



## Penerapan Metode *K-Means Clustering* Dalam Mengelompokkan Data Penjualan Obat pada Apotek M23

Nurul Azmi<sup>1</sup>, Hafsa HS<sup>2</sup>, Yuyun<sup>3</sup>, Hazriani<sup>4</sup>

<sup>1,2,4</sup>Sistem Komputer, Program Pasca Sarjana, Universitas Handayani Makassar

<sup>3</sup>Badan Riset dan Inovasi (BRIN)

<sup>1</sup>[nurulazmisyahril@gmail.com](mailto:nurulazmisyahril@gmail.com)

### Abstract

Planning for the need for the right medicines can make the procurement of medicines efficient and effective so that the medicines can be sufficiently available as needed and can be obtained when needed. At the current M23 Pharmacy, sometimes there is a shortage or overstock of medicines. To overcome these problems a data mining method is applied by analyzing drug use to produce information that can be used as drug inventory control and planning. The method used in this study is the *K-Means* method. The *K-Means* clustering method aims to group data that has the same characteristics into the same cluster and data that has different characteristics are grouped into other clusters. As for determining the best number of clusters using the Davies Bouldin Index (DBI) method. The results of this study determined that the best number of clusters was 2 clusters, the drug data grouping consisted of cluster 1 with low drug use consisting of 144 types of drugs and cluster 2 with high drug use consisting of 6 types of drugs.

Keywords: *K-Means*, Clustering, Davies Bouldin Index Method, Drug

### Abstrak

Perencanaan akan kebutuhan obat – obatan yang tepat dapat membuat pengadaan obat – obatan menjadi efisien dan efektif sehingga obat – obatan dapat tersedia dengan cukup sesuai kebutuhan serta dapat diperoleh pada saat diperlukan. Pada Apotek M23 saat ini, terkadang persediaan atau stok obat sering kali terjadi kekurangan atau kelebihan stok obat. Untuk mengatasi permasalahan tersebut diterapkan suatu metode data mining dengan cara menganalisa pada pemakaian obat untuk menghasilkan informasi yang dapat dijadikan sebagai pengendalian dan perencanaan persediaan obat. Metode yang digunakan pada penelitian ini yaitu metode *K-Means*. Metode *K-Means clustering* bertujuan untuk mengelompokkan data yang mempunyai karakteristik yang sama ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik berbeda dikelompokkan ke dalam *cluster* lain. Adapun untuk menentukan jumlah *cluster* terbaik dengan menggunakan metode *Davies Bouldin Index (DBI)*. Hasil penelitian ini menentukan bahwa jumlah *cluster* terbaik yaitu 2 *cluster*, pengelompokkan data obat terdiri atas *cluster* 1 dengan pemakaian obat rendah yang beranggotakan 144 jenis obat dan *cluster* 2 pemakaian obat tinggi yang beranggotakan 6 jenis obat.

Kata kunci: *K-Means*, Clustering, Metode *Davies Bouldin Index*, obat

### 1. Pendahuluan

Apotek merupakan salah satu pelayanan Kesehatan Masyarakat dalam bidang Kesehatan terutama sebagai penyedia obat – obatan. Salah satu factor penting untuk kelangsungan proses jual beli pada apotek yaitu adanya persediaan obat – obatan. Perencanaan akan kebutuhan obat -obatan dapat tersedia dengan cukup sesuai kebutuhan serta dapat diperoleh pada saat diperlukan[1]. Pada apotek M23 saat ini, terkadang persediaan atau stok obat sering kali terjadi kekurangan atau kelebihan stok obat. Sehingga untuk mengatasi permasalahan tersebut perlu diterapkan suatu metode data mining dengan cara menganalisa pada pemakaian obat untuk menghasilkan informasi yang dapat dijadikan sebagai pengendalian dan persediaan obat. Pada penelitian ini, penulis menerapkan metode *K-*

*Means* yang merupakan algoritma yang masuk kedalam algoritma *unsupervised learning* sehingga di dalam algoritma ini tidak perlu dilakukan proses *training*. Metode *K-Means clustering* bertujuan untuk mengelompokkan data yang mempunyai karakteristik yang sama ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik berbeda dikelompokkan ke dalam *cluster* lain [2].

