



Clustering Obat Untuk Menentukan Pola Pemasaran Efektif di Apotek Amarta Sehat

Bagus Aji Pangestu¹, Nikko Aldiana Kristiawan², Nina Sulistiyowati³

^{1,2,3}Universitas Singaperbangsa Karawang

Received: 8 Agustus 2022
Revised: 13 Agustus 2022
Accepted: 16 Agustus 2022

Abstract

Medicine is a substance that is used to diagnose, eliminate, and cure diseases, injuries, or others in humans. Handling and prevention of various diseases cannot be separated from therapeutic actions with drugs. Drug grouping serves to classify drugs into several groups to determine the characteristics of a drug or not. By knowing the characteristics of each existing drug, it can be easier to determine an effective marketing pattern. The use of data mining can help to cluster drugs by utilizing existing sales data. In this study the methodology used is CRISP-DM with the stages carried out namely Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. The dataset used is Amarta Sehat Pharmacy data from January-December 2021. The K-Means algorithm is used for cluster formation using Jupyter Notebook tools with the python programming language. The elbow method serves to determine the best number of clusters (K), the recommendation from the elbow method produces the 5 most optimal clusters and is also calculated by evaluating the Sum of square error with an optimal cluster value of 7154215036292.542. The results of drug Clustering obtained to determine an effective marketing pattern at the Amarta Sehat Pharmacy are 11 drugs that are classified as high-selling drugs, 76 drugs are classified as best-selling drugs, 131 drugs are classified as drugs with the category of selling well. quite in demand, 399 drugs into the category of drugs that are not in demand, and 326 drugs into the category of drugs that are not in demand.

Keywords: Data mining, K-Means, CRISP-DM, Metode Elbow

(*) Corresponding Author: Bagusajipangestu@gmail.com

How to Cite: Pangestu, B., Kristiawan, N., & Sulistiyowati, N. (2022). Clustering Obat Untuk Menentukan Pola Pemasaran Efektif di Apotek Amarta Sehat. *Jurnal Ilmiah Wahana Pendidikan*, 8(16), 115-126. <https://doi.org/10.5281/zenodo.7058995>.

INTRODUCTION

Perkembangan sistem informasi yang semakin maju dapat kita rasakan dari berbagai bidang kehidupan, misalnya bidang pendidikan, hiburan, kesehatan, serta bisnis yang semuanya menuntut untuk menggunakan sistem informasi (Fakhriza & Umam, 2021). Perkembangan sistem informasi yang semakin maju diikuti juga oleh persaingan dunia bisnis yang semakin pesat dan ketat membuat tingkat persaingan bisnis semakin terbuka dalam memenuhi tuntutan kebutuhan pelanggan yang juga semakin tinggi (Aziz et al., 2020). Setiap pebisnis berlomba untuk menawarkan berbagai macam keunggulan dan harga produk yang terjangkau oleh pasar dengan tujuan meningkatkan penjualan. Untuk itu, pebisnis harus memiliki

kemampuan analisis pasar yang baik untuk menciptakan strategi bisnisnya dengan analisis data.

Salah satu contoh bisnis yang sedang berkembang saat ini adalah Apotek, kebutuhan obat di era pandemi covid-19 sangat tinggi, seperti permintaan vitamin, suplemen, dan obat peningkat kekebalan tubuh mengalami kenaikan yang signifikan. Sebelum pemerintah membagikan vaksin untuk masyarakat, obat-obatan dipilih untuk menjaga kesehatan selama pandemi. Hal itu berdampak pada penjualan di Apotek Amarta Sehat yang meningkat tinggi. Setelah angka kasus covid-19 semakin melandai dan jumlah orang yang terpapar sudah menurun, penjualan di Apotek Amarta Sehat juga mulai kembali seperti biasanya atau cenderung menurun. Hal ini mungkin berdampak karena jumlah Apotek yang semakin bertambah banyak, untuk itu pebisnis apotek harus menentukan strategi pemasaran yang baik untuk bisa meningkatkan penjualan, untuk menentukan suatu pola pemasaran bisa memanfaatkan data penjualan setiap harinya. Saat ini Apotek Amarta Sehat masih memasukan data penjualan secara manual melalui buku dan setiap bulannya akan di pindah pada *Microsoft Excel*, data tersebut hanya dibiarkan begitu saja dan hanya menjadi arsip penjualan tidak dilakukan analisa untuk mendapatkan informasi yang bermanfaat untuk meningkatkan penjualan.

