VOL. 3. NO. 04 JULI 2026

JOURNAL OF COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE (JCSAI)

ANALISIS SEGMENTASI PRODUK MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING BERDASARKAN DATA PENJUALAN OBAT BULANAN PADA APOTEK PERJUANGAN

Saepul Huda Mubarok[1], Nining Rahaningsih[2], Irfan Ali[3], Indra Wiguna Marthanu[4].

Program Studi Teknik Informatika[1]
Program Studi Komputerisasi Akuntansi[24]
Program Studi Rekayasa Perangkat Lunak[3]

STMIK IKMI Cirebon
https://ikmi.ac.id/page/18/?lang=de
saepulhudamubarok574@gmail.com

(*) Corresponding Author : saepulhudamubarok574@gmail.com
Published : 30 Juni 2026

Abstract—*This study analyzes product segmentation at Apotek Perjuangan using the K-Means clustering algorithm based on monthly drug sales data. The primary objective is to identify product groups (fast-moving, mid-moving, slow-moving) to support stock management decisions, procurement priorities, and cost efficiency. The data used consists of attributes such as quantity sold, unit price, and total sales; the pre-processing stage includes data cleaning, type conversion, and feature standardization (StandardScaler). The optimal number of clusters was determined using the Elbow Method (WCSS) and the Silhouette Coefficient; the evaluation results showed that K = 3 provided the highest silhouette score (0.8934) and was therefore selected as the optimal value. The implementation of K-Means (with K-Means++ initialization) resulted in three clusters that can be interpreted as low-, medium-, and high-turnover products. Empirical findings show an inverse relationship between unit price and sales volume—lower-priced products tend to have a greater sales contribution—driving recommendations for different procurement strategies based on segments. Practical implications of this study include prioritizing procurement for fast-moving products, reducing purchases for slow-moving products, and designing stock policies that reduce the risk of shortages and overstocks, thus reducing storage costs. Limitations of the study include the use of dummy data and a small sample size; therefore, suggestions for development include implementation on a larger real-world dataset, comparison with other algorithms (e.g., DBSCAN, FCM), and integration of temporal and geographic aspects for longitudinal analysis. Overall, this study confirms that K-Means is an effective and practical tool for pharmaceutical product segmentation that can improve the accuracy of data-driven inventory decision-making.*

Keywords : *K-Means Clustering, Product Segmentation, Inventory Management, Drug Sales Data*

Abstrak—Penelitian ini menganalisis segmentasi produk pada Apotek Perjuangan menggunakan algoritma K-Means clustering berdasarkan data penjualan obat bulanan. Tujuan utama adalah mengidentifikasi kelompok produk (fast-moving, mid-moving, slow-moving) untuk mendukung keputusan pengelolaan stok, prioritas pengadaan, dan efisiensi biaya. Data yang digunakan terdiri dari atribut jumlah terjual, harga satuan, dan total penjualan; tahap pra-pemrosesan meliputi pembersihan data, konversi tipe, dan standardisasi fitur (StandardScaler). Penentuan jumlah cluster optimal dilakukan dengan Elbow Method (WCSS) dan Silhouette Coefficient; hasil evaluasi menunjukkan K = 3 memberikan skor silhouette tertinggi (0.8934) sehingga dipilih sebagai nilai optimal. Implementasi K-Means (dengan inisialisasi K-Means++) menghasilkan tiga klaster yang dapat diinterpretasikan sebagai produk berperputaran rendah, sedang, dan tinggi. Temuan empiris memperlihatkan pola hubungan invers antara harga satuan dan volume penjualan produk berharga lebih rendah cenderung memiliki kontribusi penjualan lebih besar—mendorong rekomendasi strategi pengadaan berbeda berdasarkan segmen. Implikasi praktis penelitian ini mencakup prioritas pengadaan untuk produk fast-moving, pengurangan pembelian untuk slow-moving, dan desain kebijakan stok yang mengurangi risiko kekurangan maupun overstock sehingga menekan biaya penyimpanan. Keterbatasan penelitian meliputi penggunaan data dummy dan ukuran sampel kecil; oleh karena itu saran pengembangan mencakup penerapan pada dataset nyata yang lebih besar, perbandingan dengan algoritma lain (mis. DBSCAN, FCM), serta integrasi aspek temporal dan geografis untuk analisis longitudinal. Secara keseluruhan, studi ini menegaskan bahwa K-Means merupakan alat yang efektif dan

praktis untuk segmentasi produk farmasi yang dapat meningkatkan ketepatan pengambilan keputusan inventori berbasis data.