Beberapa penelitian sebelumnya yang berkaitan dengan metode *K-Means*, seperti penelitian yang dilakukan oleh Nita Mirantika [3], menerapkan algoritma *K-Means clustering* dalam pengelompokan penyebaran *covid-19* di Provinsi Jawa Barat. Penelitian yang dilakukan oleh Ismail Virgo [4], dalam penelitiannya menggunakan algoritma *K-Means* dalam klasterisasi tingkat kehadiran Dosen. Delia Anggarwati

[5], menerapkan algoritma *K-Means* dalam prediksi penjualan karoseri.

Perbedaan penelitian sebelumnya dengan penelitian saat ini yaitu terletak pada penentuan jumlah *cluster* terbaik. Pada penelitian sebelumnya dalam menentukan jumlah *cluster* dilakukan secara random sesuai keinginan dari penulis sedangkan pada penelitian saat ini dalam menentukan jumlah *cluster* terbaik dengan menggunakan metode *Davies Bouldin Index (DBI)*.

Berdasarkan dari uraian diatas, maka penelitian yang akan dilakukan bertujuan untuk mengetahui *clusterisasi* data pemakaian obat dengan output yang dihasilkan adalah kelompok jumlah pemakaian obat. Hasil dari penelitian ini nantinya dapat digunakan pihak apotek sebagai acuan perencanaan kebutuhan obat – obatan untuk kedepannya.

## 2. Metode Penelitian

### 2.1. Objek Penelitian

Apotek M23 yang beralamat di Rukomas Odessa blok C2/1 Batam Kota. Dengan data yang diperoleh melalui staff dari Apotek M23. Adapun data yang digunakan sebanyak 150 jenis obat dengan jumlah pemakaian per bulan yaitu dari bulan Januari sampai dengan Bulan April.

### 2.2. Persiapan Data

pada proses persiapan data, langkah – langkah yang dilakukan yaitu [6],

#### 1. Selection

*Selection* digunakan untuk menentukan variabel yang akan diambil agar tidak ada kesamaan dan terjadi perulangan yang tidak diperlukan dalam proses data mining

#### 2. Processing

Pada *processing* terdapat dua tahap, yaitu: *Data Cleaning* yang Menghilangkan data yang tidak diperlukan dan *Data integration* yang Dilakukan terhadap atribut yang mengidentifikasi entitas yang unik.

#### 3. Transformasi

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.

### 2.3. Modeling

*K-Means* merupakan algoritma *clustering*. Tujuannya yaitu untuk membagi data menjadi beberapa kelompok. Algoritma ini menerima masukan yang berupa data. Pada algoritma, computer mengelompokkan data – data menjadi masukannya tanpa mengetahui terlebih dahulu target kelasnya. Masukan yang diterima yaitu data atau objek dan K kelompok (*cluster*) yang diinginkan. Algoritma ini akan mengelompokkan data atau objek ke dalam K kelompok tersebut dimana pada

setiap *cluster* terdapat titik pusat (*centroid*) yang mempresentasikan *cluster* tersebut [7].

Secara umum langkah – langkah algoritma *K-Means* Pada Ditentukan dengan jumlah K *cluster*, kemudian Pilih titik pusat secara acak, dimana titik ini akan menjadi pusat (*centroid*) dari masing – masing kelompok (*cluster*). [6]

2. Hitung jarak dan alokasikan masing – masing data ke *centroid* / rata – rata terdekat. Perhitungan nilai *centroid* dengan persamaan *Euclidean Distance* :

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

Dimana :

$D(x, y)$  adalah jarak antara titik data  $x$  dan  $y$ .

$n$  adalah jumlah dimensi (fitur) dari data.

$x_i$  dan  $y_i$  adalah nilai fitur ke- $i$  dari titik data  $x$  dan  $y$ .

3. Tentukan *centroid* baru / rata – rata dari data yang ada di masing – masing *cluster* dengan rumus:

$$m_k = \frac{1}{N_k} \sum_{x \in c_k} x \quad (2)$$

Dimana:

$m_k$  adalah pusat baru klaster ke- $k$ .

$N_k$  adalah jumlah titik data dalam klaster ke- $k$ .

$\sum_{x \in c_k} x$  adalah penjumlahan dari semua titik data dalam klaster ke- $k$ .