Dengan permasalahan yang sudah ada tentu bisa kita pecahkan dengan teknik *Data Mining*. *Data Mining* merupakan proses penambangan data dengan memanfaatkan sekumpulan data yang berukuran sangat besar melalui berbagai tahapan untuk mendapatkan berbagai informasi yang bermanfaat (Cobit & Utami, 2019). Dari data penjualan setiap hari di Apotek Amarta Sehat akan dilakukan sebuah analisa untuk mendapatkan informasi serta dapat meningkatkan penjualan. Salah satu metode dalam *data mining* untuk menangani permasalahan yang digunakan pada penelitian ini adalah *Clustering*. *Clustering* dipilih untuk menemukan jumlah *cluster* pada data penjualan. *Clustering* menentukan terlebih dahulu nilai *cluster* yang diinginkan, *Clustering* merupakan metode pengelompokan data yang dimulai dengan mengelompokan dua atau lebih objek yang memiliki kesamaan. Selanjutnya proses diteruskan ke objek lain yang memiliki kedekatan keduanya (Bramasta & Halilintar, 2021).

K-Means merupakan salah satu algoritma yang sering digunakan pada proses *Clustering*. Dengan menggunakan algoritma ini, data yang telah didapat bisa dikelompokkan kedalam beberapa *cluster* berdasarkan kemiripan dari data-data tersebut, sehingga data-data yang memiliki karakteristik yang berbeda dapat dikelompokkan dalam *cluster* yang lain yang memiliki karakteristik yang sama (Siregar, 2018). *K-Means* mempunyai kelemahan dalam menentukan jumlah nilai K (*cluster*) yang dilakukan secara acak mengakibatkan validasi *cluster* yang didapatkan tidak maksimal, untuk hal ini kita bisa menggunakan metode elbow untuk solusinya. Metode elbow mampu merekomendasikan nilai kluster yang paling optimal dan dengan bantuan evaluasi *Sum of Square Error*.

Ada beberapa penelitian terdahulu yang bisa menjadi referensi dalam penelitian ini. Penelitian yang dilakukan oleh Muhammad Hilman Fakhriza, Khaerul Umam (2021) melakukan Implementasi *K-Means* Untuk Pengelompokan Produk Terbaik PT. Koko Pelli. Adapun algoritma yang digunakan untuk pengelompokan produk terbaiknya menggunakan *K-Means*.

Penelitian *K-Means* lainnya juga pernah dilakukan oleh M. Syukron Nawawi, Falentino Sembiring, Adhithia Erfina (2021) tentang Implementasi Algoritma *K-Means Clustering* Menggunakan Orange Untuk Penentuan Produk Busana Muslim Terlaris. Penelitian ini menggunakan visualisasi Orange.

Muhammad Dahria, Rudi Gunawan, Zulkifli Lubidz (2019) melakukan Analisis Produk terlaris menggunakan metode *K-Means Clustering* pada PT. Sukanda Jaya. Hasil akhirnya yaitu membuat sistem berupa aplikasi untuk menentukan produk terbaik di PT. Sukanda Jaya

Penelitian lain yang serupa juga pernah dilakukan dengan menggunakan Algoritma C45 oleh Ahmad Zakir, Yermia Ndruru, Erdian Hadinata (2020) tentang Penerapan *Data Mining* untuk klasifikasi data penjualan makanan terlaris. Selanjutnya menggunakan Algoritma Naïve Bayes telah dilakukan juga oleh Herry Derajad Wijaya, Saruni Dwiasnati (2020) yang melakukan implementasi *Data Mining* dengan algoritma Naïve bayes pada penjualan obat. Penelitian ini menggunakan tools *Rapidminer* dengan nilai *accuracy* 88%.

METODOLOGI PENELITIAN

Cross Industry Standard Process for Data Mining (CRISP-DM) adalah metode yang sering digunakan dalam *data mining*. CRISP-DM terbagi menjadi 6 tahapan yaitu :

1. *Business Understanding*

Pada tahap ini bertujuan untuk memahami situasi dan permasalahan yang ada pada objek untuk menentukan tujuan.

2. *Data Understanding*

Pada tahap ini bertujuan untuk melakukan pengumpulan dan eksplorasi data untuk melihat kualitas data. Eksplorasi data dilakukan dengan pengecekan *missing value* dan *duplicate data* pada data yang akan digunakan.

3. *Data Preparation*

Pada tahap ini dilakukan pembersihan data dari *duplicate data* pada data yang akan digunakan. Setelah itu dilakukan *data selection* yaitu menghapus atribut yang tidak diperlukan untuk tahap pembuatan model *Clustering*.