Kata Kunci : K-Means Clustering, Segmentasi Produk, Manajemen Inventori, Data Penjualan Obat

INTRODUCTION

Manajemen persediaan obat merupakan salah satu aspek krusial dalam rantai pasok farmasi yang berpengaruh langsung terhadap ketersediaan layanan kesehatan. Ketidakefisienan dalam pengelolaan stok dapat menyebabkan kekurangan obat (*drug shortages*) yang berdampak serius terhadap keselamatan pasien dan efisiensi operasional rumah sakit [1]. Pengoptimalan sistem inventori melalui pendekatan analitik dan teknologi menjadi kebutuhan strategis dalam menjamin keberlanjutan pasokan serta menekan biaya penyimpanan. Dalam konteks tersebut, simulasi dan optimasi berbasis data telah terbukti mampu menurunkan biaya inventori di apotek rawat jalan melalui prediksi permintaan yang lebih akurat [2]

Perkembangan teknologi informasi, khususnya dalam bidang kecerdasan buatan dan *machine learning*, membuka peluang baru dalam pengelolaan rantai pasok farmasi. Model pembelajaran mesin telah digunakan secara luas untuk melakukan *demand forecasting* yang lebih adaptif terhadap dinamika pasar dan fluktuasi kebutuhan konsumen [3] Pendekatan ini memperkuat peran sistem prediktif dalam meningkatkan ketepatan perencanaan produksi dan distribusi produk farmasi [4] Sejalan dengan itu, *predictive big data analytics* memberikan kemampuan untuk mengidentifikasi pola permintaan yang kompleks, sehingga mendukung keputusan strategis di sepanjang rantai pasok [5].

Dalam lingkungan farmasi modern, penerapan teknologi informasi yang memanfaatkan *data-driven analytics* telah menjadi pendorong utama efisiensi operasional Sistem analitik berbasis data besar tidak hanya memungkinkan pengelolaan stok secara real-time, tetapi juga memperkuat budaya pengambilan keputusan berbasis bukti di tingkat operasional farmasi Dengan demikian, integrasi *Health Information Technology (HIT)* dan analisis prediktif memberikan kontribusi signifikan terhadap ketahanan rantai pasok farmasi [6].

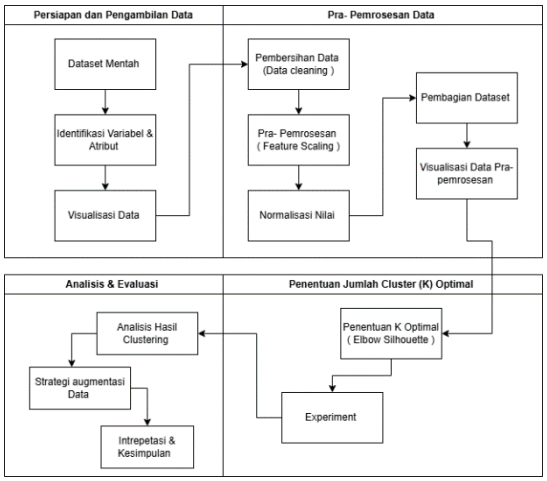
Salah satu pendekatan yang banyak diterapkan dalam pengelolaan data farmasi adalah algoritma *K-Means clustering*, yang berfungsi untuk mengelompokkan data transaksi atau pola resep berdasarkan kesamaan karakteristik (Dermawan et al., 2024). Melalui teknik klasterisasi ini, rumah sakit dan apotek dapat mengidentifikasi kategori obat dengan tingkat perputaran tertentu, sehingga dapat merumuskan strategi pengadaan yang lebih efisien (Fitriyani et al., 2024). Pendekatan serupa juga terbukti efektif dalam menentukan kebutuhan produk ritel farmasi berbasis data penjualan historis [7][i].

Selain efisiensi operasional, klasterisasi dengan algoritma *K-Means* juga berperan dalam mendukung strategi pemasaran farmasi berbasis segmentasi pelanggan [8]. Dengan mengintegrasikan data transaksi pelanggan, perusahaan dapat mengembangkan pendekatan personalisasi penjualan serta memperkirakan kebutuhan obat berdasarkan pola konsumsi konsumen. Oleh karena itu, penelitian terkini berfokus pada pengembangan model prediktif dan klasterisasi yang dapat meningkatkan ketepatan perencanaan persediaan serta optimalisasi rantai pasok farmasi secara menyeluruh [9].

