

# KLASTERISASI DATA PADA TOKO OBAT MENGUNAKAN METODE *K-MEANS* UNTUK ANALISA KETERSEDIAAN BARANG

## DATA CLUSTERING IN DRUG STORE USING THE K-MEANS METHOD FOR ANALYSIS OF STOCK AVAILABILITY

Yusuf<sup>1)</sup>, Lukman Junaedi<sup>2)</sup>,

E-mail : <sup>1)</sup>[yusuf.16@fik.narotama.ac.id](mailto:yusuf.16@fik.narotama.ac.id) , <sup>2)</sup>[lukman.junaedi@narotama.ac.id](mailto:lukman.junaedi@narotama.ac.id)

<sup>1)</sup>Program Studi Teknik Informatika, Universitas Narotama Surabaya

<sup>2)</sup> Program Studi Sistem Informasi, Universitas Narotama Surabaya

### Abstrak

Toko obat setiap harinya menyediakan ketersediaan obat namun banyaknya pembelian obat setiap harinya menyebabkan toko obat sering kali kehabisan stok. Masalah ini disebabkan karena kurangnya pengecekan terhadap data yang berjumlah besar. Penelitian ini digunakan untuk proses klasterisasi data obat yang cepat laku dan kurang laku. Dengan menggunakan metode K-Means, salah satu metode klasterisasi tidak terpandu yang mampu mengelompokkan data berdasarkan pola tertentu. Proses klasterisasi berhasil mendeteksi bahwa terdapat 69 jenis obat yang tidak ada permintaan sama sekali, dan 35 obat yang cepat laku. Selain itu penelitian ini memerlukan penelitian lanjutan untuk mendapatkan analisa hasil klaster dengan nilai centroid yang lebih baik.

**Kata kunci:** *k-means, klasterisasi, obat, persediaan*

### Abstract

Drug stores provide the availability of drugs every day, but the number of purchases of drugs every day causes drug stores to often run out of stock. This problem is caused by the lack of checking of large amounts of data. This research is used for the clustering process of drug data which sells quickly and does not sell well. By using the K-Means method, an unsupervised clustering method that is able to group data based on a certain pattern. The clustering process was successful in detecting that there were 69 types of drugs that were not in demand at all, and 35 drugs were selling quickly. In addition, this research requires further research to obtain a cluster result analysis with a better centroid value.

**Keywords:** *k-means, clusterization, medicine, inventory*

## 1. PENDAHULUAN

Toko obat merupakan salah satu toko yang setiap harinya menyediakan kebutuhan obat yang mampu memenuhi keluhan kesehatan masyarakat. Akan tetapi semakin banyak masyarakat ingin membeli kebutuhan obat sehingga menyebabkan terjadinya kekosongan stok obat. Masalah ini timbul karena sulitnya menghitung data yang besar dan kurangnya pengecekan terhadap data yang ada [1] dan kesalahan dalam menghitung

jumlah ketersediaan. Dengan demikian hal ini perlu dibutuhkan perencanaan kebutuhan obat yang nantinya permintaan masyarakat dapat terpenuhi dengan baik, mudah dan memberikan jarak tempuh pembelian yang dekat bagi masyarakat. Untuk menyelesaikan masalah tersebut, pihak apotek harus membuat suatu perencanaan kebutuhan obat yang tepat untuk mengoptimalkan kebutuhan obat untuk kebutuhan masyarakat. Maka dibutuhkan sistem dan metode yang mampu mengelompokkan barang berdasarkan merk atau model tertentu sehingga memudahkan saat pembuatan laporan, pengadaan barang dan bisa lebih efisiensi. Untuk menghindari resiko pengambilan stok barang yang kurang laku terambil kembali maka dibuatlah penelitian ini yang dapat berpengaruh dalam mengetahui tinggi rendahnya angka penjualan di toko obat [2].

Dengan demikian pengelompokan sesuai kebutuhan berdasarkan data distribusi obat obat pada bulan-bulan sebelumnya dapat menjadi acuan perencanaan obat untuk bulan berikutnya. Sehingga diharapkan ketersediaan obat pada apotek dapat memenuhi kebutuhan permintaan masyarakat atau instansi lainnya dengan mudah. Dengan memperhatikan beberapa hal di atas, maka peneliti tertarik untuk melakukan penelitian dengan judul “*Clustering Data Penjualan Untuk Ketersediaan Barang Pada Toko Obat Menggunakan Metode K-means*”. Klasterisasi data merupakan salah satu metode dalam data mining yang dapat digunakan untuk memetakan data kedalam kelompok yang lebih kecil berdasarkan kesamaan karakteristik yang dimilikinya [3]. Dengan begitu klasterisasi obat dapat diharapkan dapat menjadi salah satu sumber pengambilan keputusan pada ketersediaan obat pada apotek.