4. *Modeling*

Pada tahap ini dilakukan pemodelan terhadap data yang ada. Pembuatan model dilakukan menggunakan algoritma *K-Means*. Model dibuat dengan 2 skenario. Skenario 1 dengan mencoba nilai K secara acak yaitu sebanyak 3 cluster dan skenario 2 mencoba nilai K sebanyak 5 *cluster* sesuai rekomendasi metode Elbow.

5. *Evaluation*

Pada tahap ini dilakukan evaluasi dari setiap model skenario yang telah dibuat menggunakan *Sum of Square Error* untuk dapat mengetahui nilai cluster yang paling optimal, semakin kecil nilai SSE semakin optimal *cluster* yang direkomendasikan.

6. *Deployment*

Pada tahap ini hasil *cluster* akan dideskripsikan untuk pelaporan kepada *owner* apotek untuk informasi obat apa saja yang masuk ke dalam kategori laris, dll. yang nantinya bisa menjadi strategi penjualan yang efektif.

KAJIAN TEORI

Data Mining

Data Mining merupakan sebuah proses penggalian atau penyaringan data dengan memanfaatkan kumpulan data yang berukuran besar melalui berbagai tahapan untuk mendapatkan informasi yang bermanfaat dari data tersebut. *Data mining* dalam prosesnya menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengidentifikasi informasi yang bermanfaat yang terkait dengan database yang besar (Handoko, 2016). Dari pengertian yang sudah dijelaskan, dapat ditarik kesimpulan bahwa *Data mining* adalah proses penggalian data dengan menggunakan teknik statistika, matematika, kecerdasan buatan, yang berguna untuk mencari informasi yang bermanfaat.

Clustering

Clustering merupakan proses pembagian data dalam suatu himpunan ke dalam beberapa kelompok yang memiliki data yang sama dengan suatu kelompok besar yang memiliki data yang sama dengan data yang lainnya. Potensi *Clustering* bisa digunakan untuk mengetahui struktur dalam data yang dipakai lebih lanjut dalam berbagai aplikasi secara luas seperti klarifikasi, pengolahan gambar, dan pengenalan pola (Sadewo et al., 2017). Metode *Clustering* yang paling banyak dipelajari adalah metode partisi dan metode hirarki. Metode partisi bertujuan untuk menemukan pengelompokan yang terdapat dalam data dengan mengoptimalkan fungsi dan tujuan yang dapat meningkatkan partisi. Sedangkan, metode hirarki merupakan metode yang melakukan pendekatan dengan mengembangkan struktur berbasis pohon biner yang disebut dendogram (Wulandari, 2020). Tujuan *Clustering* yaitu mengelompokkan objek atau persamaan karakteristik yang dimiliki. Hasil *cluster* suatu objek harus memiliki internal (*within cluster*) homogenitas yang tinggi dan memiliki eksternal (*between cluster*) heterogenitas yang tinggi. Jika pengelompokan berhasil, objek dalam satu *cluster* akan selalu mendekat, dan jika *cluster* berbeda maka akan saling menjauh satu sama lain (Halim & Widodo, 2017). Menurut (Fakhriza & Umam, 2021) dengan menggunakan *Clustering*, kita bisa mengklasifikasikan persebaran daerah yang padat, menemukan pola-pola distribusi secara keseluruhan, dan menemukan keterkaitan yang menarik antara atribut data. Beberapa kebutuhan *Clustering* dalam data mining meliputi skalabilitas, kemampuan untuk menangani dimensionalitas yang tinggi, menangani data yang *noise*, dan dapat diterjemahkan dengan mudah.

K-Means

K-Means merupakan salah satu metode *cluster analysis* nor hirarki yang berguna untuk mempartisi objek yang ada kedalam satu atau lebih besar cluster atau kelompok objek berdasarkan karakteristiknya, maka objek yang memiliki karakteristik yang sama akan digabung dengan karakteristik yang sama juga dan objek yang memiliki karakteristik yang berbeda akan ditempatkan di *cluster* yang lainnya (Nur Khormarudin, 2016). *K-Means* adalah metode *Clustering* berbasis jarak yang membagi data ke dalam sejumlah cluster dan algoritma ini hanya bekerja pada atribut numeric. Algoritma *K-Means* termasuk *partitionin Clustering* yang memisahkan data ke daerah yang terpisah (Rizki et al., 2021). Menurut (Sutoyo, n.d.) Prinsip utama *K-Means* adalah menyusun K atau pusat massa (centroid) dari

sekumpulan data berdimensi. Algoritma *K-Means* membutuhkan parameter input sebanyak k dan membagi sekumpulan n objek kedalam k cluster sehingga tingkat kemiripan antar anggota *cluster* tinggi sedangkan tingkat kemiripan dengan anggota pada cluster lain rendah. Kemiripan anggota dengan cluster disebut *centroid cluster*. Pada algoritma *K-Means* mempunyai beberapa tahapan atau langkah dalam prosesnya, yaitu:

