cluster* yang dibentuk tergolong dalam kategori baik, maka dilakukan percobaan validasi dari pembentukan jumlah *k* sebanyak 2 sampai dengan 5 *k* (*cluster)*. Berikut adalah tabel hasil percobaan validasi *cluster* dengan menggunakan *silhouette coefficient*:

Tabel 10. Hasil Pengujian *Silhouette Coefficient*

Data Uji	**Nilai k (Jumlah *Cluster*)**	**Rata-Rata *Silhouette Coefficient***
217	2	0,346969
	3	0,409373
	4	0,138031
	5	0,025724

Berdasarkan hasil evaluasi *cluster* yang telah dilakukan, kualitas *cluster* dengan jumlah *cluster* sama dengan 3 memiliki kualitas yang paling baik karena nilai *silhouette coefficient* dengan *k* = 3 merupakan nilai yang paling mendekati 1 yaitu sebesar 0,409373 sehingga pembentukan 3 *cluster* telah menghasilkan pengelompokan yang baik.

3.4. Rekapitulasi Hasil *Cluster*

Berdasarkan hasil dari proses *clustering* dengan algoritma *K-Medoids* menggunakan aplikasi *rapidminer* maupun perhitungan secara manual, maka dihasilkan informasi sebagai berikut:

Tabel 11. Rekapitulasi *Cluster* 0, 1, 2

	Atribut/Variabel		**C0**	**C1**	**C2**
Kamar	1.	Klinik Rehab Medik	2	-	20
	2.	Klinik Kandungan	2	-	19
	3.	Klinik Dalam	3	-	17
	4.	Gawat Darurat	5	1	13
	5.	Klinik Anak	-	-	16
	6.	Klinik Syaraf	3	-	12
	7.	Klinik Bedah	-	-	12
	8.	Klinik Gigi	2	-	10
	9.	Klinik Kulit Kelamin	-	-	12
	10.	Klinik Mata	-	-	12
	11.	Klinik Orthopedic	-	-	11
	12.	Klinik THT	-	-	11
	13.	Klinik Bedah Anak	-	-	6
	14.	Klinik Psikiatri	-	-	5
	15.	Klinik Bedah Mulut	-	-	4
	16.	Klinik Kemuning	-	1	3
	17.	Klinik Urologi	-	-	4
	18.	Klinik Dahlia	1	-	2
	19.	Klinik Akupuntur	-	-	2
	20.	Klinik Bedah Syaraf	-	-	2
	21.	Klinik Konservasi Gigi	-	-	2
	22.	Kllinik vct	-	-	2
Laki-laki	1.	0 – 500	11	-	195
	2.	501 – 1000	4	-	2
	3.	1001 – 1500	2	-	-
	4.	1501 – 2000	1	-	-
	5.	2001 – 2500	-	-	-
	6.	> 2500	-	2	-
Perempuan	1.	0 – 500	-	-	195
	2.	501 – 1000	11	1	2
	3.	1001 – 1500	4	-	-
	4.	1501 – 2000	2	-	-
	5.	2001 – 2500	1	-	-
	6.	> 2500	-	1	-

Berdasarkan tabel diatas maka dapat disimpulkan bahwa klinik gawat darurat mendominasi pada *cluster* 0 sebanyak 5 data, lalu pada *cluster* 1 hanya terdapat 2 data yang diisi oleh klinik kemuning dan klinik gawat darurat, sedangkan pada *cluster* 2 didominasi oleh klinik rehab medik sebanyak 20 data. Untuk lebih lengkapnya, informasi yang didapatkan dapat dilihat pada Tabel 12. di bawah ini:

Tabel 12. Kesimpulan

Cluster	Kesimpulan
0	Jumlah Anggota : 18 ✓ Kamar yang sering didatangi pasien adalah klinik syaraf, klinik gawat darurat dan klinik dahlia. ✓ Penyakit yang sering terjadi adalah *Sequelae of stroke*, *Dyspepsia* dan *End-stage renal disease*. ✓ Pasien didominasi oleh perempuan. ✓ Jumlah pasien laki-laki sebanyak 9503. ✓ Jumlah pasien perempuan sebanyak 19432.
1	Jumlah Anggota : 2 ✓ Kamar yang sering didatangi pasien adalah klinik gawat darurat dan klinik kemuning. ✓ Penyakit yang sering terjadi adalah *Fever* dan *HIV disease resulting in unspecified infectious or parasitic disease*. ✓ Pasien didominasi oleh laki-laki. ✓ Jumlah pasien laki-laki sebanyak 7068. ✓ Jumlah pasien perempuan sebanyak 4277.
2	Jumlah Anggota : 197 ✓ Kamar yang sering didatangi pasien adalah klinik rehab medik, klinik kandungan, dan klinik dalam. ✓ Penyakit yang sering terjadi adalah *Fracture of bone following insertion of orthopaedic implant, Chronic obstructive pulmonary disease,* dan *Open wound of unspecified body region*. ✓ Pasien didominasi oleh perempuan. ✓ Jumlah pasien laki-laki sebanyak 17738. ✓ Jumlah pasien perempuan sebanyak 21812.

Hasil Penelitian pada klasterisasi data penyakit pasien di RSUD Kota Bandung dengan menggunakan Algoritma *K-Medoids* menunjukan bahwa *cluster* 2 menjadi *cluster* paling tinggi dengan jumlah penyakit 197 data, lalu disusul oleh *cluster* 0 dengan jumlah penyakit 18 data dan *cluster* 1 menjadi *cluster* terendah dengan jumlah penyakit 2 data. Angka tersebut menunjukan hasil yang sama baik dari perhitungan manual di *Microsoft Excel* maupun hasil dari *Rapidminer*.

