

Terbit online pada laman : <http://teknosi.fti.unand.ac.id/>

Jurnal Nasional Teknologi dan Sistem Informasi

| ISSN (Print) 2460-3465 | ISSN (Online) 2476-8812 |



Artikel Penelitian

Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru

Gustientiedina^a, M.Hasmil Adiya^a, Yenny Desnelita^b

^aTeknik Informatika, Sekolah Tinggi Ilmu Komputer Pelita Indonesia, Pekanbaru, 28127, Indonesia

^bSistim Informasi, Sekolah Tinggi Ilmu Komputer Pelita Indonesia, Pekanbaru, 28127, Indonesia

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 29 September 2018

Revisi Akhir: 06 Februari 2019

Diterbitkan Online: 30 April 2019

KATA KUNCI

Data Mining,
Clustering,
K-Means,
Obat-obatan

KORESPONDENSI

Telepon: +628127674092

E-mail: gustientiedina@lecturer.pelitaindonesia.ac.id

A B S T R A C T

Perencanaan kebutuhan obat-obatan yang tepat akan membuat pengadaan obat-obatan menjadi efektif dan efisien sehingga tersedia obat-obatan dengan jenis dan jumlah yang cukup sesuai dengan kebutuhan serta dapat diperoleh pada saat yang diperlukan. *Clustering* dalam data mining dapat digunakan untuk menganalisa pemakaian obat-obatan, perencanaan dan pengendalian obat-obatan di rumah sakit. Metode yang akan dipakai untuk *clustering* data obat-obatan adalah algoritma *K-Means* yang merupakan metode data *clustering* non hirarki yang mempartisi data ke dalam *cluster* sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok lain. Tujuan dari penelitian ini adalah untuk mengelompokkan data obat-obatan di Rumah Sakit Umum Daerah Pekanbaru yang dapat digunakan sebagai referensi dalam pengambilan keputusan dalam perencanaan dan pengendalian pasokan medis di rumah sakit tersebut.

1. PENDAHULUAN

Data mining menganalisa sejumlah besar kumpulan data observasi, menemukan suatu hubungan tidak terduga serta dapat merangkum data dengan cara yang baru sehingga dapat berguna dan dimengerti bagi pengguna [1]. Penggunaan sistem informasi terdistribusi mengarah pada pembangunan koleksi data besar di berbagai bidang. Data obat terus berkembang dengan pembaruan informasi di mana data yang ada digabungkan dengan data masa depan, sehingga ukuran data akan terus bertambah seiring dengan seringnya pembaruan. Kemudian akan ada aliran data yang besar, aliran data adalah aliran informasi yang berkelanjutan karena volume besar arus data penting untuk membangun algoritma data mining yang berfungsi secara *efisien* dengan sejumlah besar data.

Perencanaan akan kebutuhan obat-obatan adalah salah satu aspek yang penting untuk menentukan pengelolaan obat-obatan, hal tersebut akan mempengaruhi pengadaan, pendistribusian dan pemakaian obat-obatan. Perencanaan akan kebutuhan obat-obatan yang tepat dapat membuat pengadaan obat – obatan menjadi *efektif* dan *efisien* sehingga obat-obatan dapat tersedia dengan cukup sesuai kebutuhan serta dapat diperoleh pada saat yang diperlukan. *Clustering* merupakan bidang penelitian dalam analisis dan data mining. Pada *clustering* partisi dari objek data yang mempunyai karakteristik sama akan dikelompokkan pada satu kelompok dan data yang memiliki karakteristik berbeda akan dikelompokkan pada kelompok yang lainnya. Salah satu metode pada *clustering* yang terkenal daripada algoritma *clustering* lainnya adalah K-means karena kesederhanaan algoritma dan efisiensinya [2]. Kesederhanaan pada algoritma tersebut yang membuat algoritma K-means dapat diaplikasikan

diberbagai bidang, seperti penelitian yang dilakukan oleh penelitian sebelumnya [3] tentang penerapan data mining dalam Pengelompokan Penderita Thalassaemia.

Data mining merupakan proses dengan kecerdasan buatan, teknik statistik, matematika, dan machine learning untuk mengidentifikasi dan mengekstraksi dan suatu informasi yang bermanfaat dan pengetahuan terkait dari berbagai database yang besar [4]. *Clustering* termasuk salah satu metode dari data mining [5] dan *clustering* telah menjadi instrumen yang valid untuk memecahkan masalah kompleks ilmu komputer dan statistik [6]. *Clustering* adalah proses pengelompokan titik-titik data kedalam dua kelompok atau lebih sehingga titik-titik data yang termasuk didalam kelompok yang sama lebih mirip satu sama lain daripada didalam kelompok yang berbeda, hanya berdasarkan informasi yang tersedia dengan poin data [7].

K-Means yaitu salah satu dari metode pengelompokan data nonhierarki (sekatan) yang dapat mempartisi data kedalam bentuk dua kelompok ataupun lebih. Metoda tersebut akan mempartisi data kedalam suatu kelompok dimana data yang berkarakteristik sama akan dimasukkan kedalam satu kelompok sama sedangkan data yang memiliki karakteristik yang berbeda akan dikelompokkan kedalam kelompok lainnya. Tujuan dari pengelompokan yaitu untuk meminimalkan dari fungsi objektif yang diset dalam proses pengelompokan, pada umumnya akan berusaha meminimalkan variasi didalam suatu kelompok dan memaksimalkan variasi antar kelompok [8].