1. Menentukan banyaknya jumlah *cluster* yang dibentuk.
2. Memilih data untuk pusat *cluster* atau *centroid* secara random.
3. Golongkan data sesuai dengan nilai jarak paling kecil tiap data dan hitung hingga nilai *centroid* yang dihasilkan stabil atau tidak berubah. Perhitungan dilakukan dengan menggunakan rumus *Euclidean Distance* dengan rumus sebagai berikut:

$$d(x,y) = |x - y| = \sqrt{\sum_{i=0}^n (x_i - y_i)^2}$$

Penjelasan:

d = jarak yang ada pada data antar titik x dan y

x = titik objek pada sebuah data

y = titik centroid pada data

I = banyaknya atribut pada data

n = Semua jumlah objek

Python

Python merupakan Bahasa pemrograman tingkat tinggi dengan sistem dinamis, mudah dipelajari karena sintaksnya yang sederhana. Karena berfokus pada keterbacaan kode. *Python* mendukung modul dan paket, yang mendorong modularitas program. *Python* dapat digunakan di banyak domain aplikasi dan mendukung banyak protokol internet dan mudah digunakan untuk socket interface (Destyara Zanneta Hidayatullifa, 2019).

Metode Elbow

Metode Elbow adalah metode untuk menghasilkan informasi dalam menentukan jumlah cluster terbaik dengan cara melihat grafik perbandingan antara jumlah yang akan membentuk siku pada suatu titik (Winarta & Kurniawan, 2021). Metode Elbow memberikan ide dengan memberikan cara untuk memilih nilai cluster kemudian menambah nilai cluster tersebut untuk dijadikan model data dalam penentuan cluster terbaik (Muningsih & Kiswati, 2018). Jumlah cluster k yang dihasilkan dari pengujian dengan *K-Means* dievaluasi menggunakan teknik SSE.

SSE

SSE merupakan cara dalam melakukan validasi cluster melalui jumlah kuadrat setiap anggota cluster menuju pusatnya. Semakin jauh jarak yang membentuk titik siku, maka jumlah cluster tersebut menjadi yang paling optimal (Jollyta et al., 2019). Rumus SSE adalah sebagai berikut :

$$SSE = \sum_{k=1}^k \sum_{x_i \in S_k} \|x_i - C_k\|_2^2$$

Keterangan:

- K = total *cluster*
- X_i = data ke- i
- C_k = awal *cluster*
- n = total data

HASIL DAN PEMBAHASAN

Business Understanding

Pada tahap ini diketahui permasalahan dalam objek untuk menentukan tujuan bisnis. Objek pada penelitian ini yaitu data penjualan obat di Apotek Amarta Sehat tahun 2021. Data yang sebelumnya hanya menjadi arsip penjualan saja, dapat dianalisa dan menjadi informasi yang bermanfaat untuk strategi bisnis. Hasil analisa bisa membantu pemilik apotek untuk menentukan suatu pola pemasaran yang efektif, yang dilihat dari pengelompokan obat yang sudah dibuat. Maka tujuan dalam penelitian ini adalah untuk melakukan *Clustering* obat untuk menentukan pola pemasaran efektif di Apotek Amarta Sehat menggunakan algoritma *K-Means*.

Data Understanding

Hasil dari tahap ini adalah pengumpulan dan eksplorasi data. Data yang digunakan dalam penelitian ini berasal dari data penjualan Apotek Amarta Sehat dari bulan Januari-Desember tahun 2021 yang dapat dilihat pada **Gambar 1**. Eksplorasi data dilakukan untuk melihat apakah ada data yang *missing value* dan *duplicated* data. Pada *dataset* yang digunakan tidak terdapat *missing value*, namun untuk *duplicated* data ditemukan pada atribut “nama_obat” sebanyak 17498, dapat dilihat pada **Gambar 2**, data berduplikat dikarenakan dalam datanya diinput secara sehari-hari dengan obat yang sama, tidak berisi data yang *unique* saja.