Berdasarkan berbagai penelitian terdahulu, terlihat bahwa integrasi *machine learning* dan algoritma klasterisasi berbasis data merupakan langkah strategis untuk menciptakan sistem manajemen inventori farmasi yang cerdas dan berkelanjutan. Dengan memanfaatkan pendekatan tersebut, organisasi kesehatan dapat meningkatkan efisiensi biaya, mengurangi risiko kekurangan obat, serta memastikan distribusi obat yang tepat waktu dan sesuai kebutuhan pasien. Oleh karena itu, topik ini menjadi sangat relevan untuk dikaji lebih lanjut dalam konteks penelitian akademik, khususnya pada bidang Teknik Informatika dan Sistem Informasi yang berorientasi pada analisis data dan optimasi operasional.

MATERIALS AND METHODS

Alur metodologi penelitian dalam pengolahan data berbasis *clustering*. Proses ini dimulai dari tahap persiapan hingga evaluasi hasil untuk memastikan bahwa model yang dibangun mampu menghasilkan pengelompokan data yang optimal. Setiap tahapan memiliki peran penting, mulai dari memahami data awal, melakukan pembersihan, hingga menentukan parameter terbaik dalam proses *clustering*. Alur ini dirancang secara sistematis agar hasil analisis dapat dipercaya dan memiliki nilai interpretasi yang baik.



Gambar 1. Alur Penelitian

Berdasarkan gambar tersebut, proses dimulai dari tahap Persiapan dan Pengambilan Data, di mana dataset mentah dikumpulkan, kemudian dilakukan identifikasi variabel dan atribut yang relevan, serta visualisasi awal untuk memahami karakteristik data. Selanjutnya masuk ke tahap Pra**-**Pemrosesan Data, yang meliputi pembersihan data (*data cleaning*), *feature scaling*, dan normalisasi nilai agar data siap digunakan dalam proses analisis.

Setelah data bersih, dilakukan Pembagian Dataset dan visualisasi ulang untuk memastikan kualitas data sebelum pemodelan. Tahap berikutnya adalah Penentuan Jumlah Cluster **(**K**)** Optimal menggunakan metode seperti *Elbow* dan *Silhouette*, yang bertujuan menentukan jumlah kelompok terbaik. Hasil tersebut kemudian diuji melalui tahap Eksperimen.

Terakhir, pada tahap Analisis dan Evaluasi, dilakukan analisis terhadap hasil clustering, penerapan strategi augmentasi data jika diperlukan, serta interpretasi hasil untuk menghasilkan kesimpulan penelitian. Keseluruhan alur ini menunjukkan pendekatan yang terstruktur dalam membangun model clustering yang akurat dan dapat diandalkan.

RESULTS AND DISCUSSION

Tahapan analisis ini dilakukan untuk menentukan jumlah *cluster* (*k*) terbaik dalam algoritma *K-Means* dengan mempertimbangkan dua metrik utama, yaitu WGSS (Within-Group Sum of Squares / Inertia) dan *Silhouette Score*. Pertama, proses dimulai dengan menjalankan model K-Means menggunakan berbagai nilai *k* mulai dari 2 hingga 8. Untuk setiap percobaan, dihitung nilai WGSS yang menunjukkan seberapa rapat data dalam setiap cluster. Semakin kecil WGSS, semakin baik struktur cluster karena data dalam satu kelompok semakin mirip. Selanjutnya, dihitung juga *Silhouette Score* untuk menilai kualitas pemisahan antar-cluster. Nilai silhouette yang mendekati 1 menunjukkan bahwa setiap data berada pada cluster yang tepat dan terpisah baik dari cluster lainnya. Tahapan ini membantu dalam mengevaluasi kombinasi *k* mana yang menghasilkan cluster paling optimal. Setelah semua nilai dihitung, hasilnya dibandingkan untuk menentukan titik optimal berdasarkan keseimbangan antara penurunan WGSS dan kenaikan Silhouette Score.

Dari hasil evaluasi pada tabel, *k = 2* menghasilkan WGSS sebesar 4.001182 dan Silhouette Score 0.683298, menunjukkan struktur cluster yang cukup baik namun belum optimal. Ketika *k = 3*, terjadi penurunan WGSS yang signifikan menjadi 0.116085 dan Silhouette Score meningkat menjadi 0.893381 — nilai tertinggi dalam seluruh percobaan. Hal ini menunjukkan bahwa tiga cluster memberikan pemisahan terbaik. Namun, ketika *k = 4*, meskipun WGSS turun sedikit menjadi 0.077447, nilai Silhouette Score justru menurun drastis ke 0.656135, menunjukkan bahwa penambahan cluster tidak meningkatkan kualitas pemisahan data. Pada *k = 5*, WGSS mencapai angka sangat kecil (mendekati nol), tetapi Silhouette Score kembali turun menjadi 0.419739, menandakan overlap antar-cluster semakin besar. Nilai *k* lebih tinggi, yaitu 6 sampai 8, memang membuat WGSS semakin kecil tetapi silhouette terus turun ke 0.364183, 0.308628, hingga 0.103704, sehingga kualitas cluster semakin buruk karena data menjadi terlalu terpecah.

