

# Klasterisasi Data Obat Farmasi Berdasarkan Jumlah Persediaan Dengan Menggunakan Metode K-Means

Heri Supriyanto<sup>1\*</sup>, Mohammad Al Hafidz<sup>2</sup>, Ari Cahaya Puspitaningrum<sup>3</sup>,  
Rayhan Abdillah Putra Firmansyah<sup>4</sup>, Rafi Zuhdi<sup>5</sup>

<sup>1,2,3,4,5</sup>Program Studi Sistem Informasi, Universitas Hayam Wuruk Perbanas, Surabaya, Jawa Timur

Email: <sup>1\*</sup>heri.supriyanto@hayamwuruk.ac.id, <sup>2</sup>mohammad.hafidz@hayamwuruk.ac.id,

<sup>3</sup>ari.cahaya@hayamwuruk.ac.id, <sup>4</sup>202102021003@mhs.hayamwuruk.ac.id, <sup>5</sup>2102021009@mhs.hayamwuruk.ac.id

(Naskah masuk: 10 Jul 2024, direvisi: 2 Sep 2024, diterima: 9 Sep 2024)

## Abstrak

Instalasi Farmasi memiliki peran penting terhadap pelayanan kesehatan di sebuah fasilitas kesehatan. Farmasi bertanggung jawab atas pengelolaan, pengadaan, penyimpanan, distribusi, dan penggunaan persediaan obat. Persediaan obat merupakan bagian penting dalam memastikan ketersediaan, aksesibilitas, dan penggunaan obat yang efektif serta aman bagi pasien. Tujuan penelitian ini untuk melakukan klasterisasi data obat yang berguna untuk meningkatkan efisiensi proses manajemen persediaan obat, sehingga dapat menghindari kelebihan atau kekurangan yang dapat mengganggu kelancaran layanan pemberian obat dan mencegah terjadinya kerugian penjualan obat. Pengelompokan data dilakukan dengan memanfaatkan data Persediaan Obat dari data masa lalu yaitu data transaksi pembelian dan penjualan dengan memanfaatkan teori *Data Mining* dengan menggunakan metode *Clustering* yaitu *K-Means*. Dataset pada penelitian ini sebanyak 1.389 dengan 6 variabel. Sebelum dilakukan klasterisasi dilakukan proses optimasi jumlah klaster dengan dua metode yaitu Metode *Elbow* dan Metode *Gap* Statistik. Hasil kedua metode tersebut menunjukkan nilai optimasi k klaster  $k = 3$ . Hasil klasterisasi yaitu Klaster 1 sebanyak 41 data obat yang menunjukkan golongan obat Generik. Klaster 2 sebanyak 116 data obat yang menunjukkan obat Paten. Kedua klaster tersebut menunjukkan tingkat penjualan yang kurang cepat (*slow moving*). Sedangkan pada Klaster 3 sebanyak 1.232 data obat yang menunjukkan gabungan dari golongan obat generik dan paten yang memiliki tingkat penjualan yang cukup cepat (*fast moving*).

**Kata Kunci:** Data Mining, Clustering, *K-Means*, Persediaan Obat.

## *Clustering Pharmaceutical Drug Data Based on Total Inventory Using the K-Means Method*

### Abstract

The Pharmacy Installation plays an important role in health services in a health facility. Pharmacy is responsible for the management, procurement, storage, distribution, and utilisation of drug supplies. Drug inventory is an important part of ensuring the availability, accessibility, and effective and safe use of drugs for patients. The purpose of this study is to cluster drug data that is useful for improving the efficiency of the drug inventory management process, so as to avoid excess or shortages that can interfere with the smooth delivery of drug services and prevent drug sales losses. Data clustering is done by utilising Drug Inventory data from past data, namely purchase and sales transaction data by utilising Data Mining theory using the Clustering method, namely *K-Means*. The dataset in this study is 1,389 with 6 variables. Before clustering, the process of optimising the number of clusters is carried out with two methods, namely the Elbow Method and the Statistical Gap Method. The results of the two methods show the optimisation value of  $k$  clusters  $k = 3$ . The clustering results are Cluster 1 as much as 41 drug data showing the Generic drug class. Cluster 2 is 116 drug data showing Patent drugs. Both clusters show a less rapid sales rate (*slow moving*). While in cluster 3 there are 1,232 drug data showing a combination of generic and patent drugs which have a fairly fast sales rate (*fast moving*).

**Keywords:** Data Mining, Clustering, *K-Means*, Medicine Inventory.

## I. PENDAHULUAN

Fasilitas kesehatan seperti rumah sakit, puskesmas, dan klinik memiliki Unit Instalasi Farmasi. Unit tersebut bertugas untuk bertanggung jawab atas pengelolaan, pengadaan, penyimpanan, distribusi, dan penggunaan persediaan obat. Instalasi Farmasi memiliki peran penting dalam memastikan bahwa obat-obatan yang diperlukan pasien tersedia dan bahwa obat-obatan yang diberikan terjamin pada kualitas dan keamanan berdasarkan penetapan standar yang sudah ditentukan. Selain itu, farmasi perlu memastikan rantai pasok obat dari sisi pengadaan, pengeluaran, dan penggunaan obat untuk mengoptimalkan hubungan antara biaya, manfaat, dan risiko dalam mengelola persediaan obat [1].

Kondisi saat ini rantai pasok sudah mengalami kemajuan yang pesat dalam bentuk digital sehingga menjadikan keberadaan data sangat penting. Data dapat diolah dan bermanfaat untuk pengambilan keputusan dan kebijakan pada sebuah organisasi [2]. Saat melakukan pembelian obat, tentu diperlukan sebuah data masa lalu untuk dapat memberikan informasi terkait data obat yang akan dibeli. Dengan adanya sebuah perencanaan yang baik maka persediaan obat akan tersedia dengan jumlah yang cukup pada waktu tertentu dan tersedia pada saat dibutuhkan [3].

Persediaan obat merupakan bagian sistem pelayanan kesehatan yang memiliki peranan krusial dalam memastikan ketersediaan, aksesibilitas, dan penggunaan obat yang efektif serta aman bagi pasien [4]. Ketersediaan obat yang memadai merupakan faktor penting dalam mengobati suatu penyakit. Ketersediaan obat yang sesuai juga dapat meningkatkan kualitas pelayanan kesehatan [5]. Keterbatasan persediaan obat dapat menyebabkan terhambatnya proses pengobatan, peningkatan risiko komplikasi, dan bahkan kematian pada pasien karena pemberian obat memiliki aturan yang tidak sama tergantung pada umur, jenis kelamin, hingga durasi pengobatan. Berdasarkan hal tersebut, salah satu upaya diperlukan untuk dilakukan pengelompokan data obat untuk memberikan gambaran dan masukan atas data dalam pengambilan keputusan dalam proses perencanaan dan pengelolaan kebutuhan obat dimasa yang akan datang.

