

Clustering Data Obat-obatan Menggunakan Algoritma K-means**Nur Khafidhoh^{1*}, Munawarah²**¹Fakultas Teknologi Informasi

Universitas KH.A. Wahab Hasbullah

Email: nurkhafidhohunwaha@gmail.com²Fakultas Teknologi Informasi

Universitas KH.A. Wahab Hasbullah

Email: munawarahunwaha@yahoo.com

©2019 –EPiC Universitas KH. A. Wahab Hasbullah Jombang ini adalah artikel dengan akses terbuka dibawah lisensi CC BY-NC-4.0 (<https://creativecommons.org/licenses/by-nc/4.0/>).

ABSTRAK

Obat-obatan yang beredar di Indonesia memiliki banyak variasi kategori kegunaannya. Berdasarkan kategori obat-obatan diperlukan perencanaan kebutuhan yang tepat untuk membuat pengadaan obat-obatan menjadi efektif dan efisien sehingga ketersediaan obat-obatan dengan jenis jumlah yang cukup sesuai dengan kebutuhan serta dapat diperoleh pada saat yang diperlukan. Penyediaan kebutuhan obat-obatan dibutuhkan metode untuk memudahkan pengelompokan obat-obatan yang pemakaiannya rendah, sedang, tinggi yang dibutuhkan oleh masyarakat. Dalam data mining dapat menggunakan metode Clustering untuk pengelompokan dan algoritma K-means digunakan untuk menganalisa pemakaian obat-obatan, perencanaan ketersediaan dan pengendalian obat-obatan di apotik. Metode pada penelitian ini dengan 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. Hasil dari penelitian ini adalah untuk mengelompokkan data obat-obatan di apotik yang dapat digunakan sebagai salah satu alat bantu penunjang keputusan bagi pihak apotik untuk mengajukan permintaan obat pada periode yang akan datang penyediaan.

Kata kunci : Obat-obatan, Data mining, Clustering, K-means

PENDAHULUAN

Data obat terus berkembang dengan pembaruan informasi dimana data yang ada digabungkan dengan data masa depan, sehingga ukuran data akan terus bertambah seiring dengan seringnya pembaruan. Perencanaan akan kebutuhan obat-obatan adalah salah satu aspek yang penting untuk menentukan pengelolaan obat-obatan, hal tersebut akan mempengaruhi pengadaan pendistribusian dan dilakukan untuk menjamin ketersediaan serta pemerataan obat dengan jenis dan jumlah yang mencukupi agar obat dapat diperoleh dengan cepat pada tempat dan waktu yang tepat pada instansi-instansi yang terkait dengan pelayanan kesehatan, baik itu rumah sakit, puskesmas, dinas kesehatan dan lain sebagainya. Perencanaan akan kebutuhan obat-obatan yang tepat dan membuat pengadaan obat-obatan menjadi efektif dan efisien sehingga obat-obatan dapat tersedia dengan cukup sesuai

kebutuhan serta dapat diperoleh pada saat yang diperlukan. Metode dalam data mining salah satunya *Clustering* yang merupakan bidang penelitian dalam analisis partisi dari objek data yang mempunyai karakteristik sama akan dikelompokkan dan data yang memiliki karakteristik berbeda akan dikelompokkan pada kelompok yang lainnya. Data mining menganalisa kumpulan data sejumlah besar kemudian observasi, menemukan suatu hubungan tidak terduga serta dapat merangkum data dengan cara yang baru sehingga dapat berguna dan dimengerti bagi pengguna. Penggunaan sistem informasi terdistribusi mengarah pada pembangunan koleksi data besar diberbagai bidang. Salah satu algoritma yang digunakan metode clustering adalah K-Means karena kesederhanaan algoritma dan efisiensinya (Patel & Mehta, 2011). Kesederhanaan metode ini membuat algoritma K-means dapat

diaplikasikan pada berbagai bidang (K.Arai and A.R. Barakbah, 2007).

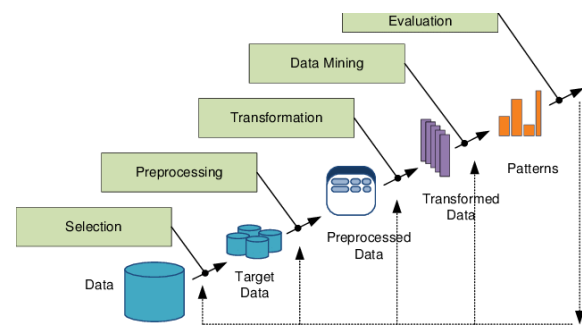
K-Means merupakan salah satu dari metode pengelompokan data non hierarki (sekatan) yang dapat mempartisi data kedalam bentuk dua kelompok ataupun lebih. Metode tersebut akan memartisi data kedalam suatu kelompok dimana data yang berkarakteristik sama akan dimasukkan kedalam satu kelompok sama sedangkan 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.