Secara keseluruhan, meskipun WGSS terus menurun saat jumlah cluster meningkat, Silhouette Score menunjukkan bahwa *k = 3* adalah pilihan terbaik karena memberikan pemisahan cluster yang paling jelas, stabil, dan logis. Dengan demikian, penggunaan tiga cluster adalah konfigurasi paling optimal berdasarkan data evaluasi tersebut.

Sebelum melakukan proses clustering, langkah penting yang harus dilakukan adalah menentukan jumlah cluster optimal (K). Pada penelitian ini digunakan dua metode evaluasi, yaitu:

1. *Within-Cluster Sum of Squares* (WCSS / Inertia**)** untuk metode *Elbow*
2. *Silhouette Coefficient* untuk mengukur kualitas pemisahan cluster

Kedua metode digunakan secara bersamaan untuk memastikan bahwa pemilihan jumlah *cluster* lebih akurat dan tidak hanya bergantung pada satu indikator evaluasi. Tabel Ringkasan Kualitas *Cluster* seperti ditunjukan pada tabel 1

Tabel 1 Ringkasan Kualitas *Cluster*

	k	K WGSS [Inertia]	Silhouette Score
0	2	2 4.001182	0.683298
1	3	3 0.116085	0.893381
2	4	4 0.077447	0.656135
3	5	5 0 0.8809	0.419739
4	6	6 0.025930	0.364183
5	7	7 0.013051	0.308628
6	8	8 0.002061	0.103704

Proses analisis dilakukan pada data yang telah distandardisasi (X_scaled). Berikut ringkasan hasil perhitungan WCSS dan Silhouette Score:

1) Nilai WCSS menurun seiring bertambahnya jumlah cluster, sebagaimana karakteristik K-Means.
2) Namun, penurunan WCSS yang terlalu kecil pada K > 3 menunjukkan tidak adanya peningkatan signifikan dalam pemisahan kelompok.
3) Silhouette Coefficient menunjukkan skor tertinggi pada K = 3, yaitu **0.**8934**,** menandakan kualitas cluster yang sangat baik.
4) Skor silhouette pada K > 3 cenderung menurun tajam, yang berarti cluster menjadi kurang terpisah dan kurang kompak.

Berdasarkan kedua indikator tersebut:

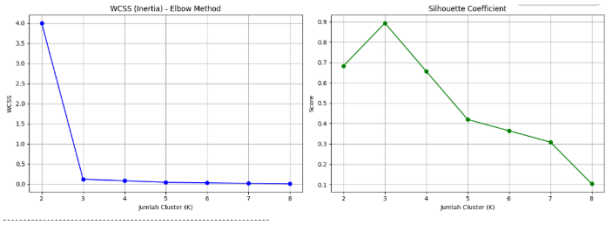
1. K optimal adalah 3, karena memiliki nilai *Silhouette Score* tertinggi (0.8934) dan menghasilkan pemisahan cluster terbaik.
2. Dengan demikian, penelitian ini menggunakan K = 3 pada tahap implementasi algoritma *K-Means*

Hasil analisis penentuan jumlah cluster optimal menunjukkan bahwa metode Elbow dan Silhouette memberikan indikasi yang konsisten terhadap pemilihan nilai K. Pada grafik *Elbow*, terjadi penurunan WCSS yang sangat tajam dari K=2 ke K=3, sementara setelah K=3 penurunan nilai WCSS menjadi sangat kecil, menandakan titik "siku" berada pada K=3. Hal ini diperkuat oleh grafik *Silhouette Coefficient*, yang menunjukkan skor tertinggi pada K=3 dengan nilai sekitar 0.89, mencerminkan kualitas pemisahan cluster yang sangat baik dan struktur kelompok yang paling jelas dibandingkan jumlah cluster lainnya. Pada K > 3, skor *Silhouette* terus menurun, menunjukkan bahwa menambah jumlah cluster justru membuat hasil clustering menjadi kurang efektif. Dengan demikian, kedua grafik tersebut secara kuat menegaskan bahwa jumlah *cluster* yang paling optimal adalah K=3, karena mampu menghasilkan pemisahan kelompok yang paling jelas dan stabil.