4. Jika *centroid* berubah, lakukan iterasi selanjutnya seperti langkah – langkah tersebut, apabila masih ada data yang berpindah *cluster* atau ada perubahan nilai *centroid*. Jika tidak ada maka hentikan proses *clustering*.

### 2.4 Davies Bouldin Index

*Davies Bouldin Indeks (DBI)* juga disebut reliabilitas klasifikasi indeks. *Davies Bouldin Index (DBI)* diperkenalkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979 didefinisikan sebagai rasio rata-rata jarak dalam dan antar-klaster untuk setiap klaster dengan klaster tetangga terdekatnya [8], Evaluasi menggunakan *DBI* ini memiliki skema evaluasi *cluster* internal, dimana baik atau tidaknya hasil *cluster* dilihat dari kuantitas dan kedekatan antar hasil *cluster* [9]. *DBI* adalah fungsi rasio dari jumlah antara *cluster scatter* sampai dengan *cluster separation* [10]. Pendekatan pengukuran *DBI* yaitu memaksimalkan jarak *intern cluster* serta meminimalkan jarak *intra cluster*. Semakin kecil nilai *DBI* menunjukkan skema *cluster* yang paling optimal [10].

## 3. Hasil dan Pembahasan

### 3.1. Penentuan Jumlah cluster

Berdasarkan metode penelitian di atas, hasil dan pembahasan diawali dari proses perhitungan K dengan menggunakan *Davies Bouldin Index (DBI)* untuk menentukan jumlah *cluster* terbaik dirujuk pada Tabel 1.

Tabel 1. Nilai *Davies Bouldin Index (DBI)*

Jumlah Cluster	Nilai <i>Davies Bouldin Index (DBI)</i>
2	0,448
3	0,508
4	0,466
5	0,510
6	0,459
7	0,518
8	0,484
9	0,450
10	0,514

Pada tabel 1. Nilai terendah dari *Davies Bouldin Index (DBI)* yaitu dengan jumlah *cluster* 2 dengan nilai *Davies Bouldin Index (DBI)* 0,448 sehingga jumlah 2 *cluster* dapat menjadi *cluster* optimal.

### 3.2. Proses perhitungan *K-Means*

Dalam proses perhitungan *K-Means* setelah mengetahui jumlah *cluster* terbaik dirujuk pada Tabel 2. Penentuan *centroid* secara acak (random)

Tabel 2. Penentuan *centroid* awal

Cluster	Jan	Feb	Mar	Apr
C1	10	10	10	10
C2	50	50	50	50

Pada tabel 2. Menunjukkan penentuan titik pusat pada masing – masing parameter secara acak.

Menghitung jarak tiap data dengan masing- masing *cluster* pusat dengan menggunakan persamaan *Euclidean Distance*:

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

$$C1 (1) = \sqrt{((10-10)^2+(70-10)^2+(61-10)^2+(96-10)^2)} = 133$$

$$C1 (2) = \sqrt{((50-10)^2+(10-10)^2+(15-10)^2+(25-10)^2)} = 43$$

$$C1 (3) = \sqrt{((1-10)^2+(1-10)^2+(1-10)^2+(1-10)^2)} = 18$$

$$C1 (4) = \sqrt{((10-10)^2+(12-10)^2+(3-10)^2+(5-10)^2)} = 9$$

$$C1 (5) = \sqrt{((11-10)^2+(5-10)^2+(20-10)^2+(14-10)^2)} = 12$$

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

$$C2 (1) = \sqrt{((10-50)^2+(70-50)^2+(61-50)^2+(96-50)^2)} = 56$$

$$C2 (2) = \sqrt{((50-50)^2+(10-50)^2+(15-50)^2+(25-50)^2)} = 59$$

$$C2 (3) = \sqrt{((1-50)^2+(1-50)^2+(1-50)^2+(1-50)^2)} = 98$$

$$C2 (4) = \sqrt{((10-50)^2+(12-50)^2+(3-50)^2+(5-50)^2)} = 85$$

$$C2 (5) = \sqrt{((11-50)^2+(5-50)^2+(20-50)^2+(14-50)^2)} = 76$$

Anggota dipilih dari terkecil diantara 2 *cluster*, jika terkecil berada pada bagian C1 maka termasuk anggota C1 dimana pada hasil perhitungan diatas anggota dari C1 sebanyak 143 data, dan jika terkecil berada pada bagian C2 maka termasuk sebagai anggota dari C2 yaitu sebanyak 7 data, data ini dirujuk pada Tabel 3.