Algoritma K-Means merupakan salah satu dari algoritma yang banyak digunakan dalam pengelompokan karena kesederhanaan dan efisiensi [9] dan diakui sebagai salah satu dari 10 algoritma data mining teratas oleh IEEE [10].

Berdasarkan dari uraian diatas, maka penelitian yang akan dilakukan bertujuan untuk mengetahui clusterisasi data obat dengan teknik data mining, dimana pengolahan data menggunakan algoritma k-means. Batasan masalah dari penelitian yang dilakukan adalah data yang akan digunakan yaitu data obat-obatan dari Rumah Sakit Umum Daerah Kota Pekanbaru. Analisis dilakukan dengan menggunakan Algoritma K-Means. Output yang dihasilkan adalah kelompok obat-obatan yang pemakaiannya rendah, sedang, dan tinggi. Tujuan penelitian ini untuk pengelompokan data obat-obatan pada rumah sakit berdasarkan pemakaian obat yang rendah, sedang, dan tinggi.

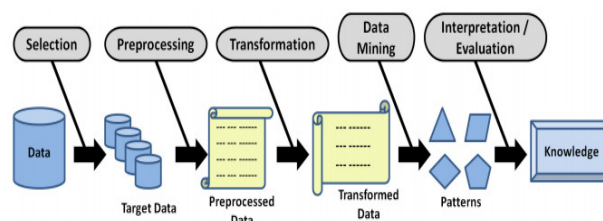
Hasil penelitian ini bermanfaat bagi bagian yang terkait dengan perencanaan kebutuhan obat-obatan pada rumah sakit, yang membutuhkan data jenis-jenis kelompok obat-obatan mana sajakah yang pemakaiannya rendah, sedang, dan tinggi setiap tahun nya, sehingga dapat digunakan sebagai acuan perencanaan kebutuhan obat-obatan untuk kedepannya.

2. METODE

Objek penelitian ini menggunakan data dari Laporan Pemakaian dan Lembar Permintaan Obat (LPLPO) RSUD Arifin Achmad Pekanbaru tahun 2017.

2.1. Knowledge Discovery In Database (KDD)

KDD adalah metode yang digunakan untuk dapat memperoleh pengetahuan yang berasal dari database yang ada. Hasil pengetahuan yang diperoleh dapat dimanfaatkan untuk basis pengetahuan (*knowledge base*) yang digunakan dalam keperluan mengambil keputusan. Secara lebih detail, proses KDD seperti pada gambar berikut ini yang diadopsi dari [5]:



Gambar 3. Proses Knowledge Discovery In Database (KDD)

1. *Selection*
Selection digunakan untuk menentukan variabel yang akan diambil agar tidak ada kesamaan dan terjadi perulangan yang tidak diperlukan dalam pengolahan data mining.
2. *Preprocessing*
Pada preprocessing terdapat dua tahap, yaitu sebagai berikut :
 - a. *Data Cleaning*
Menghilangkan data yang tidak diperlukan seperti menangani *missing value*, *noise* data serta menangani data – data yang tidak konsisten dan *relevant*.
 - b. *Data Integration*
Dilakukan terhadap atribut yang mengidentifikasi entitas yang unik.
3. *Transformation*
Merubah data sesuai format *ekstension* yang sesuai dalam pengolahan data mining karena beberapa metode pada data mining memerlukan format khusus sebelum dapat diproses pada data mining.
4. *Data mining*
Proses utama pada metode yang diterapkan untuk mendapatkan pengetahuan baru dari data yang diproses. Pada penelitian ini diterapkan teknik *clustering* yaitu metode K-Means Clustering.
5. *Evaluation/ Interpretation*
Mengidentifikasi pola – polayang menarik kedalam *knowdge base* yang diidentifikasi. Pada tahap ini, menghasilkan pola – polakhas maupun model prediksi yang dievaluasi untuk menilai kajian yang ada sudah memenuhi target yang diinginkan.
6. *Knowledge*
Pola-pola yang dihasilkan akan dipresentasikan kepada pengguna. Pada tahapan ini pengetahuan baru yang dihasilkan bisa dipahami semua orang yang akan dijadikan acuan pengambilan keputusan.

2.2. Proses Algoritma K-Means

1. Tentukan k sebagai jumlah cluster yang akan dibentuk, menggunakan metode elbow criterion dengan rumus sebagai berikut :

$$SSE = \sum_{k=1}^K \sum_{x_i \in S_k} \|N_i - C_k\|^2 \quad (1)$$

2. Tentukan k titik pusat cluster (*centroid*) awal yang dilakukan secara *random*. Penentuan pada *centroid* awal dilakukan secara acak atau *random* dari objek yang tersedia sebanyak k cluster, untuk menghitung *centroid cluster* ke- i berikutnya, digunakan rumus sebagai berikut:

$$v = \frac{\sum_{i=1}^n X_i}{n} : i = 1, 2, 3, \dots, n \quad (2)$$

3. Menghitung jarak dari setiap objek ke masing – masing *centroid* dari masing – masing *cluster* menggunakan Euclidean Distance, dengan rumus sebagai berikut:

$$d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} : i = 1, 2, 3, \dots, n \quad (3)$$

4. Alokasikan dari masing – masing objek kedalam *centroid* paling dekat. Untuk pengalokasian objek kedalam masing – masing *cluster* pada saat iterasi secara umum dilakukan dengan cara hard k -means dimana secara tegas setiap objek dinyatakan sebagai anggota *cluster* dengan mengukur jarak kedekatan sifatnya terhadap titik pusat *cluster* tersebut.