	tanggal_penjualan	nama_obat	qty	harga	total_harga
0	1-Jan-21	triamnicolon	1.0	10000	10000
1	1-Jan-21	ciprofloxacin	1.0	8000	8000
2	1-Jan-21	voltadex	1.0	5000	5000
3	1-Jan-21	erphaflam	1.0	8000	8000
4	1-Jan-21	tolak angin	2.0	3500	7000
...
18436	31-Dec-21	coric	1.0	7000	7000
18437	31-Dec-21	metidrol	1.0	15000	15000
18438	31-Dec-21	divoltar	1.0	7000	7000
18439	31-Dec-21	supertetra	1.0	8500	8500
18440	31-Dec-21	omeprazole	1.0	7500	7500

18441 rows x 5 columns

Gambar 1. *Dataset* Penjualan Apotek

```
data.duplicated("nama_obat").sum()
```

17498

Gambar 2.. Melihat *Duplicated* data

Data Preparation

Pertama-tama dilakukan pemilihan data, Dari dataset yang sudah ada, tidak semua atribut akan digunakan karena tidak sesuai dengan kebutuhan pada proses pemodelan. Maka, atribut `tanggal_penjualan` akan dihapus seperti pada **Gambar 3**.

	nama_obat	qty	harga	total_harga
0	triamnicolon	1	10000	10000
1	ciprofloxacin	1	8000	8000
2	voltadex	1	5000	5000
3	erphaflam	1	8000	8000
4	tolak angin	2	3500	7000
...
18436	coric	1	7000	7000
18437	metidrol	1	15000	15000
18438	divoltar	1	7000	7000
18439	supertetra	1	8500	8500
18440	omeprazole	1	7500	7500

18441 rows x 4 columns

Gambar 3. Menghapus Atribut Penjualan

Setelah diketahui terdapat data duplikat pada tahap *data understanding*, selanjutnya data duplikat akan ditangani, data duplikat terjadi karena terdapat nama obat yang sama karena dataset ini berisi data penjualan sehari-hari, yang memungkinkan banyak nama obat yang terjual sama atau nama yang *unique* saja. Maka, peneliti menggabungkan nama obat yang sejenis dan menghitung total dari atribut `qty`, `harga`, dan `total harga` selama satu tahun. Setelah data duplikat ditangani, jumlah dataset Apotek Amarta Sehat yang berjumlah 18441 menjadi 943 data seperti pada **Gambar 4** berikut ini.

	qty	harga	total_harga
nama_obat			
acarbose	8	16000.0	128000
acethyl sistein	3	14000.0	42000
acetylsistein	3	14000.0	42000
acifar 400	8	13000.0	104000
actifed merah	2	57000.0	114000
...
zolagel	15	10000.0	150000
zolagel cr	2	8000.0	16000
zoracyn tab	2	22000.0	44000
zoralin	12	22000.0	264000
zoralin tab	2	22000.0	44000

943 rows × 3 columns

Gambar 4. Menangani *Duplicated* data

Modeling

Pada tahap ini dilakukan proses pemodelan algoritma *K-Means*. Pembuatan model dilakukan dengan 2 skenario. Pada skenario pertama yaitu peneliti memilih nilai K atau *cluster* dan pusat *clusternya* secara acak yaitu sebanyak 3 *cluster*. Namun, setelah dilakukan dengan melihat *cluster* yang optimal dengan menggunakan metode *Elbow*, hasil yang direkomendasikan oleh metode *Elbow* yaitu sebanyak 5 *cluster*, hal ini yang membuat skenario 2 terjadi.

Skenario pertama yang dilakukan dalam pemodelan ini yaitu memilih nilai K secara acak, dalam skenario 1 memilih *cluster* sebanyak 3 *cluster*. Setelah memilih nilai K, selanjutnya masuk ke dalam proses *K-Means*. Dapat dilihat pada **Gambar 5** berikut ini.

```
kmeans = KMeans(n_clusters=3, random_state = 42)
scaler = MinMaxScaler()
data = scaler.fit_transform(data)
kmeans.fit(data)
centroid = kmeans.cluster_centers_
print(centroid)
kmeans.labels_

[[0.02623812 0.10889316 0.02581054]
 [0.01114223 0.42687606 0.05069695]
 [0.32316156 0.10046059 0.31561313]]
```