Beberapa penelitian sebelumnya terkait klasterisasi obat pernah dilakukan. Pertama dari Izzah pada tahun 2022. Penelitian tersebut memberikan penjelasan tentang penerapan algoritma *K-Means* untuk proses perencanaan kebutuhan obat di Klinik Citra Medika. Metode penelitian yang digunakan dalam penelitian ini mencakup penggunaan algoritma *K-Means* untuk mengelompokkan data obat berdasarkan parameter Harga, Kategori, Stok Awal, dan Total Pemakaian. Sebanyak 484 data hasil *cleaning* dari total 488 sampel data jenis obat di Klinik Citra Medika diuji untuk menentukan 5 kelompok pemakaian obat. Setiap kelompok obat akan merepresentasikan prediksi jumlah order persediaan di masa mendatang berdasarkan parameter total pemakaian. Hasil penelitian menyimpulkan bahwa penggunaan algoritma *K-Means* dalam merencanakan kebutuhan penggunaan obat pada Klinik Citra Medika lebih mendekati faktual penggunaan dibandingkan dengan perencanaan yang dilakukan secara manual. Prediksi jumlah persediaan untuk setiap jenis obat di

masa yang akan datang didasarkan pada parameter total pemakaian, yang membantu dalam mengoptimalkan perencanaan persediaan obat [6].

Penelitian berikutnya adalah dari Nugroho pada tahun 2022. Penelitian tersebut berisi tentang implementasi algoritma *K-Means* untuk klasterisasi data obat di Rumah Sakit ASRI. Metode penelitian yang dipakai dalam penelitian ini adalah *Data Mining*, khususnya teknik *Clustering*. Dalam penelitian ini, laporan stok obat Rumah Sakit Ibu dan Anak ASRI Purwakarta dari bulan Januari hingga Juni 2021 yang digunakan sebagai dataset. Pada penelitian ini menghasilkan kesimpulan bahwa penggunaan algoritma *K-Means* dalam klasterisasi data obat, berhasil dikelompokkan menjadi 2 *cluster*. *Cluster* pertama terdiri dari 6 obat dengan pemakaian tinggi, sedangkan *cluster* kedua terdiri dari 933 obat dengan pemakaian rendah [7].

Penelitian terdahulu yang terakhir yaitu Holwati tahun 2023. Penelitian ini membahas tentang pengelompokan obat menggunakan algoritma *K-Means* berdasarkan ketersediaan dan persediaan barang di gudang. Hasil penelitian yang dilaporkan dalam artikel ini menunjukkan bahwa pengelompokan obat berdasarkan ketersediaan dan persediaan barang di gudang menghasilkan 2 jenis *cluster*. *Cluster* pertama menunjukkan kategori obat dengan kebutuhan tinggi, yaitu ketersediaan sebesar 71 dari 100 jenis obat. Sementara itu, *cluster* kedua menunjukkan kategori obat dengan ketersediaan sedang atau rendah, yaitu sebesar 29 dari jenis obat. Pengujian dilakukan menggunakan aplikasi *Rapid Miner* yang menyajikan gambaran di mana masing-masing klaster memiliki anggota kelompok klaster sesuai dengan perhitungan manual. *Cluster* 1 menunjukkan kebutuhan tinggi, sementara *Cluster* 2 menunjukkan ketersediaan sedang atau rendah dari data obat yang diuji [8].

Berdasarkan kasus dan penelitian terdahulu, meskipun algoritma *K-Means* efektif untuk pengelompokan data. Namun hubungan antar variabel sulit diidentifikasi hanya dengan pengelompokan berbasis jarak Euclidean. Penelitian sebelumnya sering mengabaikan aspek-aspek tertentu dari data yang bisa memperkaya proses *clustering*. Maka penelitian ini perlu menerapkan metode pengelompokan untuk memastikan setiap fitur data berkontribusi secara proporsional terhadap perhitungan jarak sehingga jumlah *k* klaster lebih optimal.

Berdasarkan hal tersebut, maka penelitian ini bertujuan untuk melakukan klasterisasi data obat yang dilakukan pada Rumah Sakit Swasta Kelas C di Kabupaten Bondowoso. Berdasarkan data di instalasi farmasi di rumah sakit tersebut, seringkali terjadi adanya stok obat berlebih dan stok obat yang kurang. Sehingga pihak farmasi seringkali membeli dari apotik luar atau mengeluarkan *copy* resep. Hal ini membuat layanan penjualan obat menjadi terganggu dan mengalami kesulitan dalam menentukan pengadaan obat untuk kebutuhan dimasa yang akan datang. Dalam melakukan pemesanan atau pembelian obat kepada *supplier*, unit pengadaan barang medis biasanya melakukan pencatatan secara manual yaitu dilakukan secara *Defacta*. *Defacta* merupakan proses pencatatan pesanan obat agar dapat memenuhi kebutuhan dan ketersediaan obat [9]. Tahapan untuk mendapatkan informasi item pembelian obat dilakukan dengan cara mendapatkan laporan dari masing-

masing unit farmasi penjualan (depo farmasi rawat inap dan farmasi rawat jalan) dimana masing-masing depo melakukan pencatatan data obat yang sudah habis dan catatan tersebut di berikan ke bagian pengadaan.

Proses pembelian atau pengadaan obat tersebut tentu tidak efisien karena disamping lama proses pengambilan keputusannya, juga tidak dapat memantau secara real kondisi stok persediaan secara keseluruhan. Sehingga yang terjadi data obat banyak yang tidak terpantau dan tidak digunakan. Sehingga banyak sekali data obat yang kadaluarsa (*expired date*) karena data pembelian yang tidak sesuai dengan kebutuhan. Berdasarkan informasi dari bagian pengadaan obat, penggunaan atau pemakaian obat biasanya tergantung pada kebiasaan dokter dalam memberikan resep dengan produk obat tertentu kepada pasien. Sehingga data obat tertentu bisa jadi tidak dipakai lagi dikarenakan dokter sudah tidak meresepkan lagi ke pasiennya. Pada Tabel 1 di bawah ini menunjukkan gambaran data obat yang sudah kadaluarsa.

Tabel 1. Data Obat Kadaluarsa

Tahun	Jumlah Jenis Obat	Jumlah Stok Persediaan	Nilai Persediaan
2019	91	1.101	24.365.322
2020	62	660	17.294.923
2021	195	2.258	29.637.433
2022	138	2.410	30.458.595
2023	56	489	15.891.325

Berdasarkan Tabel 1 di atas, data obat kadaluarsa pada tahun 2019 hingga tahun 2023 sebanyak puluhan hingga ratusan jenis obat dan nilai persediaan yang hilang mencapai puluhan juta. Hal ini menunjukkan adanya sebuah proses pembelian yang kurang optimal. Dampak dari kebijakan yang kurang baik, pihak rumah sakit mengalami banyak sekali kerugian seperti yang ditunjukkan pada Tabel 1 yaitu nominal nilai persediaan per tahun yang terlampaui tinggi. Padahal unit farmasi ini merupakan salah satu layanan aset yang paling besar di rumah sakit [10].