5. Lakukan iterasi dan kemudian tentukan posisi *centroid* baru dengan menggunakan persamaan.

6. Ulangi langkah tiga jika posisi *centroid* baru tidak sama.

3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data yang berasal dari Laporan Pemakaian dan Lembar Permintaan Obat (LPLPO) RSUD Arifin Achmad Pekanbaru tahun 2017. Parameter yang digunakan adalah nama obat, unit, jumlah item obat setiap bulan dari Januari hingga Desember. Proses *cluster* dilakukan untuk mengetahui pola kebutuhan obat rumah sakit.

Tabel 1. Data *Cleaning*

No	Nama Obat	Pemakaian
1	Acarbose 100 mg	1800
2	Acarbose 50mg	0
3	Acetylcysteine kaps	0
4	Albendazol 400 mg	0
5	Albendazol 400 mg (ASKES)	0
6	Allopurinol 100 mg	500
7	Alopurinol 300 mg	3420
8	Alprazolam 0,5 mg	2200
9	Ambroksol 30 mg	6700
10	Ambroksol syrup 15 mg/5 ml	421
...
...
295	Zink	300

3.1. Data *Integration*

Pada tahap data *integration* ini dilakukan penggabungan data-data obat yang terdapat pada laporan setiap bulannya kedalam satu tahun menjadi total pemakaian.. Untuk pemakaian obat yang tidak ada jumlah pemakaian obatnya akan diberi nilai 0, seperti yang terlihat di tabel 2.

3.2. Data *Cleaning*

Pada tahap data *cleaning* ini yang dilakukan adalah membuang atribut yang tidak relevan atau tidak konsisten. Atribut yang dibuang dari dataset yang sudah ada yaitu satuan, harga satuan, stock awal, masuk, stock akhir. Atribut yang digunakan yaitu nama obat, pemakaian, dataset yang telah di *cleaning* dapat dilihat pada tabel 1.

Tabel 2. Data *Integration*

No	Nama Obat	Bulan				Total Pemakaian
		Jan	Feb	...	Des	
1	Acarbose 100 mg	1800	0	...	4000	22100
2	Acarbose 50mg	0	13200	...	3100	52400
3	Acetylcysteine kaps	0	0	...	0	1200
4	Albendazol 400 mg	0	0	...	0	30
5	Albendazol 400 mg (ASKES)	0	0	...	0	30
6	Allopurinol 100 mg	500	0	...	0	7100
7	Alopurinol 300 mg	3420	0	...	1300	10140
8	Alprazolam 0,5 mg	2200	4900	...	100	22500
9	Ambroksol 30 mg	6700	2300	...	5300	46300
10	Ambroksol syrup 15 mg/5 ml	421	250	...	0	2357
...
...
295	Zink	300	200	...	3900	4700

3.3. Proses *Training*

1. Iterasi 1

Tentukan K jumlah pusat *cluster* secara acak (random). Pada percobaan pertama ini ditentukan 3 data secara acak sebagai titik pusat awal untuk perhitungan jarak dari seluruh kelompok *cluster* yang akan dibentuk.

Jumlah cluster = 3 (sedikit, sedang, tinggi)

Jumlah data = 295

Jumlah atribut = 1

Data acak tersebut dapat dilihat pada tabel 3.

Tabel 3. Data Pusat *Cluster* Iterasi 1

No	Nama Obat	Total Pemakaian	Cluster
131	Infus Nacl 0.9% 1000 ml	148	C1
7	Alopurinol 300 mg	10140	C2
290	Vitamin B kompleks	116800	C3

Dari tabel 3 dapat dilihat bahwa masing-masing data memiliki nilai pusat cluster yang terlihat pada tabel 4.

Tabel 4. Nilai Pusat Cluster Iterasi 1

Cluster	Pusat
C1	148
C2	10140
C3	116800

2. Hitung jarak tiap data dengan masing-masing cluster pusat dengan menggunakan persamaan (3) yaitu persamaan Euclidean Distance

Hitung Euclidean Distance dari semua data ke tiap titik pusat pertama:

$$C1 (1) = \sqrt{(148-22100)^2} = 21952$$

$$C1 (2) = \sqrt{(148-52400)^2} = 52252$$

$$C1 (3) = \sqrt{(148-1200)^2} = 1052$$

$$C1 (4) = \sqrt{(148-30)^2} = 118$$

$$C1 (5) = \sqrt{(148-30)^2} = 118$$

$$C1 (6) = \sqrt{(148-7100)^2} = 6952$$

$$C1 (7) = \sqrt{(148-10140)^2} = 9992$$

$$C1 (8) = \sqrt{(148-22500)^2} = 22352$$

$$C1 (9) = \sqrt{(148-46300)^2} = 46152$$

$$C1 (10) = \sqrt{(148-2357)^2} = 2209$$

.....

.....

$$C1 (295) = \sqrt{(148-4700)^2} = 4552$$

Hitung Euclidean Distance dari semua data ke tiap titik pusat kedua :

$$C2 (1) = \sqrt{(10140-22100)^2} = 11960$$

$$C2 (2) = \sqrt{(10140-52400)^2} = 42260$$

$$C2 (3) = \sqrt{(10140-1200)^2} = 4940$$

$$C2 (4) = \sqrt{(10140-30)^2} = 10110$$

$$C2 (5) = \sqrt{(10140-30)^2} = 10110$$

$$C2 (6) = \sqrt{(10140-7100)^2} = 3040$$

$$C2 (7) = \sqrt{(10140-10140)^2} = 0$$

$$C2 (8) = \sqrt{(10140-22500)^2} = 12360$$

$$C2 (9) = \sqrt{(10140-46300)^2} = 36160$$

$$C2 (10) = \sqrt{(10140-2357)^2} = 7783$$

.....