Gambar 5. Proses *K-Means cluster* sebanyak 3

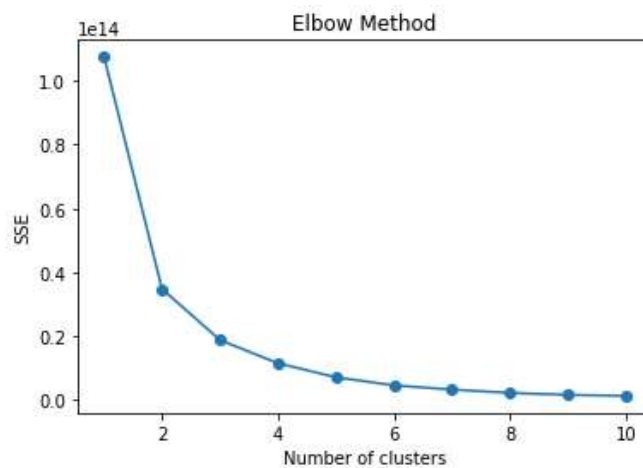
Selanjutnya pada **Gambar 6** proses yang dilakukan adalah mengetahui hasil *cluster*, pada tahap ini data sudah terbagi ke dalam masing-masing *cluster*. *Cluster* 0 sebanyak 786 data, *cluster* 1 sebanyak 106 data, dan *cluster* 2 sebanyak 51 data.

```
from collections import Counter
Counter(kmeans.labels_)

Counter({0: 786, 1: 106, 2: 51})
```

Gambar 6. Hasil Skenario 1

Setelah skenario pertama sudah selesai dilakukan dan setiap *value* sudah memperoleh setiap *clusternya*. Selanjutnya, mencoba melihat dari metode *Elbow*, hal ini membantu dalam menentukan jumlah *cluster* yang paling sesuai. Pada proses analisis untuk mendapatkan nilai K yang optimal dapat menggunakan metode *Elbow* dengan bantuan *Sum of Square Error* (SSE). Hasil dari metode *Elbow* menunjukkan *cluster* yang optimal yaitu *cluster* 5 seperti pada **Gambar 7** berikut ini



Gambar 7. Rekomendasi Nilai K dari *Elbow*

Skenario kedua terjadi setelah mendapat rekomendasi dari metode *Elbow*, maka untuk memilih nilai K yaitu sebanyak 5 *cluster* sesuai dengan rekomendasi metode *Elbow*, selanjutnya masuk ke dalam proses *Clustering* menggunakan algoritma *K-Means*. Hasil skenario 2 dapat dilihat pada **Gambar 8**.

```
from collections import Counter
Counter(kmeans.labels_)

Counter({0: 326, 2: 131, 1: 399, 3: 76, 4: 11})
```

Gambar 8. Hasil Skenario 2

Setelah skenario kedua sudah selesai dilakukan dan setiap *value* sudah memperoleh setiap *clusternya* dengan asumsi *cluster* 0 sebanyak 326 data, *cluster* 1 sebanyak 399 data, *cluster* 2 sebanyak 131 data, *cluster* 3 sebanyak 76 data, dan

cluster 4 sejumlah 11 data. Selanjutnya, masuk ke tahap *evaluation*, pada tahap ini akan mengetahui apakah *cluster* yang dipilih sudah optimal.

Evaluation

Hasil *Clustering* menggunakan algoritma *K-Means* selanjutnya akan dilakukan evaluasi dengan menggunakan *Sum of Square Error* yang ditunjukkan pada **Gambar 9**. Kualitas *cluster* yang dihasilkan dari *elbow* untuk menentukan *cluster* yang optimal dibantu dengan evaluasi menggunakan *Sum of Square Error*, dikatakan baik apabila nilai indeksnya paling kecil. Evaluasi nilai *cluster* menggunakan SSE menghasilkan hasil *Clustering* dengan $K = 5$. *Cluster* 5 dipilih karena nilai evaluasi tidak jauh dengan nilai kluster berikutnya, dan jumlah kluster yang dihasilkan tidak terlalu banyak, maka *cluster* 5 dipilih pada penelitian ini.

```
SSE for k(Cluster) = 1 is 107504675390014.31
SSE for k(Cluster) = 2 is 34825248312524.254
SSE for k(Cluster) = 3 is 18815355941898.73
SSE for k(Cluster) = 4 is 11545300942855.58
SSE for k(Cluster) = 5 is 7154215036701.358
SSE for k(Cluster) = 6 is 4622553277316.693
SSE for k(Cluster) = 7 is 3309085691057.889
SSE for k(Cluster) = 8 is 2260672377576.9385
SSE for k(Cluster) = 9 is 1638171227238.107
```

Gambar 9. Hasil Evaluasi SSE

Hasil evaluasi menggunakan *Sum of Square Error* menunjukkan bahwa *cluster* 5 adalah *cluster* yang paling optimal yaitu sebesar 7154215036292.542.