Berdasarkan permasalahan yang dijabarkan, penelitian ini memanfaatkan data transaksi masa lalu agar persediaan obat dapat dikendalikan dengan tujuan untuk menciptakan keseimbangan antara pembelian dengan obat yang tersedia. Pemanfaatan data transaksi masa lalu juga dapat digunakan untuk strategi penjualan pada instalasi farmasi, sehingga dapat menekan data persediaan obat yang kurang dan data yang tidak dipakai. Pengendalian atas ketersediaan obat ini sangat penting untuk menjamin efektifitas dan efisiensi pengelolaan obat pada proses instalasi farmasi. Solusi yang ditawarkan dalam penelitian ini dengan memanfaatkan data transaksi penjualan maupun data pembelian obat, selanjutnya diolah menggunakan konsep metode *Data Mining*. Pola yang dihasilkan atas proses pengolahan data dengan menggunakan data mining dapat digunakan sebagai masukan dalam mengambil keputusan [11]. Salah satu teori dalam proses *Data Mining* adalah algoritma *Unsupervised Learning* yang salah satunya adalah *Clustering*. Penelitian ini melakukan pengelompokan data obat farmasi berdasarkan jumlah persediaan dengan menggunakan jenis

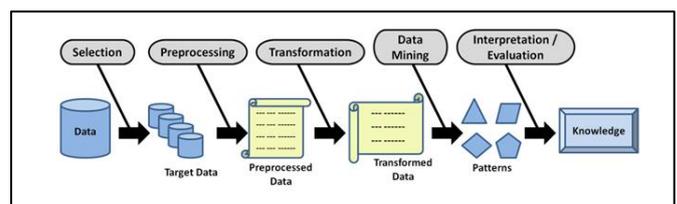
algoritma *clustering* yaitu metode *K-Means*. Metode *K-Means* merupakan alat dalam ilmu data (*data science*) yang berguna untuk membentuk struktur kluster pada beberapa kumpulan data yang ditandai dengan kesamaan terbesar dalam kluster yang sama dan ketidaksamaan terbesar di antara kluster yang berbeda [12].

Berdasarkan kasus tersebut, maka dibutuhkan untuk memahami pola kebutuhan stok persediaan yang optimal. Penentuan jumlah *k* kluster yang optimal dapat dilakukan dengan menggunakan metode *Elbow* dan Metode *Gap Statistik*. Dengan demikian, penelitian ini bertujuan untuk mengidentifikasi dan mengelompokkan data obat yang memiliki pola persediaan yang mirip menggunakan algoritma *K-Means* yang berguna untuk meningkatkan efisiensi proses manajemen persediaan obat. Sehingga dapat menghindari kelebihan atau kekurangan yang dapat mengganggu kelancaran layanan pemberian obat dan mencegah terjadinya kerugian penjualan obat.

## II. TINJAUAN PUSTAKA

### A. Data Mining

*Data mining* diartikan sebagai proses yang meliputi teori kecerdasan buatan, teknik statistik, matematika, dan *machine learning* dengan cara mengeksplorasi dan menganalisis sekumpulan data (*dataset*) untuk menemukan pola, relasi, dan informasi yang bermanfaat [13]. Terdapat dua jenis pendekatan utama pada *machine learning* yaitu *Supervised Learning* merupakan model dilatih menggunakan data yang sudah memiliki label atau output yang diketahui dan *Unsupervised Learning* merupakan model dilatih pada dataset yang tidak memiliki label atau kategori yang telah ditentukan sebelumnya. Selain itu, kategori besar Data Mining dapat diartikan sebagai proses penyaring atau menambang informasi dari kumpulan data yang besar. *Knowledge Discovery in Database (KDD)* adalah istilah lain pada *data mining*. KDD adalah proses ekstraksi pengetahuan yang bermanfaat, valid, dan dapat dipahami secara otomatis dari data mentah atau data yang belum diolah. KDD melibatkan serangkaian langkah dan teknik untuk mengidentifikasi pola, tren, dan informasi penting lainnya dari data besar dengan tujuan mendukung pengambilan keputusan yang lebih baik. Pada Gambar 1 di bawah ini menunjukkan gambaran proses KDD [14].



Gambar 1. Proses KDD

Berdasarkan Gambar 1 di atas, proses KDD terdiri dari beberapa tahapan seperti berikut:

1. *Selection*: digunakan untuk menghindari kesamaan dan perulangan yang tidak diperlukan, pilihan digunakan

untuk menentukan variabel yang akan diambil.

2. *Preprocessing*: Pada tahapan *preprocessing* terdiri atas dua proses, yaitu:
  - a. *Data Cleaning*: Menghilangkan data yang tidak diperlukan, misalnya seperti menangani *missing value*, *noise* data, serta menangani data yang tidak relevan dan tidak konsisten.
  - b. *Data Integration*: Proses ini merupakan proses menggabungkan beberapa data yang berasal dari beragam sumber agar memiliki format yang sama.
3. *Transformation*: Proses pengolahan *data mining* yang dapat mengubah format data ke ekstensi yang sesuai. Pada tahap ini perlu dilakukan persamaan *format* agar mempermudah dalam proses analisis data.
4. *Data Mining*: Tahapan ini merupakan proses utama dalam memperoleh pengetahuan baru dari kumpulan data yang diolah.
5. *Evaluation/Interpretation*: Tahapan ini merupakan proses identifikasi pola yang dianggap menarik ke dalam pengetahuan dasar yang diidentifikasi.
6. *Knowledge*: Pengguna akan melihat pola yang dibuat. Pada titik ini, semua orang yang akan digunakan untuk membuat keputusan dapat memahami informasi baru yang dibuat.

### B. Clustering

*Clustering* (pengelompokan) merupakan salah satu kategori pada *Unsupervised Learning* untuk memisahkan data menjadi beberapa kelompok berdasarkan kesamaan fitur. Dalam beberapa dekade terakhir, beberapa teknik *clustering* yang lebih baik telah diterapkan untuk berbagai aplikasi seperti biologi, psikologi, statistik, pengenalan pola, pencarian informasi, pembelajaran mesin, dan *data mining* [15]. Analisis berbasis *cluster* ini adalah salah satu teknik penggalian data yang paling populer yang sering digunakan dalam praktiknya [16].