.....

$$C2 (295) = \sqrt{(10140-4700)^2} = 5440$$

Hitung Euclidean Distance dari semua data ke tiap titik pusat ketiga :

$$C3 (1) = \sqrt{(116800-22100)^2} = 94700$$

$$C3 (2) = \sqrt{(116800-52400)^2} = 64400$$

$$C3 (3) = \sqrt{(116800-1200)^2} = 115600$$

$$C3 (4) = \sqrt{(116800-30)^2} = 116770$$

$$C3 (5) = \sqrt{(116800-30)^2} = 116770$$

$$C3 (6) = \sqrt{(116800-7100)^2} = 109700$$

$$C3 (7) = \sqrt{(116800-10140)^2} = 106660$$

$$C3 (8) = \sqrt{(116800-22500)^2} = 94300$$

$$C3 (9) = \sqrt{(116800-46300)^2} = 70500$$

$$C3 (10) = \sqrt{(116800-2357)^2} = 114443$$

.....

.....

$$C3 (295) = \sqrt{(116800-4700)^2} = 112100$$

Perhitungan diatas menghasilkan tabel jarak seperti tabel 5.

Tabel 5. Perhitungan Euclidean Distance Iterasi 1

Nama Obat	C1	C2	C3
Acarbose 100 mg	21952	11960	94700
Acarbose 50mg	52252	42260	64400
Acetylcysteine kaps	1052	8940	115600
Albendazol 400 mg	118	10110	116770
Albendazol 400 mg (ASKES)	118	10110	116770
Allopurinol 100 mg	6952	3040	109700
Alopurinol 300 mg	9992	0	106660
Alprazolam 0,5 mg	22352	12360	94300
Ambroksol 30 mg	46152	36160	70500
Ambroksol syrup 15 mg/5 ml	2209	7783	114443
...
...
Zink	4552	5440	112100

Anggota dipilih dari terkecil diantara 3 cluster jika terkecil pada bagian C1 maka termasuk sebagai anggota C1 yaitu sebanyak 163 data, jika terkecil pada bagian C2 maka termasuk sebagai anggota C2 yaitu sebanyak 113 data, dan jika terkecil pada bagian C3 maka termasuk sebagai anggota C3 yaitu sebanyak 19 data.

3. Lakukan iterasi ke 2

Tentukan posisi centroid baru dengan cara menghitung rata-rata dari data-data yang ada pada centroid yang sama atau anggota yang sama.

Rerata = Total / Jumlah Data

$$\text{Anggota C1} \rightarrow \text{Rerata} = 218617/163 = 1341.21$$

$$\text{Anggota C2} \rightarrow \text{Rerata} = 2289450/113 = 20260.62$$

$$\text{Anggota C3} \rightarrow \text{Rerata} = 2170019/19 = 114211.53$$

Dari hasil diatas dapat dilihat bahwa masing-masing data memiliki titik pusat cluster baru yang terlihat pada tabel 6.

Tabel 6. Nilai Pusat Cluster Iterasi 2

Cluster	Pusat
C1	1341.21
C2	20260.62
C3	114211.53

Selanjutnya hitung jarak tiap data dengan masing-masing cluster pusat. Perhitungannya sama dengan tahap perhitungan pada iterasi 1.

Hitung Euclidean Distance dari semua data ke tiap titik pusat pertama :

$$C1 (1) = \sqrt{(1341.21-22100)^2} = 20758.79$$

$$C1 (2) = \sqrt{(1341.21-52400)^2} = 51058.79$$

$$C1 (3) = \sqrt{(1341.21-1200)^2} = 141.21$$

$$C1 (4) = \sqrt{(1341.21-30)^2} = 1311.21$$

$$C1 (5) = \sqrt{(1341.21-30)^2} = 1311.21$$

$$C1 (6) = \sqrt{(1341.21-7100)^2} = 5758.79$$

$$C1 (7) = \sqrt{(1341.21-10140)^2} = 8798.79$$

$$C1 (8) = \sqrt{(1341.21-22500)^2} = 21.158.79$$

$$C1 (9) = \sqrt{(1341.21-46300)^2} = 44958.79$$

$$C1 (10) = \sqrt{((1341.21-2357)^2)} = 1015.79$$

.....

$$C1 (295) = \sqrt{((1341.21-4700)^2)} = 3358.79$$

Hitung Euclidean Distance dari semua data ke tiap titik pusat kedua :

$$C2 (1) = \sqrt{((20260.62-22100)^2)} = 1839.38$$

$$C2 (2) = \sqrt{((20260.62-52400)^2)} = 32139.38$$

$$C2 (3) = \sqrt{((20260.62-1200)^2)} = 19060.62$$

$$C2 (4) = \sqrt{((20260.62-30)^2)} = 20230.62$$

$$C2 (5) = \sqrt{((20260.62-30)^2)} = 20230.62$$

$$C2 (6) = \sqrt{((20260.62-7100)^2)} = 13160.62$$

$$C2 (7) = \sqrt{((20260.62-10140)^2)} = 10120.62$$

$$C2 (8) = \sqrt{((20260.62-22500)^2)} = 2239.38$$

$$C2 (9) = \sqrt{((20260.62-46300)^2)} = 26039.38$$

$$C2 (10) = \sqrt{((20260.62-2357)^2)} = 17903.62$$

.....