Tabel 3. Perhitungan *Euclidean Distance* iterasi 1

No.	NAMA	J	F	M	A	C1	C2	Clus
1	DEPO HARMONI S INJ 3BULAN 3ML LAPI	73	70	61	96	133	56	2
2	LAMESON 16MG 10X10S PFIZER	50	10	15	25	43	59	1
3	VIAGRA 100MG 1X4S MENARINI	1	1	1	1	18	98	1
4	ARCOXIA 120MG 3X10 TENSIVAS	10	12	3	5	9	85	1
5	K 10MG 5X10 TAB	11	5	20	14	12	76	1
-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-
149	YARINDO FUROSEMI DE 40 MG RB	50	45	50	50	78	5	2
150	DETTOL ANTISEPTI K 245ML	1	1	1	1	18	98	1

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

Lakukan Iterasi, Setelah penentuan *centroid* baru selanjutnya menghitung jarak tiap data dengan masing – masing *cluster* pusat dengan menggunakan *Euclidean Distance*. proses pencarian *centroid* baru dan iterasi akan terus berulang sampai sudah tidak ada lagi perubahan dalam jumlah elemen setiap *cluster*nya.

Pada penelitian dapat kita lihat di Tabel 4 bahwa proses iterasi ini terhenti pada iterasi ke-3 karena hasil dari iterasi sebelumnya sudah sama sehingga tidak perlu lagi melanjutkan ke iterasi ke-4 atau cukup berhenti di iterasi ke-3. Hasil iterasi terakhir akan digunakan untuk parameter *clustering*.

Tabel 4. Perhitungan *Euclidean Distance* iterasi 3

No.	NAMA	J	F	M	A	C1	C2	Clust
1	DEPO HARMONIS INJ 3BULAN 3ML	73	70	61	96	136	22	2
2	LAPI LAMESON 16MG 10X10S	50	10	15	25	44	90	1
3	PFIZER VIAGRA 100MG 1X4S	1	1	1	1	15	133	1
4	MENARINI ARCOXIA 120MG 3X10	10	12	3	5	8	120	1
5	TENSIVASK 10MG 5X10 TAB	11	5	20	14	14	110	1
-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-
149	YARINDO FUROSEMIDE 40 MG	50	45	50	50	81	38	2
150	RB DETTOL ANTISEPTIK 245ML	1	1	1	1	15	133	1

### 3.3. Hasil Perhitungan *K-Means*

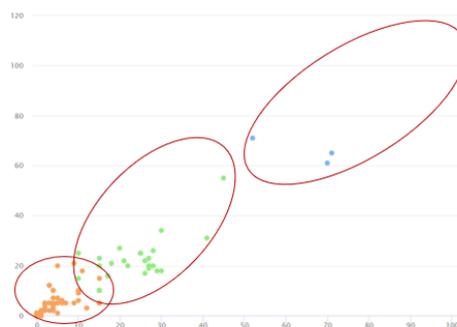
Hasil yang diperoleh dari proses *clustering* kemudian divisualisasikan dalam bentuk diagram *scatter* dapat dilihat dari Gambar 1, Dari data tersebut dapat dilihat bahwa *cluster* 1 merupakan *cluster* obat yang pemakaiannya rendah dengan memiliki 144 anggota dan kelompok pemakaian obat tinggi pada *cluster* 2 yang memiliki 6 anggota.



Gambar 1. Visualisasi data hasil *clustering*

### 3.4. Perbandingan Visualisasi jumlah *cluster*

Untuk menguji ketepatan metode *Davies Bouldin Index (DBI)* dalam menentukan jumlah *cluster* yang tepat dilakukan perbandingan dapat dilihat pada Gambar 2, pada gambar ini dijelaskan bahwa Dengan jumlah *cluster* 3 Semua data obat berada pada kelas yang sama berdasarkan *cluster* yang sama namun tidak *homogeny* dalam satu *cluster*, sedangkan pada gambar 1 dengan jumlah *cluster* 2 pengelompokan data obat terkumpul dalam kelas yang sama pada satu *cluster*.