Deployment

Setelah tahap pemodelan sudah dilakukan, selanjutnya adalah tahap penyebaran. Dalam tahap ini dilakukannya proses laporan yang disampaikan kepada pemilik *owner* Apotek Amarta Sehat sebagai rekomendasi dalam menentukan suatu pola pemasaran yang efektif, yang dapat dilihat dari *Clustering* obat di Apotek Amarta Sehat pada **Tabel 1** berikut ini. Cluster yang terbentuk sebanyak 5 dengan deskripsi C4 adalah kriteria obat sangat laris, C3 adalah kriteria obat laris, C2 adalah kriteria obat cukup laris, C1 adalah kriteria obat kurang laris, C0 adalah kriteria obat sangat kurang laris.

Tabel 1. Hasil Pengelompokan Obat Terlaris

Cluster	Nama Obat	Jumlah	Kriteria
C4	Amoxicilin, Dexteem Plus, Erphaflam, Incidal, Paramex, Promagh Tab, Pyrexin, Rhemafar, Supertetra, Tolak Angin, Voltadex.	11	Sangat Laris
C3	Allofar 100mg, Amlodipin 5mg, Cataflam 50, Divoltar, Ermetasone, ..., ..., ..., Trisela.	76	Laris

C2	Bioplacenton, Azofir, Cendo Lyteers, Polysilane Syr, Viostin, Venaron, ..., ..., ..., Zoralin.	131	Cukup Laris
C1	Allopurinol 100, Piroxicam 10mg, Rhemafar, Rhemacyl, Salycil Cito, ..., ..., ..., Wiros.	399	Kurang Laris
C0	Acetylsistein, Afibramol, Allerin, Alphara, Anacetine Plus, ..., ..., ..., Yasimox.	326	Sangat Kurang Laris

KESIMPULAN

Pada penelitian ini diketahui bahwa *Clustering* obat untuk menentukan pola pemasaran efektif di Apotek Amarta Sehat dapat dilakukan dengan teknik *data mining* dengan menggunakan algoritma *K-Means*. Data yang digunakan untuk penelitian ini yaitu data penjualan obat dari bulan Januari-Desember tahun 2021. *Clustering* obat dengan menggunakan *K-Means* menghasilkan 5 *cluster* yang telah direkomendasikan oleh Metode *Elbow*. Hasilnya, sebanyak 11 obat masuk ke dalam jenis obat dengan kategori sangat laris, 76 obat masuk ke dalam jenis obat dengan katgeori laris, 131 obat masuk ke dalam jenis obat dengan kategori cukup laris, 399 obat masuk ke dalam jenis obat dengan kategori kurang laris, dan 326 obat masuk ke dalam jenis obat dengan kategori sangat kurang laris. Dengan mengetahui hasil *Clustering* obat yang ada, dapat membantu pemilik apotek untuk menentukan suatu pola pemasaran yang efektif. Hasil Evaluasi *K-Means* menghasilkan *cluster* yang optimal adalah pada *cluster* 5. Parameter penentu *cluster* terbaik ditentukan dengan nilai *Sum of Square Error* (SSE) yang paling kecil, berikut perhitungan seluruh jarak pada masing-masing data dengan nilai titik pusat *clusternya* menghasilkan nilai SSE pada *cluster* 1 sebesar 107504675390014.31, *cluster* 2 sebesar 34825248312524.254, *cluster* 3 sebesar 18815355941898.73, *cluster* 4 sebesar 11545300942855.58, dan *cluster* 5 sebesar 7154215036292.542. Pada penelitian ini urutan *cluster* terbaik didapatkan dari *cluster* 5 dengan nilai *cluster* sebesar 7154215036292.542.

DAFTAR PUSTAKA

- F., Kesuma, F. T., & Tamba, S. P. (2020). Penerapan Data Mining Untuk Menentukan Penjualan Sparepart Toyota Dengan Metode *K-Means Clustering*. *Jurnal Sistem Informasi Dan Ilmu Komputer Prima (JUSIKOM PRIMA)*, 2(2), 67–72. <https://doi.org/10.34012/jusikom.v2i2.376>
- Bramasta, F. A., & Halilintar, R. (2021). Penerapan Data Mining Untuk Menentukan Strategi Penjualan Toko Sepatu. *Prosiding SEMNAS INOTEK ...*, 236–241. <https://proceeding.unpkediri.ac.id/index.php/inotek/article/view/1135%0Ahttps://proceeding.unpkediri.ac.id/index.php/inotek/article/download/1135/736>
- Cobit, M. F., & Utami, E. (2019). *Jurnal Informasi Dan Komputer Vol : 7 No : 2 Thn .: 2019 Analisa Infrastruktur Teknologi Informasi Jurnal Informasi Dan Komputer Vol : 7 No : 2 Thn .: 2019*. 9–18.