$$C2 (295) = \sqrt{((20260.62-4700)^2)} = 15560.62$$

Hitung Euclidean Distance dari semua data ke tiap titik pusat ketiga :

$$C3 (1) = \sqrt{((114211.53-22100)^2)} = 92111.53$$

$$C3 (2) = \sqrt{((114211.53-52400)^2)} = 61811.53$$

$$C3 (3) = \sqrt{((114211.53-1200)^2)} = 113011.5$$

$$C3 (4) = \sqrt{((114211.53-30)^2)} = 114181.5$$

$$C3 (5) = \sqrt{((114211.53-30)^2)} = 114181.5$$

$$C3 (6) = \sqrt{((114211.53-7100)^2)} = 107111.5$$

$$C3 (7) = \sqrt{((114211.53-10140)^2)} = 104071.5$$

$$C3 (8) = \sqrt{((114211.53-22500)^2)} = 91711.53$$

$$C3 (9) = \sqrt{((114211.53-46300)^2)} = 67911.53$$

$$C3 (10) = \sqrt{((114211.53-2357)^2)} = 111854.5$$

.....

$$C3 (295) = \sqrt{((114211.53-4700)^2)} = 109511.5$$

Perhitungan diatas menghasilkan tabel jarak seperti tabel 7.

Tabel 7. Perhitungan Euclidean Distance Iterasi 2

Nama Obat	C1	C2	C3
Acarbose 100 mg	20758.79	1839.38	92111.53
Acarbose 50mg	51058.79	32139.38	61811.53
Acetylcysteine kaps	141.21	19060.62	113011.5
Albendazol 400 mg	1311.21	20230.62	114181.5
Albendazol 400 mg (ASKES)	1311.21	20230.62	114181.5
Allopurinol 100 mg	5758.79	13160.62	107111.5
Alopurinol 300 mg	8798.79	10120.62	104071.5
Alprazolam 0,5 mg	21158.79	2239.38	91711.53
Ambroksol 30 mg	44958.79	26039.38	67911.53
Ambroksol syrup 15 mg/5 ml	1015.79	17903.62	111854.5
...
...
Zink	3358.79	15560.62	109511.5

Anggota dipilih dari terkecil diantara 3 cluster jika terkecil pada bagian C1 maka termasuk sebagai anggota C1 yaitu sebanyak 205 data, jika terkecil pada bagian C2 maka termasuk sebagai anggota

C2 yaitu sebanyak 72 data, dan jika terkecil pada bagian C3 maka termasuk sebagai anggota C3 yaitu sebanyak 18 data.

4. Karena hasil iterasi ke-2 tidak sama dengan iterasi ke-1 sehingga perlu dilakukan kembali perhitungan ke iterasi ke-3 dan seterusnya sampai mendapatkan hasil yang sama.

Lakukan iterasi ke 3

Tentukan posisi *centroid* baru dengan cara menghitung rata-rata dari data-data yang ada pada *centroid* yang sama atau anggota yang sama.

Rerata = Total / Jumlah Data

$$\text{Anggota C1} \rightarrow \text{Rerata} = 527552/205 = 2573.42$$

$$\text{Anggota C2} \rightarrow \text{Rerata} = 2044017/72 = 28389.13$$

$$\text{Anggota C3} \rightarrow \text{Rerata} = 2106517/18 = 117028.72$$

Dari hasil diatas dapat dilihat bahwa masing-masing data memiliki titik pusat *cluster* baru yang terlihat pada tabel 8.

Tabel 8. Nilai Pusat *Cluster* Iterasi 3

Cluster	Pusat
C1	2573.42
C2	28389.13
C3	117028.72

Selanjutnya hitung jarak tiap data dengan masing-masing *cluster* pusat. Perhitungannya sama dengan tahap perhitungan pada iterasi 1 dan iterasi 2.

Hitung Euclidean Distance dari semua data ke tiap titik pusat pertama :

$$C1 (1) = \sqrt{((2573.42-22100)^2)} = 19526.58$$

$$C1 (2) = \sqrt{((2573.42-52400)^2)} = 49826.58$$

$$C1 (3) = \sqrt{((2573.42-1200)^2)} = 1373.42$$

$$C1 (4) = \sqrt{((2573.42-30)^2)} = 2543.42$$

$$C1 (5) = \sqrt{((2573.42-30)^2)} = 2543.42$$

$$C1 (6) = \sqrt{((2573.42-7100)^2)} = 4526.58$$

$$C1 (7) = \sqrt{((2573.42-10140)^2)} = 7566.58$$

$$C1 (8) = \sqrt{((2573.42-22500)^2)} = 19926.58$$

$$C1 (9) = \sqrt{((2573.42-46300)^2)} = 43726.58$$

$$C1 (10) = \sqrt{((2573.42-2357)^2)} = 216.42$$

.....

$$C1 (295) = \sqrt{((2573.42-4700)^2)} = 2126.58$$

Hitung Euclidean Distance dari semua data ke tiap titik pusat kedua :

$$C2 (1) = \sqrt{((28389.13-22100)^2)} = 6289.13$$

$$C2 (2) = \sqrt{((28389.13-52400)^2)} = 24010.87$$

$$C2 (3) = \sqrt{((28389.13-1200)^2)} = 27189.13$$

$$C2 (4) = \sqrt{((28389.13-30)^2)} = 28359.13$$

$$C2 (5) = \sqrt{((28389.13-30)^2)} = 28359.13$$

$$C2 (6) = \sqrt{((28389.13-7100)^2)} = 21289.13$$

$$C2 (7) = \sqrt{((28389.13-10140)^2)} = 18249.13$$

$$C2 (8) = \sqrt{((28389.13-22500)^2)} = 5889.13$$

$$C2 (9) = \sqrt{((28389.13-46300)^2)} = 17910.87$$

$$C2 (10) = \sqrt{((28389.13-2357)^2)} = 26032.13$$

.....