### C. Standardisasi (Z-Score Normalization)

Standardisasi merupakan teknik yang digunakan dalam pengolahan data untuk mengubah nilai data sehingga memiliki distribusi dengan rata-rata 0 dan standar deviasi 1. Proses ini penting dalam analisis statistik dan *machine learning* karena membantu menetralkan skala variabel yang berbeda, sehingga setiap fitur memiliki pengaruh yang seimbang. Berikut Persamaan 1 ini rumus untuk menghitung Z-Score dari suatu data adalah [17]:

$$Z = \frac{X - \mu}{\sigma} \quad (1)$$

Keterangan:

$Z$  = Z-Score atau nilai standar dari data

$X$  = Nilai data yang akan distandarisasi  $\mu$

$\mu$  = Rata-rata dari seluruh data dalam *dataset*

$\sigma$  = Standar deviasi dari seluruh data dalam *dataset*

### D. Metode Elbow

Metode *Elbow* adalah teknik yang digunakan dalam analisis kluster untuk menentukan jumlah kluster optimal dalam suatu *dataset*. Nama "elbow" diambil dari bentuk grafisnya yang menyerupai siku. Teknik ini didasarkan pada evaluasi *Sum of Squared Errors* (SSE) dari *clustering* yang dilakukan dengan berbagai jumlah *cluster* [18].

Metode ini digunakan untuk menghasilkan informasi dalam menentukan nilai  $k$  atau jumlah *cluster* terbaik dengan cara melihat presentase hasil perbandingan antara jumlah *cluster* yang akan membentuk siku pada suatu titik. Untuk mendapatkan perbandingannya adalah dengan cara menghitung SSE (*Sum of Square Error*) dari masing-masing nilai  $k$ . Jika nilai  $k$  semakin besar, maka nilai dari SSE akan semakin kecil [19]. Berikut ini adalah tahapan algoritma metode *Elbow* dalam menentukan nilai  $k$  pada *K-Means* [20]:

1. Inialisasi nilai awal  $k$ ;
2. Tingkatkan nilai  $k$ ;
3. Menghitung jumlah hasil kuadrat kesalahan dari setiap nilai  $k$ ;
4. Analisis hasil jumlah kuadrat error dari nilai  $k$  yang mengalami penurunan drastis;
5. Cari dan tetapkan nilai  $k$  yang berbentuk siku.

### E. Metode Gap Statistic

Metode *Gap Statistic* adalah teknik yang digunakan untuk menentukan jumlah *cluster* optimal dalam analisis *cluster*. Metode ini membandingkan hasil *clustering* pada data asli dengan hasil *clustering* pada data acak yang memiliki distribusi yang sama. *Gap* statistik seperti terlihat pada Persamaan 2 digunakan untuk mengukur seberapa jauh hasil *clustering* pada data asli lebih baik dibandingkan dengan data acak [21]:

$$Gap_n(k) = E_n^* \log(W_{(k)}) - \log(W_{(k)}) \quad (2)$$

Keterangan:

$E_n^*$  = ukuran sample  $n$  dari distribusi referensi.

$\log(W_{(k)})$  = Hitung nilai rata-rata

Langkah-langkah metode *Gap Statistics* seperti berikut [22]:

1. *Cluster* data yang diamati, memvariasikan jumlah *cluster* dari  $k = k_1 \dots k_{max}$  dan hitung sesuai  $W_k$ .
2. Menghasilkan referensi dataset  $B$  dan setiap *cluster* dengan jumlah *cluster* bervariasi dari  $k = k_1 \dots k_{max}$  Hitung perkiraan kesenjangan statistik yang disajikan di Persamaan 2.
3. Kemudian hitung dengan rumus  $\bar{w} = \left(\frac{1}{B}\right) \sum_b \log(W_{kb}^*)$ , selanjutnya hitung standart deviasi dengan rumus  $sd_k = \sqrt{\left(\frac{1}{B}\right) \sum_b (\log(W_{kb}^*) - \bar{w})^2}$  dan tentukan hasilnya sebagai dengan persamaan  $s_k = sd_k \times \sqrt{1 + 1/B}$ .
4. Pilih jumlah *cluster* sebagai yang terkecil menggunakan rumus  $Gap_k \geq Gap(k + 1) - s_{k-1}$ .

**F. Metode K-Means**

Salah satu teknik data *clustering* non hierarki yang digunakan untuk mempartisi data ke dalam bentuk satu atau lebih kelompok adalah *K-Means*. Data yang diolah dengan menggunakan teknik *K-Means* memungkinkan memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lainnya. *K-Means* dapat dikatakan sebagai metode *clustering* berbasis jarak yang membagi data ke dalam sejumlah *cluster*. *K-Means* ini hanya bekerja pada atribut *numeric*. Metode *K-Means* dapat dikatakan metode *partitioning clustering* yang populer dan sering digunakan karena sifatnya yang relatif sederhana namun efektif [23].

*Partitioning clustering* adalah metode *clustering* di mana data dibagi menjadi beberapa klaster tanpa struktur hirarkis. Teknik *clustering* menggunakan *K-Means* dapat dikategorikan sebagai *unsupervised learning* yang digunakan untuk analisis data. Algoritma ini mengidentifikasi 'K' *centroid* dari dataset 'D' dan menetapkan titik data yang tidak tumpang tindih ke masing-masing *cluster* terdekat. Metode ini menggunakan pendekatan yang berulang, sehingga titik data dipindahkan ke *cluster* yang berbeda berdasarkan perhitungan *centroid* [24].

Berikut ini tahapan-tahapan dalam melakukan *clustering* menggunakan teknik *K-Means* sebagai berikut [25]:

1. Tentukan nilai *k* untuk jumlah *cluster* yang ingin dibuat
2. Inisialisasi *k* pusat *cluster*. Ini dapat dilakukan dengan berbagai cara. Namun cara yang paling banyak digunakan adalah dengan memilih secara *random*. Pusat *cluster* diberi nilai awal dengan angka-angka *random*.
3. Alokasikan semua data/objek ke dalam *cluster* terdekat. Kedekatan dua data/objek dapat didapatkan berdasarkan jarak kedua objek tersebut. Untuk dapat menghitung jarak semua data terhadap titik pusat *cluster* maka dapat menggunakan teori *Euclidean Distance* yang dirumuskan pada Persamaan 3 di bawah ini :

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \tag{3}$$

Keterangan:

*d* = Jarak antara objek dan

*p* = Jumlah data

*x<sub>ik</sub>* = Koordinat dari objek *i* pada dimensi *k*

*x<sub>jk</sub>* = Koordinat dari objek *j* pada dimensi *k*

4. Mengklasifikasikan/mengelompokkan setiap data berdasarkan kedekatannya dengan *centroid* (jarak terkecil).
5. Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang. Memperbaharui nilai *centroid*. Nilai *centroid* baru di peroleh dari rata-rata *cluster* yang bersangkutan dengan Persamaan 4:

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q \tag{4}$$

Keterangan:

$\mu_k$  = Titik *centroid* dari *cluster* ke - *k*

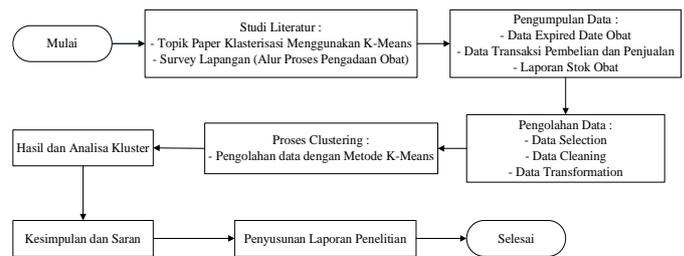
$N_k$  = Banyaknya data pada *cluster* ke - *k*

$x_q$  = data ke - *q* pada *cluster* ke - *k*

6. Melakukan perulangan dari langkah 3 hingga 5, sampai anggota tiap *cluster* tidak ada yang berubah. Jika langkah 6 telah terpenuhi, maka nilai pusat *cluster* ( $\mu_j$ ) pada iterasi terakhir akan digunakan sebagai parameter untuk menentukan klasifikasi data.

**III. METODE PENELITIAN**

Proses *cluster* data persediaan pada penelitian ini memiliki beberapa tahapan yang ditunjukkan pada gambar sebagai berikut:



Gambar 2. Tahapan Penelitian

**1. Studi Literatur**

Tahap penelitian ini melibatkan tinjauan literatur yang relevan seperti buku teks, para ahli, artikel jurnal yang relevan, dan makalah penelitian sebelumnya yang membahas teori pendekatan yang digunakan dalam penelitian tentang pengklasteran data obat pada salah satu rumah sakit. Pada penelitian ini akan dipelajari tentang proses *data mining* yang dilakukan yang tujuan untuk meningkatkan efisiensi pembelian obat sehingga jumlah pembelian dapat sesuai dengan kebutuhan.

**2. Pengumpulan Data**

Tahap pengumpulan data ini merupakan tahap yang utama karena berfungsi sebagai masukan pada proses *clustering* data. Data yang dikumpulkan berasal dari data rumah sakit dan data yang digunakan adalah data sekunder yaitu data transaksi pembelian obat, data transaksi penjualan obat, data laporan persediaan, dan data tanggal kadaluarsa obat. Periode pengumpulan data adalah pada bulan Agustus 2023 hingga Mei 2024. Pengumpulan data dilakukan dengan berbagai cara, antara lain observasi langsung, kuesioner, wawancara, dan masukan langsung dari manajemen rumah sakit.

**3. Pengolahan Data**

Langkah ini merupakan proses sebelum dilakukan pengelompokan dengan menggunakan metode *K-Means*. Pada penelitian ini menggunakan data obat yang terdiri dari berbagai jenis obat berdasarkan data penjualan dan pembelian. Dari beberapa jenis data yang dikumpulkan, proses ini melakukan transformasi data sesuai dengan kebutuhan yang dapat digunakan untuk proses pengklasteran. Kebutuhan data yang dibutuhkan seperti stok awal, stok masuk, stok keluar,

dan stok akhir. Proses ini juga dilakukan pencarian optimalisasi jumlah kluster dengan menggunakan dua metode yaitu metode *Elbow* dan Metode *Gap* Statistik.

4. Proses *Clustering*

Proses *Clustering* dilakukan terhadap data yang sudah melalui proses pengolahan sesuai dengan kebutuhan. Proses Kluster pada penelitian ini menggunakan metode *K-Means*. *K-Means* digunakan pada penelitian ini karena merupakan satu dari beragam teknik *clustering* yang berguna untuk mengelompokkan beberapa data ke dalam suatu kelompok berdasarkan kemiripan karakteristik tertentu. Proses ini menggunakan langkah-langkah berdasarkan referensi yang terdapat pada bab tinjauan pustaka pada sub bab metode *K-Means*. Salah satu proses metode ini adalah penentuan jumlah kluster. Untuk jumlah kluster akan ditentukan dari hasil perbandingan nilai *k* optimal dari metode *Elbow* dan *Gap* Statistik. Sedangkan proses klasterisasi data menggunakan tool R Studio.

5. Hasil dan Analisis *Cluster*

Proses ini dilakukan setelah mendapatkan hasil kluster. Deskripsi hasil melibatkan penjelasan tentang struktur data yang ditemukan setelah penerapan algoritma *clustering*. Sedangkan analisis *cluster* melibatkan evaluasi lebih mendalam dari hasil *clustering* untuk memahami dan menginterpretasikan makna dari kluster yang terbentuk.

IV. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data sekunder yang berasal dari laporan stok, laporan pembelian dan penjualan di Instalasi Farmasi dari bulan Agustus 2023 sampai dengan bulan Mei 2024. Jumlah jenis data obat keseluruhan sebanyak 3699. Dari jumlah data tersebut, dilakukan proses *Pre-Processing* yaitu menangani *noise*, mengoreksi data yang tidak konsisten dan menghilangkan data yang tidak diperlukan dikarenakan data obat lama yang sudah tidak memiliki stok. Proses *Pre-Processing* pada penelitian ini menggunakan teknik filter data SQL. Adapun dataset yang terbentuk dan digunakan pada penelitian ini sebanyak 1389 dengan 6 variabel yaitu kode barang, nama barang, stok awal, stok masuk, stok keluar, dan stok akhir seperti yang ditunjukkan seperti Tabel 2 berikut:

Tabel 2. *Dataset* Penelitian

No	Nama Barang	Stok Awal	Stok Masuk	Stok Keluar	Stok Akhir
1	ABBOCATH 14	3	2	4	1
2	ABBOCATH 16	6	12	13	5
3	AB VASK 10 MG TAB	0	12	10	2
4	ACARBOSE 100	1.500	2.320	2.800	1.020
5	ACARBOSE 50	35	3.507	2.470	1.072
6	ACETYLCYSTEINE KAP	4.990	18.360	22.546	804
7	ACRAN INJ (RANITIDINE)	125	204	301	28
8	ACYCLOVIR CR 5 GR	38	159	168	29

9	ACYCLOVIR TAB 400 MG	1.727	4.974	5.494	1.207
10	ADALAT OROS - TAB 30 MG	10.827	26.265	35.923	1.169
11	AKILEN TETES TELINGA	224	287	492	19
12	ALA 600	44	76	65	55
...	....	...	...	...	...
...	....	...	...	...	...
1.388	ZITANID TAB	67	90	14	143
1.389	ZYFORT INJ	20	0	0	20