## 2. METODOLOGI

Metodologi yang digunakan dalam penelitian ini menggunakan beberapa prosedur diantaranya *data collection*, *cleaning data*, klasterisasi dan analisa data.

### 2.1 Data Collection

Populasi dan sampel adalah salah satu bagian penting dalam penelitian yang harus ditentukan sejak awal.[4] pengumpulan data pada penelitian ini menggunakan data yang berasal dari apotik yang didapatkan pada transaksi penjualan pada Kota Surabaya. Metode pengumpulan data yang digunakan dalam penelitian ini adalah observasi dan wawancara [5]. Data tersebut berektensi *.csv* atau *excel* yang terdapat 4 atribut sebagai penunjang yakni nama obat, satuan obat, dan juga penjualan perhari. Pengambilan data dilakukan dengan metode observasi. Observasi merupakan aktivitas dari suatu proses atau objek dengan maksud merasakan kemudian memahami pengetahuan dari sebuah fenomena[6]. Format data *csv* ini dapat diolah dengan berbagai aplikasi *data mining* dengan mudah [7]. Sebagai acuan data yang digunakan pada penelitian diambil sampel data dari 2 bulan.:

### 2.2 Cleaning Data

Data cleansing atau yang disebut juga dengan *data scrubbing* merupakan suatu proses analisa mengenai kualitas dari data dengan mengubah.[8] Data Cleansing adalah proses mendeteksi dan memperbaiki (atau menghapus) data yang rusak atau tidak akurat dengan cara mengatur kembali data-data yang ada pada catatan, tabel, atau basis data [9]. Pada tahapan kali ini data obat akan melalui beberapa proses operasi

*cleaning* data yang akan mencakup beberapa hal yaitu identifikasi entry tanpa data dan entry berupa data yang telah hilang. proses tersebut akan sangat membantu untuk mendapatkan akurasi, *decleansing*, dan validitas suatu data. Contoh data yang belum mengalami proses *cleaning* data adalah pada tabel berikut

**Tabel 1. Tabel data yang belum mengalami proses cleaning data**

No	Nama obat	Satuan	2020				
			28	29	31	1	3
1	Acyclovir	Tablet	0		10		
2	Alcohol	Botol	2		2		1
3	Ambeven	tablet	10	10	20		
4	amlodipin	tablet	10	20	30	10	50
6	amoxsilin	Set	30	11	21	40	50
7	amplodipine	tablet	20				
8	anakonidin	tablet	1	3			
9	antangan cair	Set	33	1	23	7	2
10	Apialys	tablet	0			1	
11	Asepso	Pot	2	2			1
12	b.geliga	Botol	2		1		1
13	Berloson	tablet	1	2			2
14	Bodrex	Set	3	12	7	2	23
15	Bratidin	tablet	11	1			
16	Braxidin	tablet	2	25			
17	bye anak	tablet	5	1	1		3
18	Caladin	tablet	1	1			
19	Captopil	tablet	3	20		2	

Dapat dilihat pada tabel 1 masih banyak kolom yang kosong sehingga proses klasterisasi tidak dapat berjalan, oleh karena itu *cleaning* data bertujuan agar kolom tersebut dapat terisi.

### 2.3 Klasterisasi

Klasterisasi adalah sebuah proses untuk mengelompokkan data ke dalam beberapa *cluster* atau kelompok sehingga data dalam satu *cluster* memiliki tingkat kemiripan yang maksimum dan data antar *cluster* memiliki kemiripan yang minimum.[10] tujuan utama analisis cluster adalah mengelompokkan objek-objek berdasarkan kesamaan karakteristik di antara objek-objek tersebut[11]. Proses klasterisasi dilakukan dengan menggunakan bahasa pemrograman *Python*, dan menggunakan beberapa modul diantaranya *seaborn*, *numpy*, *pandas*, dan *matplotlib* [12]. Selanjutnya dilakukan analisa variasi masing-masing variabel dan visualisasi untuk mengetahui pola awal data sebelum proses klasterisasi. Variasi masing-masing variable