$$C2 (295) = \sqrt{((28389.13-4700)^2)} = 23689.13$$

Hitung Euclidean Distance dari semua data ke tiap titik pusat ketiga :

$$\begin{aligned}
 C3 (1) &= \sqrt{(117028.72-22100)^2}) = 94928.72 \\
 C3 (2) &= \sqrt{(117028.72-52400)^2}) = 64628.72 \\
 C3 (3) &= \sqrt{(117028.72-1200)^2}) = 115828.7 \\
 C3 (4) &= \sqrt{(117028.72-30)^2}) = 116998.7 \\
 C3 (5) &= \sqrt{(117028.72-30)^2}) = 116998.7 \\
 C3 (6) &= \sqrt{(117028.72-7100)^2}) = 109928.7 \\
 C3 (7) &= \sqrt{(117028.72-10140)^2}) = 106888.7 \\
 C3 (8) &= \sqrt{(117028.72-22500)^2}) = 94528.72 \\
 C3 (9) &= \sqrt{(117028.72-46300)^2}) = 70728.72 \\
 C3 (10) &= \sqrt{(117028.72-2357)^2}) = 114671.7 \\
 &\dots\dots\dots \\
 &\dots\dots\dots \\
 C3 (295) &= \sqrt{(117028.72-4700)^2}) = 112328.7
 \end{aligned}$$

Perhitungan diatas menghasilkan tabel jarak seperti tabel 9.

Tabel 9. Perhitungan Euclidean Distance Iterasi 3

Nama Obat	C1	C2	C3
Acarbose 100 mg	19526.58	6289.13	94928.72
Acarbose 50mg	49826.58	24010.87	64628.72
Acetylcysteine kaps	1373.42	27189.13	115828.7
Albendazol 400 mg	2543.42	28359.13	116998.7
Albendazol 400 mg (ASKES)	2543.42	28359.13	116998.7
Allopurinol 100 mg	4526.58	21289.13	109928.7
Alopurinol 300 mg	7566.58	18249.13	106888.7
Alprazolam 0,5 mg	19926.58	5889.13	94528.72
Ambroksol 30 mg	43726.58	17910.87	70728.72
Ambroksol sirup 15 mg/5 ml	216.42	26032.13	114671.7
...
...
Zink	2126.58	23689.13	112328.7

Anggota dipilih dari terkecil diantara 3 cluster jika terkecil pada bagian C1 maka termasuk sebagai anggota C1 yaitu sebanyak 216 data, jika terkecil pada bagian C2 maka termasuk sebagai anggota C2 yaitu sebanyak 63 data, dan jika terkecil pada bagian C3 maka termasuk sebagai anggota C3 yaitu sebanyak 16 data.

5. Karena hasil iterasi ke-3 tidak sama dengan iterasi ke-2 sehingga perlu dilakukan kembali perhitungan ke iterasi ke-4 dan seterusnya sampai mendapatkan hasil yang sama.

Lakukan iterasi ke 4

Tentukan posisi *centroid* baru dengan cara menghitung rata-rata dari data-data yang ada pada *centroid* yang sama atau anggota yang sama.

Rerata = Total / Jumlah Data

Aggota C1 → Rerata = $664881/217 = 3078.15$

Aggota C2 → Rerata = $2044218/63 = 32447.90$

Aggota C3 → Rerata = $1968987/16 = 123061.69$

Dari hasil diatas dapat dilihat bahwa masing-masing data memiliki titik pusat *cluster* baru yang terlihat pada tabel 10.

Tabel 10. Nilai Pusat *Cluster* Iterasi 4

Cluster	Pusat
C1	3078.15
C2	32447.90
C3	123061.69

Selanjutnya hitung jarak tiap data dengan masing-masing *cluster* pusat. Perhitungannya sama dengan tahap perhitungan pada iterasi 1, iterasi 2 dan iterasi 3.

Hitung Euclidean Distance dari semua data ke tiap titik pusat pertama :

$$\begin{aligned}
 C1 (1) &= \sqrt{(3078.15-22100)^2}) = 19021.85 \\
 C1 (2) &= \sqrt{(3078.15-52400)^2}) = 49321.85 \\
 C1 (3) &= \sqrt{(3078.15-1200)^2}) = 1875.15 \\
 C1 (4) &= \sqrt{(3078.15-30)^2}) = 3048.15 \\
 C1 (5) &= \sqrt{(3078.15-30)^2}) = 3048.15 \\
 C1 (6) &= \sqrt{(3078.15-7100)^2}) = 4021.85 \\
 C1 (7) &= \sqrt{(3078.15-10140)^2}) = 7061.85 \\
 C1 (8) &= \sqrt{(3078.15-22500)^2}) = 19421.85 \\
 C1 (9) &= \sqrt{(2573.15-46300)^2}) = 43221.85 \\
 C1 (10) &= \sqrt{(3078.15-2357)^2}) = 721.15 \\
 &\dots\dots\dots \\
 &\dots\dots\dots \\
 C1 (295) &= \sqrt{(3078.15-4700)^2}) = 1621.85
 \end{aligned}$$