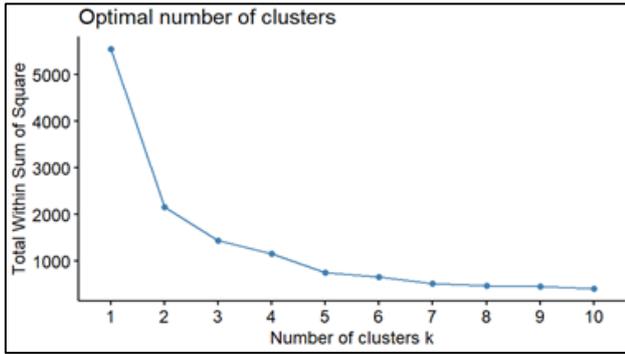
Selanjutnya adalah dilakukan Normalisasi atau Tranformasi Data. Proses ini berfungsi untuk memastikan bahwa setiap fitur berkontribusi secara proporsional terhadap perhitungan jarak. Hal ini membantu *K-Means* dalam menemukan pusat kluster (*centroid*) yang lebih representatif. Sehingga kluster yang dihasilkan lebih relevan dan akurat. Metode Transformasi Data pada penelitian ini menggunakan *Z-Score Normalization*. Dalam proses ini merubah data untuk memiliki distribusi dengan rata-rata 0 dan standar deviasi 1. Berikut ini hasil *sample* data transformasi data yang sudah dilakukan yang ditunjukkan *sample* data pada Gambar 3 di bawah ini.

	stok_awal	stok_masuk	stok_keluar	stok_akhir
1	-0.36300608	-0.365150118	-0.358063244	-0.311781491
2	-0.36180714	-0.363537992	-0.356931312	-0.307961081
3	-0.36420503	-0.363537992	-0.357308622	-0.310826388
4	0.23526644	0.008540669	-0.006409664	0.661467968
5	-0.35021736	0.199900016	-0.047913842	0.711133299
6	1.63003671	2.594390639	2.477049415	0.455165826
7	-0.31424907	-0.332585174	-0.320709484	-0.285993723
8	-0.34901842	-0.339839741	-0.337436925	-0.285038621
9	0.32598645	0.436398888	0.332415351	0.840072138
10	3.96278001	3.868776176	4.159477861	0.803778243

Showing 1 to 10 of 1,000 entries, 4 total columns

Gambar 3. *Dataset* yang Dinormalisasi

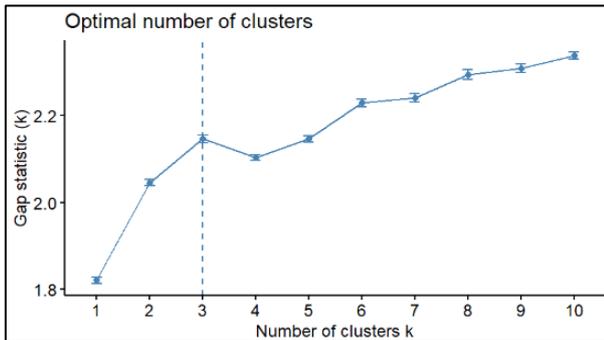
Dari dataset yang sudah dinormalisasi ini, selanjutnya dilakukan proses pencarian optimasi jumlah *k* kluster. Penentuan jumlah kluster pada penelitian ini menggunakan konsep berdasarkan pola pada data yang paling optimal yaitu dengan menggunakan dua metode yaitu metode *Elbow* dan Metode *Gap* Statistik. Tujuan dari proses penentuan jumlah *cluster* ini adalah untuk mencari jumlah *k cluster* yang paling optimal sebelum dilakukan proses *cluster* dengan metode *K-Means*. Proses metode *Elbow* dengan menentukan jumlah *cluster* dengan rentang jumlah *cluster* yang ingin diuji dalam hal ini *k=1* hingga *k=10*. Untuk setiap *k*, dihitung dengan *Within Cluster Sum of Squares* (WCSS) yang merupakan total jarak kuadrat dari setiap titik data ke *centroid* terdekatnya dalam *cluster* tersebut. WCSS berfungsi sebagai ukuran seberapa kompak *cluster* yang dihasilkan. Setelah menghitung WCSS untuk setiap *k*, selanjutnya membuat plot yang menampilkan jumlah cluster (*k*) di sumbu X dan WCSS di sumbu Y. Berikut ini hasil optimasi pencarian jumlah kluster dari metode *Elbow*.



Gambar 4. Optimasi K Klaster Dengan Metode Elbow

Berdasarkan Gambar 4 di atas, bahwa garis grafik menunjukkan grafik yang landai setelah grafik yang curam adalah antara  $k=2$  dan  $k=3$ . Jadi semakin kecil jarak dari *Sum of Squared Errors* (SSE) maka semakin optimal nilai  $k$ . Maka dari hasil ini dapat dikatakan jumlah *cluster* yang optimal adalah  $k = 3$ .

Sedangkan pada metode *Gap Statistik* membandingkan seberapa baik data asli membentuk *cluster* untuk jumlah  $k$  tertentu dibandingkan dengan bagaimana data acak akan membentuk *cluster* untuk jumlah  $k$  yang sama. *Gap Statistic* dihitung dengan mengurangi logaritma WCSS dari data asli dengan rata-rata logaritma WCSS dari data acak. Jika nilai *Gap* tinggi, ini menunjukkan bahwa *cluster* dalam data asli lebih "nyata" dan bukan hanya artefak dari data yang tersebar secara acak. Semakin besar nilai *Gap*, semakin jelas *cluster* dalam data asli, dan nilai  $k$  yang memberikan *Gap* tertinggi dianggap sebagai jumlah *cluster* yang optimal. Berikut ini hasil optimasi pencarian jumlah *cluster*:



Gambar 5. Optimasi K Klaster dengan Metode Gap Statistik

Sedangkan untuk pencarian optimasi dengan menggunakan Metode *Gap Statistik* dapat dilihat sesuai dengan Gambar 5 bahwa jumlah *cluster* optimal menunjukkan jumlah *cluster*  $k = 3$ .

Berdasarkan dari pengujian metode *Elbow* dan *Gap Statistik*, bahwa untuk jumlah  $k$  yang paling optimal memiliki nilai yang sama yaitu  $k=3$ . Maka pada penelitian ini untuk jumlah klaster yang digunakan sejumlah  $k = 3$ .

Setelah mendapatkan jumlah  $k$  optimalnya, selanjutnya dilakukan proses *cluster* dengan menggunakan Metode *K-Means*. Jumlah klaster yang dihasilkan pada klaster 1 sebanyak 41 data obat, klaster 2 sebanyak 116 data obat, dan

klaster 3 sebanyak 1232 data obat. Berikut ini contoh *sample* hasil kluster yang dihasilkan seperti Gambar 6-8 di bawah ini.