Berdasarkan dari uraian diatas, maka penelitian yang akan dilakukan bertujuan untuk mengetahui clusterisasi pada obat-obatan dengan teknik data mining, dimana pengolahan data menggunakan algoritma K-means. Batasan masalah dari penelitian yang dilakukan adalah data yang akan digunakan yaitu obat-obatan dari apotik. Analisis dilakukan dengan menggunakan Algoritma K-Means dengan Output yang dihasilkan adalah kelompok obat-obatan yang pemakaiannya rendah, sedang, tinggi. Tujuan penelitian ini untuk pengelompokan data obat-obatan pada rumah sakit berdasarkan pemakaian obat yang rendah, sedang, tinggi. Sehingga dapat digunakan sebagai acuan perencanaan pengambilan keputusan dalam penyediaan kebutuhan obat-obatan untuk periode kedepannya.

METODE

2.1 Knowledge Discovery in Database

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 :



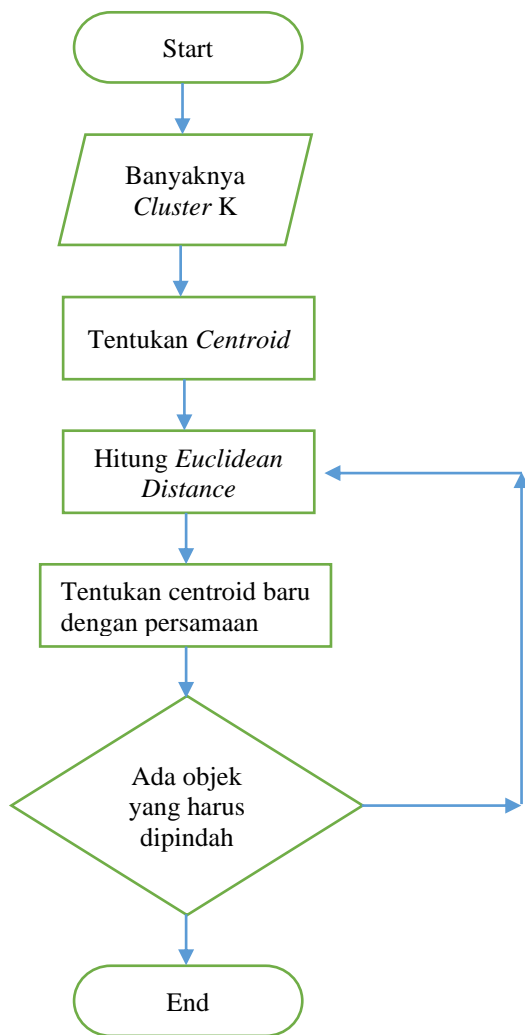
Gambar 2. Proses KDD

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 relevan.
 - b. *Data Integration* : Dilakukan terhadap atribut yang mengidentifikasi entitas yang unik.
3. *Transformation* : Merubah data sesuai format ekstention 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

Algoritma K-Means dimulai dengan pemilihan secara acak titik pusat, K disini merupakan banyaknya *Cluster* yang ingin dibentuk. Tentukan K dengan random, bisa disebut juga sebagai centroid. Hitung jarak setiap data pada masing2 centroid

menggunakan persamaan *Euclidean* hingga temukan jarak terdekat pada setiap centroid. Klasifikasikan data berdasarkan kedekatan *centroid*. Dilakukan hingga nilai centroid tidak berubah. Langkah melakukan *cluster K-means* sebagai berikut :



Gambar 2.2 Flowchart Algoritma K-Means

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

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

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:

$$v = \frac{\sum_{i=1}^n x_1}{n} : i = 1, 2, 3, \dots, n \quad (2)$$

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.

HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data yang berasal dari Laporan Pemakaian Obat. Parameter yang digunakan adalah nama obat, unit, jumlah item obat setiap bulan dari Januari hingga Desember. Proses *cluster* dilakukan untuk mengetahui pola kebutuhan obat apotik.

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.

Tabel 1. Data Obat

no	nama obat	∑ pemakaian
1	acyclovir 200 mg	6780
2	acyclovir 400 mg	540
....
232	imboost force tablet	4077
233	imboost tab	1891
234	Imodium	2093
235	Imunaga	1613.5
....
2202	zovirax cream 5g	0
2203	zyloric tab 100 mg 60's	1619

2204	zyloric tab 300 mg 100's	1165
------	-----------------------------	------

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	Jan	feb	mar	Σ pemakaian
1	acyclovir 200 mg	300	300	700	6780
2	acyclovir 400 mg	40	140	40	540
....
232	imboost force tablet	350	371	235	4077
233	imboost tab	220	170	128	1891
234	Imodium	224	186	248	2093
235	Imunaga	4	12	7	1613.5
....
2202	zovirax cream 5g	0	0	0	0
2203	zyloric tab 100 mg 60's	177	158	138	1619
2204	zyloric tab 300 mg 100's	105	95	75	1165

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 = 500 dengan atribut = 1

Tabel 3. Data Pusat Cluster Iterasi 1

no	nama obat	Σ pakai	cluster
275	licokalk	293	c1
84	candesartan 8 mg	2176	c2

383	piroxicam 20 mg	18490	c3
-----	-----------------	-------	----

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

Tabel 4. Perhitungan Euclidean Distance Iterasi 1

Nama obat	C1	C2	C3
acyclovir 200 mg (yarindo)	6487	4604	11710
acyclovir 400 mg (novell)	247	1636	17950
acyclovir 400 mg (phapros)	1527	356	16670
.....