- Destyara Zanneta Hidayatullifa. (2019). Rancang Bangun Pembuatan Sistem Pengiriman Sensor Secara Real Time Menggunakan Python dan Raspberry Pi. *Risalah Fisika*, 3(2), 43–46. <https://doi.org/10.35895/rf.v3i2.154>
- Fakhriza, M. H., & Umam, K. (2021). *ANALISIS PRODUK TERLARIS MENGGUNAKAN METODE K-MEANS Means Clustering dalam Pengelompokan*. 8–15.
- Halim, N. N., & Widodo, E. (2017). *Clustering* dampak gempa bumi di indonesia menggunakan kohonen self organizing maps. *Prosiding SI MaNIS (Seminar Nasional Integrasi Matematika Dan Nilai Islami)*, 1(1), 188–194. <http://conferences.uin-malang.ac.id/index.php/SIMANIS/article/view/62>
- Handoko, K. (2016). Penerapan Data Mining Dalam Meningkatkan Mutu Pembelajaran Pada Instansi Perguruan Tinggi Menggunakan Metode *K-Means Clustering* (Studi Kasus Di Program Studi Tkj Akademi Komunitas Solok Selatan). *Jurnal Teknologi Dan Sistem Informasi*, 02(03), 31–40. <http://teknosi.fti.unand.id/index.php/teknosi/article/view/70>
- Jollyta, D., Efendi, S., Zarlis, M., & Mawengkang, H. (2019). Optimasi Cluster Pada Data Stunting: Teknik Evaluasi Cluster Sum of Square Error dan Davies Bouldin Index. *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 1(September), 918. <https://doi.org/10.30645/senaris.v1i0.100>
- Muningsih, E., & Kiswati, S. (2018). Sistem Aplikasi Berbasis Optimasi Metode Elbow Untuk Penentuan *Clustering* Pelanggan. *Joutica*, 3(1), 117. <https://doi.org/10.30736/jti.v3i1.196>
- Nur Khormarudin, A. (2016). Teknik Data Mining: Algoritma *K-Means Clustering*. *Jurnal Ilmu Komputer*, 1–12. <https://ilmukomputer.org/category/datamining/>
- Rizki, M. Y., Maysaroh, S., & Windarto, A. P. (2021). Implementasi *K-Means Clustering* Dalam Mengelompokkan Minat Membaca Penduduk Menurut Wilayah. *JUST IT: Jurnal Sistem Informasi, Teknologi Informasi Dan Komputer*, 11(2), 41. <https://doi.org/10.24853/justit.11.2.41-49>
- Sadewo, M. G., Windarto, A. P., & Hartama, D. (2017). Penerapan Datamining Pada Populasi Daging Ayam Ras Pedaging Di Indonesia Berdasarkan Provinsi Menggunakan *K-Means Clustering*. *InfoTekJar (Jurnal Nasional Informatika Dan Teknologi Jaringan)*, 2(1), 60–67. <https://doi.org/10.30743/infotekjar.v2i1.164>
- Siregar, M. H. (2018). Data Mining Klasterisasi Penjualan Alat-Alat Bangunan Menggunakan Metode *K-Means* (Studi Kasus Di Toko Adi Bangunan). *Jurnal Teknologi Dan Open Source*, 1(2), 83–91. <https://doi.org/10.36378/jtos.v1i2.24>
- Sutoyo, M. N. (n.d.). *Algoritma K-Means*. 1, 1–7.
- Winarta, A., & Kurniawan, W. J. (2021). Optimasi cluster *K-Means* menggunakan metode elbow pada data pengguna narkoba dengan pemrograman python. *Jurnal Teknik Informatika Kaputama (JTIK)*, 5(1), 113–119.
- Wulandari, S. (2020). *Clustering* Kecamatan Di Kota Bandung Berdasarkan Indikator Jumlah Penduduk Dengan Menggunakan Algoritma *K-Means*. *Seminar Nasional Riset Dan Teknologi (SEMNAS RISTEK)*, 128–132.