Tahapan ini dilakukan untuk menentukan jumlah cluster (*k*) terbaik dalam algoritma K-Means Clustering. Langkah pertama adalah menjalankan K-Means dengan berbagai nilai *k* mulai dari 2 hingga 8. Untuk setiap nilai *k*, dihitung dua metrik evaluasi: WCSS *(Within-Cluster Sum of Squares / Inertia)* dan Silhouette Score. WCSS mengukur seberapa rapat data bekerja dalam cluster—semakin kecil nilainya, semakin baik. Sementara itu, Silhouette Score menilai kualitas pemisahan antar-cluster—semakin mendekati 1, semakin ideal. Dengan menggunakan kedua pendekatan ini, analisis dilakukan untuk mengidentifikasi titik *k* yang memberikan keseimbangan terbaik antara kompaknya cluster dan keterpisahan antar-cluster. Hasil perhitungan kemudian divisualisasikan melalui grafik metode Elbow dan grafik Silhouette untuk memperoleh keputusan jumlah cluster optimal.

Grafik sebelah kanan memperlihatkan nilai *Silhouette* Score untuk setiap jumlah cluster. Nilai tertinggi muncul pada k = 3 dengan skor mendekati 0.89, yang menunjukkan bahwa struktur cluster sangat baik, terpisah dengan jelas, dan stabil.

Setelah *k = 3*, nilai Silhouette Score menurun secara konsisten (k = 4 ke bawah), menunjukkan bahwa penambahan cluster justru memperburuk kualitas pemisahan data. Bahkan pada k = 8 nilai silhouette sangat rendah (sekitar 0.10), menandakan cluster menjadi kurang bermakna dan saling tumpang tindih.



Gambar 3 Analisis Evaluasi Cluster

Hasil analisis statistik menunjukkan bahwa proses clustering berhasil membagi data penjualan menjadi tiga kelompok dengan karakteristik berbeda. Cluster 1 merupakan kelompok dengan performa terbaik, ditandai oleh jumlah terjual yang tinggi (rata-rata 105 unit) dan nilai penjualan terbesar, yaitu sekitar 1.550.000, sehingga menggambarkan produk dengan permintaan paling kuat. Cluster 0 berada pada kategori menengah, dengan rata-rata penjualan 55 unit dan total penjualan sekitar 850.000, menunjukkan performa yang stabil namun tidak setinggi cluster terbaik. Sementara itu, Cluster 2 adalah kelompok dengan performa terendah, karena hanya memiliki rata-rata penjualan sekitar 12 unit dengan total

penjualan 223.333, sehingga mencerminkan produk dengan minat pasar paling rendah dan perlu perhatian lebih dalam strategi pemasaran atau pengembangan.

Hasil analisis statistik setelah proses *clustering* menggunakan algoritma K-Means menunjukkan tiga kelompok data yang memiliki karakteristik penjualan berbeda. Analisis dilakukan dengan menghitung nilai rata-rata, median, dan maksimum untuk dua variabel utama, yaitu jumlah terjual dan total penjualan. Pendekatan ini bertujuan memahami performa setiap cluster sehingga dapat diketahui kelompok produk mana yang memiliki kinerja tinggi, sedang, atau rendah.

Cluster pertama yang muncul adalah Cluster 1, yang menunjukkan performa penjualan paling tinggi. Rata-rata jumlah produk yang terjual mencapai 105 unit, dengan median yang sama dan nilai maksimum sebesar 110 unit. Nilai ini sejalan dengan pencapaian total penjualan yang juga paling besar, yaitu median Rp1.550.000 dan maksimum Rp1.600.000. Hal ini menegaskan bahwa cluster ini terdiri dari produk-produk dengan permintaan sangat tinggi dan kontribusi penjualan terbesar, sehingga layak dijadikan prioritas dalam strategi bisnis dan pengelolaan stok.

Selanjutnya, Cluster 0 menunjukkan kategori produk dengan performa menengah. Jumlah penjualan rata-rata sebesar 55 unit dan nilai maksimum 60 unit, menunjukkan distribusi penjualan yang stabil. Pendapatan yang dihasilkan juga berada pada level menengah dengan median Rp850.000 dan maksimum Rp900.000. Produk pada cluster ini memiliki potensi untuk ditingkatkan melalui strategi promosi atau peningkatan visibilitas agar dapat naik ke kategori high performer.