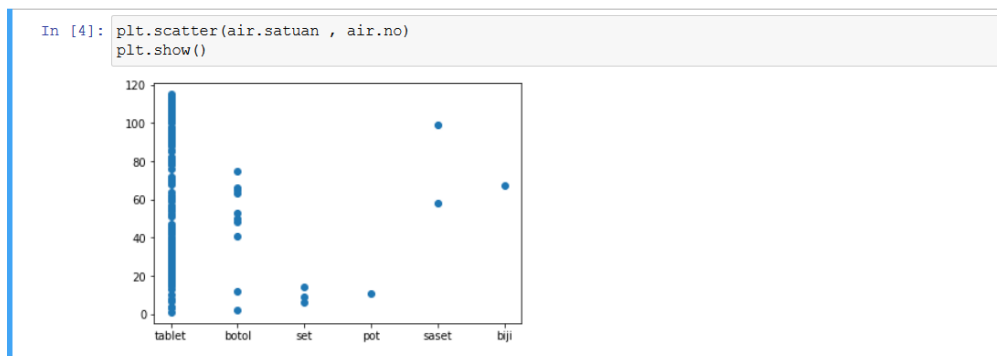
```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 107 entries, 0 to 106
Data columns (total 22 columns):
 #   Column      Non-Null Count  Dtype
---  -
 0   no          107 non-null   int64
 1   nama obat  107 non-null   object
 2   satuan      107 non-null   object
 3   28          107 non-null   int64
 4   29          107 non-null   int64
 5   31          107 non-null   int64
 6   1           107 non-null   int64
 7   3           107 non-null   int64
 8   4           107 non-null   int64
 9   5           107 non-null   int64
10   6           107 non-null   int64
11   7           107 non-null   int64
12   8           107 non-null   int64
13   10          107 non-null   int64
14   11          107 non-null   int64
15   12          107 non-null   int64
16   13          107 non-null   int64
17   14          107 non-null   int64

```

**Gambar 1. Analisa variasi data pada masing-masing variabel**

Langkah selanjutnya adalah visualisasi awal pola data untuk mengetahui pusat kluster dan langkah untuk menentukan variabel yang akan dikelompokkan. Visualisasi awal dilakukan dengan menggunakan perintah yang ada pada modul *matplotlib*. [13]



**Gambar 2. Visualisasi awal untuk mengetahui pola data**

Visualisasi data dipakai untuk mempresentasikan data yang terstruktur ataupun tidak dengan grafik atau bagan untuk menampilkan info yang tersembunyi didalam data [14]. Setelah mengetahui pola awal data maka dapat ditentukan variabel untuk klasterisasi yaitu nilai skala peubah, nilai *random state* dan juga jumlah kluster yang masing masing bernilai 1, 5, dan 3. Scatter plot sebuah grafik yang biasa digunakan untuk melihat suatu pola hubungan antara 2 variabel [15]

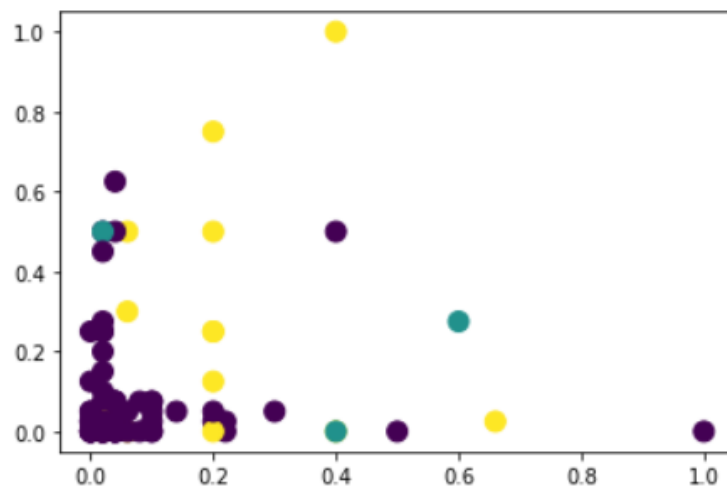
### 3. HASIL DAN PEMBAHASAN

```
In [10]: print(kmeans.cluster_centers_)
[[0.06222222 0.06472222 0.04777778 0.02472222 0.04252137 0.02049383
 0.02703704 0.03511111 0.02111111 0.01633987 0.02866667 0.00911111
 0.01822222 0.03472222 0.01333333 0.0237037 0.03259259 0.02611111
 0.01972222]
 [0.34 0.25833333 0.7 0.90833333 0.51282051 0.61481481
 0.66666667 0.86666667 0.83333333 0.60130719 0.66666667 0.14
 0.4 0.25 0.51666667 0.77777778 0.66666667 0.5
 0.5 ]
 [0.20571429 0.26607143 0.32142857 0.17857143 0.25 0.25555556
 0.28095238 0.24 0.125 0.27731092 0.21142857 0.16714286
 0.32857143 0.22678571 0.43571429 0.19285714 0.28571429 0.40178571
 0.32678571]]
```

Gambar 3. Hasil kalterisasi

Dalam hasil klasterisasi data obat bahwa dapat disimpulkan masih banyak di antara obat yang dalam kelompok tergolong kurang laku. Kecenderungan data obat yang kurang laku 69 obat tersebut adalah beberapa hari tidak ada permintaan sama sekali pada obat tersebut, sedangkan obat yang termasuk laku rata 35 buah. Dengan diketahuinya produk pada setiap *cluster*, maka dapat menyusun strategi penjualan untuk persediaan stok barang dan strategi pemasaran untuk memasarkan produk yang paling cepat laku.