Gambar 2. Visualisasi data hasil *clustering (K=3)*

## 4. Kesimpulan

Metode *Davies Bouldin Index (DBI)* dapat digunakan untuk mencari optimasi dalam penentuan jumlah *cluster* yang akan diterapkan dalam algoritma *clustering* seperti *K-Means*.

Pada proses perhitungan yang telah dilakukan tentang penerapan metode *K-Means* untuk *clustering* data obat, dapat diambil kesimpulan bahwa penerapan metode *Davies Bouldin Index (DBI)* sesuai dengan penelitian ini menghasilkan jumlah *cluster* terbaik adalah 2 *cluster* dan dengan menerapkan metode *K-Means* dihasilkan bahwa jumlah pemakaian obat sebanyak 150 jenis obat, dengan proses perhitungan sebanyak 3 iterasi dihasilkan ada 2 *cluster* yaitu pemakaian obat rendah dan pemakaian obat tinggi. Hasil *cluster* pemakaian obat rendah sebanyak 144 jenis obat dengan rentan jumlah pemakaian antara 1 – 40, sedangkan hasil *cluster* pemakaian obat tinggi sebanyak 6 jenis obat dengan rentan pemakaian antara 40-105.

Hasil tersebut dapat salah satu pertimbangan atau *knowledge base* dalam pengendalian dan persediaan obat – obatan pada Apotek M23 untuk bulan-bulan berikutnya.

### Daftar Rujukan

- [1] N. A. O. Saputri and M. Elvirasari, “Penerapan Metode K-Means Clustering Untuk Menentukan Jumlah Penjualan Obat Yang Banyak Terjual Pada Apotek Murbay Sekayu,” *Informanika J.*, vol. 7, no. 2, pp. 44–51, 2021.
- [2] M. W. Talakua, Z. A. Leleury, and A. W. Taluta, “Analisis Cluster Dengan Menggunakan Metode K-Means Untuk Pengelompokan Kabupaten/Kota Di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014,” *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 11, no. 2, pp. 119–128, 2017, doi: 10.30598/barekengvol11iss2pp119-128.
- [3] N. Mirantika, “Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Covid-19 di Provinsi Jawa Barat,” *Nuansa Inform.*, vol. 15, no. 2, pp. 92–98, 2021, doi: 10.25134/nuansa.v15i2.4321.
- [4] I. Virgo, S. Defit, and Y. Yuhandri, “Klasterisasi Tingkat Kehadiran Dosen Menggunakan Algoritma K-Means Clustering,” *J. Sistim Inf. dan Teknol.*, vol. 2, pp. 23–28, 2020, doi: 10.37034/jsisfotek.v2i1.17.
- [5] D. Anggarwati, O. Nurdiawan, I. Ali, and D. A. Kurnia, “Penerapan Algoritma K-Means Dalam Prediksi Penjualan Karoseri,” *J. Data Sci. Inform.*, vol. 1, no. 2, pp. 58–62, 2021.
- [6] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, “Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.
- [7] P. M. C. Abrianto, “Penerapan Metode K-Means Clustering Untuk Pengelompokan Pasien Penyakit Liver,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 2, no. 2, pp. 247–255, 2018.
- [8] Y. Sopyan, A. D. Lesmana, and C. Juliane, “Analisis Algoritma K-Means dan Davies Bouldin Index dalam Mencari Cluster Terbaik Kasus Perceraian di Kabupaten Kuningan,” *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1464–1470, 2022, doi: 10.47065/bits.v4i3.2697.
- [9] D. Jollyta, S. Efendi, M. Zarlis, and H. Mawengkang, “Optimasi Cluster Pada Data Stunting: Teknik Evaluasi Cluster Sum of Square Error dan Davies Bouldin Index,” *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 918, 2019, doi: 10.30645/senaris.v1i0.100.
- [10] Widiarina, “Algoritma Cluster Dinamik Untuk Optimasi Cluster Pada Algoritma K-Means Dalam Pemetaan Nasabah Potensial Algoritma Cluster Dinamik Untuk Optimasi Cluster Pada Algoritma K-Means Dalam,” *Tesis Magister Ilmu Komputer, Nusa Mandiri*, vol. 1, no. 1, pp. 33–36, 2013.