Hitung Euclidean Distance dari semua data ke tiap titik pusat kedua :

$$\begin{aligned}
 C2 (1) &= \sqrt{(32447.90-22100)^2}) = 10347.9 \\
 C2 (2) &= \sqrt{(32447.90-52400)^2}) = 19952.1 \\
 C2 (3) &= \sqrt{(32447.90-1200)^2}) = 31247.9 \\
 C2 (4) &= \sqrt{(32447.90-30)^2}) = 32417.9 \\
 C2 (5) &= \sqrt{(32447.90-30)^2}) = 32417.9 \\
 C2 (6) &= \sqrt{(32447.90-7100)^2}) = 25347.9 \\
 C2 (7) &= \sqrt{(32447.90-10140)^2}) = 22307.9 \\
 C2 (8) &= \sqrt{(32447.90-22500)^2}) = 9947.9 \\
 C2 (9) &= \sqrt{(32447.90-46300)^2}) = 13852.1 \\
 C2 (10) &= \sqrt{(32447.90-2357)^2}) = 30090.9 \\
 &\dots\dots\dots \\
 &\dots\dots\dots \\
 C2 (295) &= \sqrt{(32447.90-4700)^2}) = 27747.9
 \end{aligned}$$

Hitung Euclidean Distance dari semua data ke tiap titik pusat ketiga :

$$\begin{aligned}
 C3 (1) &= \sqrt{(123061.69-22100)^2}) = 100961.7 \\
 C3 (2) &= \sqrt{(123061.69-52400)^2}) = 70661.69 \\
 C3 (3) &= \sqrt{(123061.69-1200)^2}) = 121861.7 \\
 C3 (4) &= \sqrt{(123061.69-30)^2}) = 123031.7 \\
 C3 (5) &= \sqrt{(123061.69-30)^2}) = 123031.7 \\
 C3 (6) &= \sqrt{(123061.69-7100)^2}) = 115961.7 \\
 C3 (7) &= \sqrt{(123061.69-10140)^2}) = 112921.7 \\
 C3 (8) &= \sqrt{(123061.69-22500)^2}) = 100561.7 \\
 C3 (9) &= \sqrt{(123061.69-46300)^2}) = 76761.69 \\
 C3 (10) &= \sqrt{(123061.69-2357)^2}) = 120704.7 \\
 &\dots\dots\dots \\
 &\dots\dots\dots \\
 C3 (295) &= \sqrt{(123061.69-4700)^2}) = 118361.7
 \end{aligned}$$

Perhitungan diatas menghasilkan tabel jarak seperti tabel 11. Anggota dipilih dari terkecil diantara 3 cluster jika terkecil pada bagian C1 maka termasuk sebagai anggota C1 yaitu sebanyak 224

data, jika terkecil pada bagian C2 maka termasuk sebagai anggota C2 yaitu sebanyak 55 data, dan jika terkecil pada bagian C3 maka termasuk sebagai anggota C3 yaitu sebanyak 16 data.

Tabel 11. Perhitungan Euclidean Distance Iterasi 4

Nama Obat	C1	C2	C3
Acarbose 100 mg	19021.85	10347.9	100961.7
Acarbose 50mg	49321.85	19952.1	70661.69
Acetylcysteine kaps	1875.15	31247.9	121861.7
Albendazol 400 mg	3048.15	32417.9	123031.7
Albendazol 400 mg (ASKES)	3048.15	32417.9	123031.7
Allopurinol 100 mg	4021.85	25347.9	115961.7
Alopurinol 300 mg	7061.85	22307.9	112921.7
Alprazolam 0,5 mg	19421.85	9947.9	100561.7
Ambroksol 30 mg	43221.85	13852.1	76761.69
Ambroksol syrup 15 mg/5 ml	721.15	30090.9	120704.7
...
...
Zink	1621.85	27747.9	118361.7

6. Iterasi ke 4 dengan iterasi ke-3 hasilnya sudah sama maka tidak perlu lagi melanjutkan ke iterasi ke-5 atau cukup berhenti di iterasi ke-4.

Clusterisasi data obat yang dilakukan dengan algoritma k-means didapatkan hasil cluster nya setelah melakukan iterasi ke-4 yaitu terdapat kelompok obat yang pemakaian sedikit terdapat pada cluster 1 yang memiliki 224 anggota, kelompok obat yang pemakaian sedang terdapat pada cluster 2 yang memiliki 55 anggota, dan kelompok obat yang pemakaian tinggi terdapat pada cluster 3 yang memiliki 16 anggota.

Tabel 12. Anggota Cluster 1 (Obat Pemakaian Sedikit)

No	Nama Obat	Pemakaian
1	Acetylcysteine kaps	1200
2	Albendazol 400 mg	30
3	Albendazol 400 mg (ASKES)	30
4	Allopurinol 100 mg	7100
5	Alopurinol 300 mg	10140
6	Ambroksol syrup 15 mg/5 ml	2357
7	Amikasin 250 mg	660
8	Amikasin 500 mg inj	1370
9	Aminofilin 24 mg/ml inj	1566
10	Amitriptilin 25 mg	5000
...
...
224	Zink	4700

Tabel 13. Anggota Cluster 2 (Obat Pemakaian Sedang)

No	Nama Obat	Pemakaian
1	Acarbose 100mg	22100
2	Acarbose 50mg	52400
3	Alprazolam 0,5 mg	22500
4	Ambroksol 30 mg	46300