Klaster 1:

kode_barang	nama_barang	hasil.clust
00000025	ADALAT OROS -TAB 30 MG	1
00000064	ALLOPURINOL TAB 100 MG	1
00003083	AMLODIPIN 5 MG JKN	1
00000098	AMOXICILLIN TAB 500 MG	1
00000124	ANTASIDA DOEN TAB	1
00000142	AQUABIDEST 25 ML	1
00000176	ASPILETS TAB	1
00000338	BLOOD LANCET 26	1
00000409	CALCIUM LACT TAB	1
00000510	CEFIXIME KAPS 100 MG	1
00000509	CEFIXIME KAPS 200 MG	1
00000576	CETIRIZIN TAB 10 MG	1
00000824	DEKETOPIROFEN TAB	1
00000853	DIGOXIN TAB	1
00000886	DOMPERIDON TAB	1
00003104	GLIMEPIRIDE 2 MG JKN	1
00001308	HANSAPLAST SILVER CARE	1
00001436	ISOSORBID DINITRATE 5 MG	1

Gambar 6. Hasil Kluster 1

Klaster 2 :

kode_barang	nama_barang	hasil.clust
00000009	ACETYLCYSTEINE KAP	2
00000065	ALLOPURINOL TAB 300 MG	2
00000068	ALPRAZOLAM TAB 0.5 MG	2
00000078	AMBROXOL 30 MG TAB	2
00000079	AMINOPHYLLIN TAB	2
00000091	AMITRIPTYLLIN TAB 25 MG	2
00000207	AMLODIPIN 10 MG	2
00000138	APRON	2
00000165	ASAM TRANEKSAMAT TAB 500 MG	2
00003084	BAMGETOL	2
00003543	BETAHISTINE 24 MG	2
00000283	BETAHISTINE 6 MG	2
00003784	BRICLOT 90 MG TAB	2
00000503	CEFADROXIL CAP 500 MG	2
00000521	CEFOTAXIM VIAL 1 GR	2
00000529	CEFTRIAXONE VIAL 1 GR	2
00000600	CILOSTAZOL TAB 100 MG	2
00000609	CIPROFLOXACIN KAPS 500 MG	2

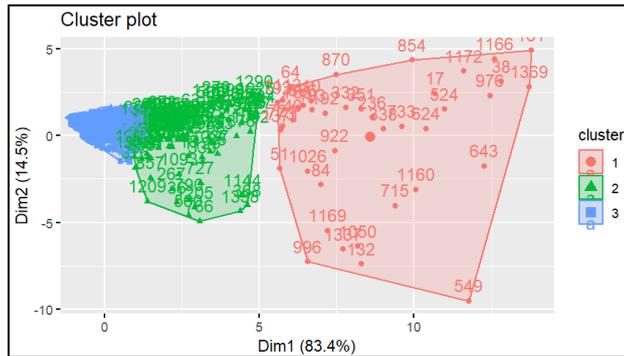
Gambar 7. Hasil Kluster 2

Klaster 3 :

kode_barang	nama_barang	hasil.clust
00000000	ABBOCATH 14	3
00000005	ABBOCATH 16	3
00000001	AB VASK 10 MG TAB	3
00000006	ACARBOSE 100	3
00000007	ACARBOSE 50	3
00000012	ACRAN INJ (RANITIDINE)	3
00000021	ACYCLOVIR CR 5 GR	3
00000023	ACYCLOVIR TAB 400 MG	3
00003262	AKILEN TETES TELINGA	3
00000037	ALA 600	3
00003241	ALBUFORCE CAP	3
00000049	ALGANAX 0,5	3
00000058	ALKOHOL 70 % 20 LT	3
00000061	ALKOHOL 96 % 1 LT	3
00000063	ALLORIS TAB	3
00003292	ALOCLAIR PLUS GEL 8 ML	3
00003685	ALOCLAIR PLUS SPRAY 15 ML	3
00000066	ALVELL 70 MG	3

Gambar 8. Hasil Kluster 3

Adapun hasil kluster yang dihasilkan yang digambarkan dalam bentuk grafik *cluster plot* seperti berikut:



Gambar 9. Hasil Cluster Dengan Metode K-Means.

Berdasarkan hasil *cluster* yang dilakukan (Gambar 9), klaster 1 merupakan golongan obat generik. Klaster 2 merupakan golongan obat paten. Klaster 1 dan 2 bersifat *slow moving* atau tingkat penjualan kurang cepat. Sedangkan untuk klaster 3 merupakan gabungan dari dua jenis golongan obat tersebut yang bersifat *fast moving* yang berarti perputaran penjualannya memiliki waktu relatif cepat.

## V. KESIMPULAN

Proses klusterisasi pada penelitian ini diawali dengan cara melakukan normalisasi atau tranformasi data yang bertujuan untuk memastikan setiap fitur dapat berkontribusi secara proporsional terhadap hasil kluster. Setelah proses tersebut, dilanjutkan mencari optimasi terbaik nilai  $k$  kluster dengan melihat hasil  $k$  optimal dari dua metode yaitu Metode Elbow dan Metode Gap Statistik. Hasil  $k$  optimal yang diperoleh dari kedua metode itu adalah  $k = 3$ , sehingga jumlah kluster yang digunakan pada penelitian ini adalah  $k=3$ .

Hasil dari kluster yang dilakukan pada penelitian ini menghasilkan kluster 1 sebanyak 41 data obat yang menunjukkan golongan obat Generik, kluster 2 sebanyak 116 data obat yang menunjukkan obat Paten. Kedua kluster menunjukkan memiliki tingkat penjualan yang kurang cepat (*slow moving*). Sedangkan pada kluster 3 sebanyak 1232 data obat yang menunjukkan gabungan dari golongan obat generik dan paten yang memiliki tingkat penjualan yang cukup cepat (*fast moving*). Dengan adanya informasi ini unit pengelola data obat dapat mengetahui informasi terkait kecepatan perputaran atau frekuensi penjualan dan pembelian, selain itu juga dapat bermanfaat untuk memastikan ketersediaan obat-obatan *fast moving* yang selalu diperlukan.

## REFERENSI

[1] Â. M. Bagattini, J. L. A. Borges, R. Riera, and D. C. M. F. de Carvalho, "Automation of a tertiary hospital pharmacy drug dispensing system in a lower-middle-income country: A case study and preliminary results," *Explor. Res. Clin. Soc. Pharm.*, vol. 6, no. November 2021, pp. 4–7, 2022, doi: 10.1016/j.rcsop.2022.100151.

[2] H. Supriyanto, "Perbandingan Metode Supervised Learning Untuk Peramalan Time Series Pada Kunjungan

Pasien Rawat jalan," *Junal SimanteC*, vol. 10, no. 2, pp. 67–76, 2022.