Menimbang dari hasil penelitian yang telah dilakukan, diharapkan dengan adanya penelitian ini dapat memberikan informasi mengenai penyakit pasien yang sering terjadi berdasarkan kamar dan jumlah pasien, sehingga pihak rumah sakit dapat mengambil tindakan kebijakan dalam antisipasi pengobatan dan pencegahan penyakit. Selain itu, hasil penelitian ini dapat dijadikan sebagai bahan evaluasi pihak rumah sakit dalam merencanakan pembangunan fasilitas baru, penambahan tenaga medis spesialis ataupun penambahan jenis obat.

4. Kesimpulan

Berdasarkan hasil penelitian mengenai pengelompokan data penyakit pasien dengan menggunakan algoritma *K-Medoids* maka diperoleh beberapa kesimpulan sebagai berikut:

1. Pengelompokan penyakit pasien dengan atribut Kode ICD, Nama Penyakit, Kamar, Laki-laki dan Perempuan menggunakan algoritma *K-Medoids* dengan jumlah sampel sebanyak 217 data yang dibentuk menjadi 3 *cluster* menghasilkan hasil *clustering* dimana pada *cluster* 0 terdapat 18 data yang mana kamar yang sering didatangi pasien adalah klinik syaraf, klinik gawat darurat dan klinik dahlia dengan penyakit yang sering terjadi adalah *Sequelae of stroke* dan pasien didominasi oleh perempuan. Kemudian pada *cluster* 1 hanya terdapat 2 data yang mana kamar yang sering diatangi pasien adalah klinik kemuning dan klinik gawat darurat dengan penyakit yang sering terjadi adalah *Fever* dan *HIV disease resulting in unspecified infectious or parasitic disease* dan didominasi oleh pasien laki-

laki. Sedangkan pada *cluster* 2 terdapat 197 data yang mana kamar yang sering didatangi pasien adalah klinik rehab medik, klinik kandungan, dan klinik dalam dengan penyakit yang sering terjadi adalah *Fracture of bone following insertion of orthopaedic implant, Chronic obstructive pulmonary disease,* dan *Open wound of unspecified body region* dan pasien didominasi oleh perempuan.

2. Berdasarkan nilai evaluasi *Silhouette Coefficient* pembentukan 3 *cluster* telah dinilai baik karena dengan percobaan evaluasi dari pembentukan jumlah *k* (*cluster*) 2 – 5 *k* (*cluster*), jumlah *k* (*cluster)* sama dengan 3 merupakan nilai yang paling mendekati 1 yaitu sebesar 0,409373.
3. Algoritma *K-Medoids* telah bekerja dengan baik karena setiap objek pada setiap *cluster* memiliki mutu yang baik, dimana setiap objek telah dikelompokan sesuai dengan tingkat kemiripan yang tinggi.
4. Dari hasil *cluster* didapat informasi tentang penyakit yang sering terjadi berdasarkan kamar dan jumlah pasien sehingga dapat membantu pihak rumah sakit untuk mengambil tindakan kebijakan dalam antisipasi pengobatan dan pencegahan penyakit. Selain itu, hasil penelitian ini dapat dijadikan sebagai bahan evaluasi pihak rumah sakit dalam merencanakan pembangunan fasilitas baru, penambahan tenaga medis spesialis ataupun penambahan jenis obat.

Referensi

Amrin, Satriadi, I., & Rosanto, O. (2019). Algoritma C4.5 Untuk Diagnosa Penyakit Tuberkulosis. *Jurnal Khatulistiwa Informatika*, *VII*(2), 79–84.

Asmiatun, S., Wakhidah, N., & Putri, A. N. (2020). Penerapan Metode K-Medoids Untuk Pengelompokkan Kondisi Jalan Di Kota Semarang 1,2. *Jurnal Teknik Informatika Dan Sistem Informasi*, *6*(2), 171–180.

Juninda, T., Mustakim, & Andri, E. (2019). Penerapan Algoritma K-Medoids untuk Pengelompokan Penyakit di Pekanbaru Riau. *Seminar Nasional Teknologi Informasi Komunikasi Dan Industri (SNTIKI)*, *November*, 42–49.

Muningsih, E., & Kiswati, S. (2015). Penerapan Metode K-Means Untuk Clustering Produk Online Shop Dalam Penentuan Stok Barang. *Jurnal Bianglala Informatika*, *3*(1). (17 Mei 2020)

Silitonga, P. D. P., & Sri, I. (2017). Klusterisasi Pola Penyebaran Penyakit Pasien Berdasarkan Usia Pasien Dengan Menggunakan K-Means Clustering. *Jurnal Times*, *VI*(2), 2005–2008.

Sundari, S. S., & Ariani, N. (2019). Penerapan Data Mining Untuk Pengelompokan Penyakit Dengan Algoritma Fuzzy C-Means (Studi Kasus : UPT Puskesmas Salawu). *Jurnal VOI (Voice of Informatics)*, *8*(2), 63–76.

Triyanto, W. A. (2015). Algoritma K-Medoids Untuk Penentuan Strategi Pemasaran. *Jurnal SIMETRIS*, *6*(1), 183–188.