5	Aminofilin 200 mg	26900
6	Amitriptilin 25 mg	22200
7	Amoksisilin 500 mg	26300
8	Betahistin 6mg	24930
9	Bisoprolol 5 mg tab	69330
10	Candesartan 16 mg	20880
...
...
55	Vitamin C 50 mg	46300

Tabel 14. Anggota Cluster 3 (Obat Pemakaian Tinggi)

No	Nama Obat	Pemakaian
1	Amlodipin 10 mg	144380
2	Amlodipin 5 mg	112500
3	Asam Folat 1 mg	105000
4	Asam Mefenamat 500 mg	105300
5	Clopidogrel 75 mg	79050
6	Furosemida 40 mg	83400
7	Infus Nacl 0.9% 500 ml	97867
8	Infus Ringer Laktat	81700
9	Isosorbid dinitrat	181400
10	Kalsium Laktat 500 mg	107000
...
...
16	Vitamin B kompleks	116800

4. KESIMPULAN

Dari hasil clusterisasi pada data obat – obatan dapat ditarik kesimpulan bahwa kelompok obat yang termasuk pemakaian sedikit rata rata permintaan obat setiap tahunnya kurang dari 18000 buah, dan obat yang termasuk pemakaian sedang rata rata permintaan obat setiap tahunnya diantara 18000–70000 buah, sedangkan obat yang masuk kedalam kelompok obat yang pemakaian tinggi rata – rata permintaan obat setiap tahunnya diatas 70000 buah.

Dari analisa cluster diatas mungkin perlu dilakukan lagi penelitian lanjutan agar clusterisasi data obat dapat dilakukan secara lebih valid dengan menetapkan nilai centroid terbaik.

UCAPAN TERIMA KASIH

Ucapan terima kasih kepada Direktorat Riset dan Pengabdian Masyarakat Direktorat Jendral Penguatan Riset dan Pengembangan Kementerian Riset, Teknologi dan Pendidikan Tinggi Republik Indonesia telah mendanai penelitian ini dan Sekolah Tinggi Ilmu Komputer Indonesia.

DAFTAR PUSTAKA

- [1] R. Mythily, A. Banu, and S. Raghunathan, "Clustering models for data stream mining," *Procedia Comput. Sci.*, vol. 46, no. Icict 2014, pp. 619–626, 2015.
- [2] T. H. Sardar and Z. Ansari, "An analysis of MapReduce

- efficiency in document clustering using parallel K-means algorithm,” *Futur. Comput. Informatics J.*, pp. 1–10, 2018.
- [3] Sulastrri H, Gufroni AI. Penerapan Data Mining Dalam Pengelompokan Penderita Thalassaemia. *Jurnal Teknologi dan Sistem Informasi*. 2017 Sep 26;3(2):299-305.
- [4] Gurunescu, F. (2011). *Data Mining Concepts, Models and Techniques* (Volume 12 ed.). Verlag Berlin Heidelberg.
- [5] F. Gullo, “From patterns in data to knowledge discovery: What data mining can do,” *Phys. Procedia*, vol. 62, pp. 18–22, 2015.
- [6] A. Amelio and A. Tagarelli, “Data Mining: Clustering,” *Ref. Modul. Life Sci.*, 2018.
- [7] N. Nidheesh, K. A. Abdul Nazeer, and P. M. Ameer, “An enhanced deterministic K-Means clustering algorithm for cancer subtype prediction from gene expression data,” *Comput. Biol. Med.*, vol. 91, pp. 213–221, 2017.
- [8] Prasetyo, Eko, *Data Mining : Mengolah Data Menjadi Informasi Menggunakan Matlab*, Andi, 2014
- [9] J.A. Harding, M. Shahbaz, S. Srinivas, A. Kusiak, Data mining in manufacturing: a review, *J. Manuf. Sci. Eng.* 128 (2006).
- [10] X. Wu, et al., Top 10 algorithms in data mining, *Knowl. Inf. Syst.* 14 (2008) 1–37

NOMENKLATUR

Ni	arti dari normaeuclid (L2)
Ck	arti dari pusat cluster Sk yang dihitung berdasarkan rata-rata jarak titik-titik kluster ke pusat kluster
v	arti dari centroid pada cluster
xi	arti dari objek x ke-i
yi	arti dari data y ke-i
n	arti dari banyaknya objek/jumlah objek yang menjadi anggota cluster

BIODATA PENULIS



Gustientiedina, S. Kom, M. Kom. lahir di Sawahlunto, 2 Agustus 1965. Menyelesaikan pendidikan tingkat atas di SMA Negeri 1 Sawahlunto. Pendidikan S-1 (Teknik Informatika STMIK-AMIK RIAU) lulus tahun 2014 dan S-2 (Ilmu Komputer Universitas Putra Indonesia “YPTK”) lulus tahun 2016. Sedang kuliah S3 (Pendidikan Teknologi Kejuruan UNP).



M. Hasmil Adiya, ST, MAB. lahir di Padang, 9 Maret 1976. Menyelesaikan pendidikan S-1 (Teknik Elektro ITENAS Bandung) lulus tahun 2000 dan S-2 (Bisnis Administrasi ITB) lulus tahun 2005.



Yenny Desnelita, S. Kom., M.Kom. lahir di Sawahlunto, 10 Juni 1069. Menyelesaikan pendidikan tingkat atas di SMA Negeri 1 Sawahlunto. Pendidikan S-1 (Manajemen Informatika UPI Padang) lulus tahun 1993 dan S-2 (Ilmu Komputer UGM) lulus tahun 2004. Sedang kuliah S3 (Pendidikan Teknologi Kejuruan UNP).