Sementara itu, Cluster 2 berisi produk dengan performa terendah. Rata-rata jumlah terjual hanya sekitar 12 unit, dengan nilai maksimum 15 unit. Total penjualan juga jauh lebih rendah dibandingkan cluster lainnya, yaitu median Rp220.000 dan maksimum Rp250.000. Cluster ini menggambarkan produk dengan permintaan rendah dan kontribusi pendapatan yang kecil. Produk dalam kelompok ini perlu dilakukan evaluasi lebih lanjut, seperti peninjauan strategi pemasaran, penempatan produk, atau kemungkinan penyesuaian stok.

Secara keseluruhan, analisis ini menunjukkan segmentasi produk yang jelas berdasarkan performa penjualan. Cluster 1 mewakili produk unggulan, Cluster 0 menunjukkan performa menengah yang stabil, sementara Cluster 2 menggambarkan produk berkinerja rendah. Hasil ini sangat berguna untuk pengambilan keputusan strategis, seperti pengelolaan stok, pengalokasian anggaran promosi, hingga prioritas bisnis dalam pengembangan produk.

```
--- Analisis Statistik Cluster ---
        Jumlah Terjual                          Total Penjualan                \
                 count        mean median  max           count          mean
Cluster
0                    3   55.000000   55.0   60               3  8.500000e+05
1                    3  105.000000  105.0  110               3  1.550000e+06
2                    3   12.333333   12.0   15               3  2.233333e+05


            median      max
Cluster
0         850000.0   900000
1        1550000.0  1600000
2         220000.0   250000
```

Gambar 3 Analisis statistik cluster

Tahapan ini dilakukan setelah jumlah cluster optimal ditentukan menggunakan metode Elbow dan Silhouette, yang menunjukkan bahwa K = 3 adalah nilai terbaik. Proses clustering dimulai dengan mengekstraksi dua variabel penting, yaitu Jumlah Terjual dan Total Penjualan, kemudian algoritma K-Means mengelompokkan produk berdasarkan kesamaan pola pada dua variabel tersebut. Setelah cluster terbentuk, setiap data diberi label cluster 0, 1, atau 2. Untuk memvisualisasikan hasilnya, dibuat grafik scatter plot yang menunjukkan persebaran data penjualan berdasarkan jumlah terjual pada sumbu X dan total penjualan pada sumbu Y. Selain itu, centroid atau titik pusat cluster ditampilkan sebagai tanda silang (X) berukuran besar untuk menggambarkan posisi rata-rata setiap cluster. Tahap visualisasi ini penting untuk memverifikasi apakah hasil clustering sudah memisahkan grup data dengan baik dan sesuai dengan logika bisnis.

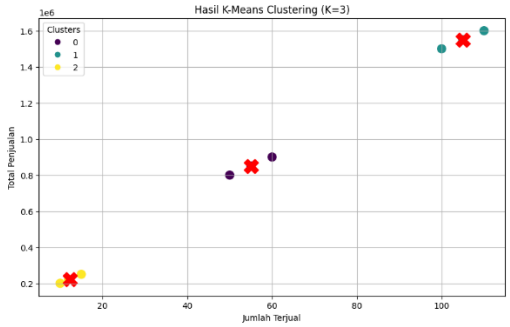
Grafik menunjukkan tiga cluster yang terbentuk dari analisis K-Means. Setiap cluster diwakili oleh warna berbeda: ungu untuk cluster 0, hijau kebiruan untuk cluster 1, dan kuning untuk cluster 2. Titik-titik kecil pada grafik mewakili data penjualan tiap produk, sedangkan tanda silang merah berukuran besar menunjukkan posisi centroid masing-masing cluster. Data dikelompokkan berdasarkan nilai jumlah terjual (sumbu X) dan total penjualan (sumbu Y), sehingga terlihat jelas pola pemisahan antar cluster.

Cluster 1 (warna hijau) berada pada bagian kanan atas grafik, menunjukkan kelompok produk dengan jumlah terjual tinggi (sekitar 100–110 unit) dan total penjualan sangat besar (sekitar 1,5–1,6 juta**)**. Cluster ini menggambarkan produk berperforma tinggi baik dari sisi volume maupun pendapatan. Sementara itu, cluster 0 (warna ungu) berada pada posisi tengah grafik dengan jumlah terjual sekitar 50–60 unit dan total penjualan sekitar 800–900 ribu. Cluster ini berisi produk

dengan performa menengah, stabil namun tidak setinggi cluster 1.