- [3] T. Sutarsih, D. Apresziyanti, H. Wulandari, and A. N. Hasyiyati, "Statistik Telekomunikasi Indonesia," *Badan Pus. Stat.*, vol. 1999, no. December, pp. 1–362, 2020.
- [4] T. Nurainun, "Perancangan Sistem Informasi Pengendalian Persediaan Bahan Pokok (Studi Kasus Swalayan Buyung Family Pekanbaru)," *J. Tek. Ind. J. Has. Penelit. dan Karya Ilm. dalam Bid. Tek. Ind.*, vol. 4, no. 2, p. 139, 2020, doi: 10.24014/jti.v4i2.6761.
- [5] D. A. Wulansari, "Faktor Keberhasilan Penanggulangan Tuberculosis Dengan Strategi Dots ( Directly Observed Treatment Shortcourse )," vol. 15, no. 2, pp. 1–8, 2023.
- [6] L. ' Izzah and A. Jananto, "Penerapan Algoritma K-Means Clustering Untuk Perencanaan Kebutuhan Obat Di Klinik Citra Medika," *Progresif J. Ilm. Komput.*, vol. 18, no. 1, p. 69, 2022, doi: 10.35889/progresif.v18i1.769.
- [7] M. R. Nugroho, I. E. Hendrawan, and P. P. Purwanto, "Penerapan Algoritma K-Means Untuk Klasterisasi Data Obat Pada Rumah Sakit ASRI," *Nuansa Inform.*, vol. 16, no. 1, pp. 125–133, 2022, doi: 10.25134/nuansa.v16i1.5294.
- [8] Holwati, E. Widodo, and W. Hadikristanto, "Pengelompokan Untuk Penjualan Obat Dengan Menggunakan Algoritma K-Means," *Bull. Inf. Technol. ...*, vol. 4, no. 2, pp. 408–413, 2023, [Online]. Available: <https://journal.fkpt.org/index.php/BIT/article/view/848> %0Ahttps://journal.fkpt.org/index.php/BIT/article/download/848/468
- [9] M. N. Susanto and V. Y. Permanasari, "Penerapan Metode ABC Indeks Kritis dalam Pengelolaan Persediaan Obat di Instalasi Farmasi Rumah Sakit XYZ Pekanbaru, Riau Tahun 2018," *J. Adm. Rumah Sakit Indones.*, vol. 5, no. 2, pp. 72–84, 2019, doi: 10.7454/arsi.v5i2.3196.
- [10] V. Andita, W. Hermawat, and N. S. Hartati, "Pengaruh Jumlah Pelayanan Rawat Jalan, Rawat IGD Dan Rawat Inap Terhadap Tingkat Pendapatan Rumah Sakit Umum Daerah (RSUD) Cideres Kabupaten Majalengka," *J. Manaj. Sekol. Tinggi Ekon. Cirebon*, vol. 14, no. 2, pp. 370–378, 2019.
- [11] Yuda Irawan, "Penerapan Data Mining Untuk Evaluasi Data Penjualan Menggunakan Metode Clustering Dan Algoritma Hirarki Divisive Di Perusahaan Media World Pekanbaru," *J. Teknol. Inf. Univ. Lambung Mangkurat*, vol. 4, no. 1, pp. 13–20, 2019, doi: 10.20527/jtiulm.v4i1.34.
- [12] K. P. Sinaga and M. S. Yang, "Unsupervised K-Means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.
- [13] F. Gorunescu, *Data Mining Concepts, Models and Techniques*. Springer-Verlag Berlin Heidelberg, 2011.
- [14] F. Gullo, "From patterns in data to knowledge discovery: What data mining can do," *Phys. Procedia*, vol. 62, pp. 18–22, 2015, doi: 10.1016/j.phpro.2015.02.005.
- [15] A. Zaki, I. Irwan, and I. A. Sembe, "Penerapan K-Means Clustering dalam Pengelompokan Data (Studi Kasus Profil Mahasiswa Matematika FMIPA UNM)," *J. Math.*

- Comput. Stat.*, vol. 5, no. 2, p. 163, 2022, doi: 10.35580/jmathcos.v5i2.38820.
- [16] D. Das, P. Kayal, and M. Maiti, "A *K-Means* clustering model for analyzing the Bitcoin extreme value returns," *Decis. Anal. J.*, vol. 6, no. June 2022, p. 100152, 2023, doi: 10.1016/j.dajour.2022.100152.
- [17] M. Butwall, "Data Normalization and Standardization: Impacting Classification Model Accuracy," *Int. J. Comput. Appl.*, vol. 183, no. 35, pp. 6–9, 2021, doi: 10.5120/ijca2021921669.
- [18] P. Bholowalia and A. Kumar, "EBK-Means: A Clustering Technique based on Elbow Method and *K-Means* in WSN," *Int. J. Comput. Appl.*, vol. 105, no. 9, pp. 975–8887, 2014.
- [19] R. A. Febrianty, W. Witanti, and P. N. Sabrina, "Segmentasi Penjualan Obat Di Apotek Menggunakan Metode *K-Means*," *Pros. Semin. Nas. Sist. Inf. dan Teknol.*, vol. 4, pp. 200–206, 2020.
- [20] R. Nainggolan, R. Perangin-Angin, E. Simarmata, and A. F. Tarigan, "Improved the Performance of the *K-Means* Cluster Using the Sum of Squared Error (SSE) optimized by using the Elbow Method," *J. Phys. Conf. Ser.*, vol. 1361, no. 1, 2019, doi: 10.1088/1742-6596/1361/1/012015.
- [21] M. Mohajer, K.-H. Englmeier, and V. J. Schmid, "A comparison of Gap statistic definitions with and without logarithm function," 2011, [Online]. Available: <http://arxiv.org/abs/1103.4767>
- [22] A. Hasibuan, D. R. Kembuan, C. T. M. Manoppo, and M. T. Hermanto, "Optimization of *K-Means* algorithm in grouping data using the statistical gap method," *J. Intell. Decis. Support Syst.*, vol. 6, no. 3, pp. 112–120, 2023.
- [23] M. Benri, H. Metisen, and S. Latipa, "Analisis Clustering Menggunakan Metode *K-Means* Dalam Pengelompokan Penjualan Produk Pada Swalayan Fadhila," *J. Media Infotama*, vol. 11, no. 2, pp. 110–118, 2015, [Online]. Available: <https://core.ac.uk/download/pdf/287160954.pdf>
- [24] P. Anitha and M. M. Patil, "RFM model for customer purchase behavior using *K-Means* algorithm," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 5, pp. 1785–1792, 2022, doi: 10.1016/j.jksuci.2019.12.011.
- [25] S. Handoko, F. Fauziah, and E. T. E. Handayani, "Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode *K-Means* Clustering," *J. Ilm. Teknol. dan Rekayasa*, vol. 25, no. 1, pp. 76–88, 2020, doi: 10.35760/tr.2020.v25i1.2677.