0	1	acyclovir	tablet	0	0	10	0	0	2	0	...	0	0	1	0	0	1	0	0	0	
1	2	alkohol	botol	2	0	2	1	0	1	1	...	1	1	1	1	1	0	1	0	0	2
2	3	ambeven	tablet	10	10	20	0	0	10	10	...	10	0	20	2	0	10	0	0	10	0
3	4	amlodipin	tablet	10	20	30	10	50	20	10	...	10	0	20	0	0	10	20	0	10	10
4	6	amoxsilin	set	30	11	21	40	50	43	20	...	51	20	20	20	10	20	30	10	20	40



Gambar 3. Hasil pola gambar klaterisasi

#### 4. KESIMPULAN DAN SARAN

Produk yang paling banyak terjual dapat diketahui dengan menggunakan *K-Means*, dengan mengelompokan data penjualan, jumlah transaksi, dan rata-rata. Untuk rata-rata diambil dari data jumlah transaksi selama 2 bulan, Dari hasil klasterisasi data obat bahwa dapat disimpulkan masih banyak di antara obat yang dalam kelompok tergolong kurang laku. Kecendrungan data obat yang kurang laku 69 obat tersebut adalah beberapa hari tiadak ada permintaan sama sekali pada obat tersebut, sedangkan obat yang termasuk laku rata 35 buah dari hasil analisa cluster diatas mungkin perlu dilakukan lagi penelitian lanjutan agar clustering data obat dapat dilakukan secara lebih valid dengan nilai centroid yang terbaik.

#### 5. DAFTAR RUJUKAN

- [1] U. I. Arsyah, "Prediksi Ketersediaan Obat dengan Jaringan Syaraf Tiruan Backpropagation (Studi Kasus : PT. Kalbe Farma Medan)," p. 7.
- [2] M. H. Adiya and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru," vol. 05, no. 01, p. 8, 2019.
- [3] T. Taslim and F. Fajrizal, "Penerapan algoritma k-mean untuk clustering data obat pada puskesmas rumbai," *Digit. Zone J. Teknol. Inf. Dan Komun.*, vol. 7, no. 2, pp. 108–114, Nov. 2016, doi: 10.31849/digitalzone.v7i2.602.
- [4] Yuvalianda, "Membedakan Populasi dan Sampel Dalam Penelitian," Mar. 27, 2020. <https://yuvalianda.com/populasi-dan-sampel/>.
- [5] Y. Darmi and A. Setiawan, "Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk," vol. 12, no. 2, p. 10, 2016.
- [6] romadecade, "Pengertian Observasi," 2020. <https://www.romadecade.org/pengertian-observasi/#!>
- [7] Anggri Yulio P, "Manipulasi file csv dengan Python," 2017. <https://devtrik.com/python/manipulasi-file-csv-dengan-python/>.
- [8] Bivisyani Questibrilia, "Data Cleansing: Pengertian dan Tahapannya," Feb. 10, 2020. <https://www.jojonomic.com/blog/data-cleansing/>.
- [9] Gamal Hadi, "Apa yang dimaksud dengan Data Cleansing (Data Scrubbing)," 2018. <https://www.dictio.id/t/apa-yang-dimaksud-dengan-data-cleansing-data-scrubbing/15064>.
- [10] Edy Irwansyah, S.T.,M.Si., "CLUSTERING," 2006. <https://socs.binus.ac.id/2017/03/09/clustering/>.
- [11] Anwar Hidayat, "Penjelasan Lengkap Tentang Analisis Cluster," 2014. <https://www.statistikian.com/2014/03/analisis-cluster.html>.
- [12] Yasir Abdur Rohman, "Pengenalan NumPy, Pandas, Matplotlib," 2019. <https://medium.com/@yasirabd/pengenalan-numpy-pandas-matplotlib-b90bafd36c0>.

- [13] Abder-Rahman Ali, “Memperkenalkan Matplotlib Library Pada Python,” Sep. 14, 2016. <https://code.tutsplus.com/id/tutorials/introducing-matplotlib--cms-26543>.
- [14] finereport\_en, “Apa Itu Visualisasi Data? Definisi, Pentingnya, Jenis, dan Software,” 2020. <https://www.finereport.com/en/data-visualization/visualisasi-data.html>.
- [15] Prasetyowijaya, “Penjelasan tentang Scatter plot,” <https://www.prasetyowijaya.com/>, Feb. 08, 2009. <https://www.prasetyowijaya.com/2009/02/scatter-plot.html>.

*Halaman ini sengaja dikosongkan*