Cluster 2 (warna kuning) menempati bagian kiri bawah grafik, dengan jumlah terjual rendah di kisaran 12–15 unit dan total penjualan sekitar 200–250 ribu. Produk dalam cluster ini termasuk kelompok berperforma rendah, sehingga memerlukan evaluasi strategi promosi atau manajemen stok. Jarak antar cluster yang terlihat jelas pada grafik menunjukkan bahwa algoritma K-Means berhasil memisahkan data ke dalam tiga kelompok yang berbeda secara logis dan konsisten dengan hasil analisis statistik sebelumnya.

Grafik hasil *K-Means Clustering* dengan K=3 menunjukkan bahwa data penjualan terbagi menjadi tiga kelompok yang jelas berdasarkan hubungan antara jumlah terjual dan total penjualan. Cluster 1 menempati posisi paling atas dengan jumlah terjual tinggi (sekitar 100–110 unit) dan total penjualan terbesar, mencapai 1,5 juta, sehingga mewakili produk berperforma terbaik. Cluster 0 berada di tengah grafik dengan jumlah terjual menengah (50–60 unit) dan total penjualan sekitar 800–900 ribu, menggambarkan produk yang stabil namun tidak dominan. Di sisi lain, Cluster 2 terletak di bagian bawah grafik dengan jumlah terjual sangat rendah (10–15 unit) dan total penjualan sekitar 200–250 ribu, menandakan kelompok produk dengan kinerja paling rendah. Visualisasi ini mempertegas perbedaan pola penjualan antarcluster dan membantu dalam menentukan strategi pemasaran atau pengelolaan produk sesuai performanya masing-masing.



Gambar 4 Grafik hasil K-Means Clustering

Tahapan ini dilakukan setelah proses *clustering* dengan algoritma K-Means selesai dan tiga cluster berhasil terbentuk. Setelah setiap produk mendapatkan label cluster, langkah selanjutnya adalah menyusun ringkasan anggota cluster untuk mengetahui produk mana yang termasuk dalam setiap kelompok. Tahapan ini mencakup penggabungan kembali data asli (nama produk, jumlah terjual, dan total penjualan) dengan hasil cluster sehingga dapat terlihat sebaran produk berdasarkan performa penjualannya. Analisis ringkasan anggota cluster sangat penting karena memberikan gambaran nyata mengenai karakteristik setiap produk dalam masing-masing cluster, apakah termasuk kategori penjualan tinggi, menengah, atau rendah. Informasi ini menjadi dasar untuk menyusun strategi bisnis, pengendalian stok, ataupun rekomendasi penelitian.

Gambar menampilkan daftar produk yang telah dikelompokkan ke dalam tiga cluster berdasarkan pola jumlah terjual dan total penjualan. Pada Cluster 0, terdapat tiga produk yaitu Prod D, Prod E, dan Prod F, masing-masing memiliki jumlah terjual antara 50 hingga 60 unit, dengan total penjualan berada pada kisaran 800.000 hingga 900.000 rupiah. Hal ini menunjukkan bahwa cluster ini berisi produk dengan performa penjualan menengah, stabil dalam jumlah penjualan dan memiliki kontribusi pendapatan yang cukup besar namun tidak setinggi cluster tertinggi.

Selanjutnya, Cluster 1 berisi produk dengan performa penjualan paling tinggi, yaitu Prod G, Prod H, dan Prod I. Ketiga produk ini memiliki jumlah terjual antara 100 hingga 110 unit, dan total penjualan mencapai 1.500.000 hingga 1.600.000 rupiah. Cluster ini menggambarkan kategori produk high performer, yaitu produk yang memiliki permintaan paling besar dan memberikan kontribusi pendapatan tertinggi bagi perusahaan. Keberadaan produk dalam cluster ini sangat penting dan perlu menjadi prioritas dalam strategi pemasaran maupun pengelolaan stok agar tidak terjadi kekurangan persediaan.

Sementara itu, produk dengan performa paling rendah tergabung dalam Cluster 2, yaitu Prod A, Prod C, dan Prod B. Produk-produk ini hanya terjual sebanyak 10 hingga 15 unit dengan total penjualan berkisar antara 200.000 hingga 250.000 rupiah. Cluster ini menunjukkan kelompok produk low performer, di mana permintaan terhadap produk relatif kecil. Produk dalam cluster ini perlu ditinjau lebih lanjut, apakah rendahnya penjualan disebabkan oleh kurangnya promosi, daya saing harga, atau memang tidak banyak diminati oleh konsumen.