Anggota dipilih dari terkecil diantara 3 cluster jika terkecil pada bagian C1 maka termasuk sebagai anggota C1 yaitu sebanyak 236 data, jika terkecil pada bagian C2 maka termasuk sebagai anggota C2 yaitu sebanyak 250 data, dan jika terkecil pada bagian C3 maka termasuk sebagai anggota C3 yaitu sebanyak 14 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
4. Jika 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.
5. Dan jika hasil iterasi ke-3 masih tidak sama dengan iterasi ke-2 sehingga perlu dilakukan kembali perhitungan ke iterasi ke-4 dan seterusnya sampai mendapatkan hasil yang sama.
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.

Hasil Clusterisasi data obat yang dilakukan dengan algoritma k-means didapatkan setelah melakukan iterasi ke-4 yaitu terdapat kelompok obat yang pemakaian sedikit terdapat pada cluster 1 yang memiliki 239 anggota, kelompok obat yang pemakaian sedang terdapat pada cluster 2 yang memiliki 39 anggota, dan kelompok obat yang pemakaian tinggi terdapat pada cluster 3 yang memiliki 22 anggota.

Tabel 5. Anggota Cluster 1 (Sedikit)

nama obat	Σ pemakaian
acyclovir 200 mg	120

acyclovir 400 mg	140
acyclovir cream 5g	133
akurat strip	64
alleron tab	160
allopurinol 100 mg (nova)	90
alpara tab	180
.....

Tabel 6. Anggota Cluster 2 (Sedang)

nama obat	Σ pemakaian
acyclovir 400 mg	660
alkohol swab avico pcs	700
allopurinol 300 mg	790
bodrex	802
bodrex extra	532
bodrex migra	601
.....

Tabel 7. Anggota Cluster 3 (Tinggi)

nama obat	total pemakaian
allopurinol 100 mg (hj)	910
alofar 100 mg	1100
decolsin cap 100's	964
delstrum	1117
imboost force tablet	962
itramol syr 60 ml	877
lanamol tab 500 mg	900
loratadine 10 mg (promed)	1040
neozep forte tab	990
neurobion forte	1130
.....

SIMPULAN DAN SARAN

Dari hasil penelitian clusterisasi data obat dapat disimpulkan bahwa banyak diantara obat yang ada masuk dalam kelompok cluster kurang. Kecenderungan obat yang tergolong kurang ini adalah ada beberapa bulan yang tidak ada permintaan sama sekali terhadap obat tersebut. Sedangkan kelompok obat yang termasuk tinggi rata rata permintaan obatnya setiap bulan diatas 300 buah, sedangkan obat yang masuk dalam kelompok cluster sangat tinggi rata rata permintaan setiap bulannya adalah diatas 2000 buah. Dari hasil analisa cluster diatas mungkin perlu dilakukan lagi penelitian lanjutan agar clusterisasi data obat dapat dilakukan secara lebih valid dengan menetapkan nilai centroid terbaik.

DAFTAR PUSTAKA

A. Amelio and A. Tagarelli, "Data Mining: Clustering," *Ref. Modul. Life Sci.*, 2018.

Ade Bastian, Haruan Sujadi, dan Gigin Febrianto. Penerapan Algoritma *K-Means Cluster Analysis* Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka), vol.14 no.01, 2018

F. Gullo, "From patterns in data to knowledge discovery: What data mining can do," *Phys. Procedia*, vol. 62, pp. 18–22, 2015.

Gurunesu, F. *Data Mining Concepts, Models and Techniques* (Volume 12 ed.). Verlag Berlin Heidelberg. 2011

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.

Prasetyo, Eko, *Data Mining : Mengolah Data Menjadi Informasi Menggunakan Matlab*, Andi, 2014

R. Mythily, A. Banu, and S. Raghunathan, "Clustering models for data stream mining," *Procedia Comput. Sci.*, vol. 46, no. Ict 2014, pp. 619–626, 2015.

Sulastri H, Gufroni AI. Penerapan Data Mining Dalam Pengelompokan Penderita Thalassaemia. *Jurnal Teknologi dan Sistem Informasi*. 2017 Sep 26;3(2):299-305.

Taslim, Fajrizal, "Penerapan algoritma k-mean untuk clustering data obat pada puskesmas rumbai" Vol.7, Nomor 2. 108-114, 2016.

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.

X. Wu, et al., *Top 10 Algorithms In Data Mining*, Knowl. Inf. Syst. 14 hal. 1–37. 2008