Secara keseluruhan, ringkasan anggota cluster memberikan gambaran yang jelas mengenai segmentasi produk berdasarkan performa penjualan. Cluster 1 menempati posisi sebagai kelompok produk terbaik, cluster 0 berada di kategori menengah, sementara cluster 2 merupakan kelompok produk dengan kinerja terendah. Informasi ini sangat bermanfaat untuk mengambil keputusan strategis, seperti prioritas

stok, pengembangan produk, hingga kebijakan pemasaran yang lebih tepat sasaran.

CONCLUSION

Berdasarkan hasil penelitian yang telah dilakukan mengenai penerapan algoritma K-Means clustering untuk mengelompokkan data penjualan obat di Apotek Perjuangan,

Algoritma K-Means dapat diterapkan secara efektif untuk mengelompokkan data transaksi obat setelah melalui tahapan metodologis yang sistematis, mulai dari persiapan data, pembersihan data, pra-pemrosesan (feature scaling), penentuan jumlah cluster optimal, hingga implementasi model. Penggunaan data dummy memungkinkan seluruh prosedur penelitian dijalankan sesuai metodologi CRISP-DM.

Hasil evaluasi penentuan jumlah cluster menunjukkan bahwa K = 3 merupakan jumlah cluster paling optimal, ditunjukkan oleh nilai Silhouette Coefficient tertinggi sebesar 0.8934, yang menandakan kualitas pemisahan cluster sangat baik. Hal ini membuktikan bahwa data memiliki struktur alami yang mendukung pembentukan tiga kelompok utama.

Hasil klasterisasi menghasilkan tiga kelompok produk obat, yaitu Cluster 0 (Slow-Moving Item): produk dengan jumlah penjualan rendah dan kontribusi pendapatan kecil. Cluster 1 (Medium-Moving Item): produk dengan pola penjualan stabil dan tingkat perputaran sedang. Cluster 2 (Fast-Moving Item): produk dengan volume penjualan tinggi serta kontribusi besar terhadap total pendapatan apotek Segmentasi ini memberikan gambaran yang jelas mengenai tingkat perputaran stok dan karakteristik produk.

Hasil klasterisasi terbukti mendukung pengambilan keputusan dalam manajemen persediaan obat, terutama dalam menentukan prioritas pengadaan, optimasi jumlah stok, dan pengendalian biaya penyimpanan. Pembagian produk menjadi cluster fast-, medium-, dan slow-moving memberikan dasar analitik dalam merencanakan strategi inventori yang lebih efisien dan terukur.

REFERENCE

[1] T. Abu Zwaida, C. Pham, and Y. Beauregard, “Optimization of inventory management to prevent drug shortages in the hospital supply chain,” *Applied Sciences*, vol. 11, no. 6, p. 2726, 2021, doi: 10.3390/app11062726.

[2] C.-N. Chen *et al.*, “Applying simulation optimization to minimize drug inventory costs: A study of a case outpatient pharmacy,” *Healthcare*, vol. 10, no. 3, p. 556, 2022, doi: 10.3390/healthcare10030556.

[3] K. Douaioui, R. Oucheikh, O. Benmoussa, and C. Mabrouki, “Machine learning and deep learning models for demand forecasting in supply chain management: A critical review,” *Applied System Innovation*, vol. 7, no. 5, p. 93, 2024, doi: 10.3390/asi7050093.

[4] K. P. Fourkiotis and A. Tsadiras, “Applying machine learning and statistical forecasting methods for enhancing pharmaceutical sales predictions,” *Forecasting*, vol. 6, no. 1, pp. 170–186, 2024, doi: 10.3390/forecast6010010.

[5] M. Seyedan and F. Mafakheri, “Predictive big data analytics for supply chain demand forecasting: Methods, applications, and research opportunities,” *Journal of Big Data*, vol. 7, p. 53, 2020, doi: 10.1186/s40537-020-00329-2.

[6] S. Al-Hourani and D. Weraikat, “A systematic review of artificial intelligence (AI) and machine learning (ML) in pharmaceutical supply chain (PSC) resilience: Current trends and future directions,” *Sustainability*, vol. 17, no. 14, p. 6591, 2025, doi: 10.3390/su17146591.

[7] R. I. Manarung, E. Widodo, and A. M. Rifai, “Sales Data Clustering Using the K-Means Algorithm to Determine Retail Product Needs,” *International Journal Software Engineering and Computer Science (IJSECS)*, vol. 5, no. 1, pp. 226–234, 2025, doi: 10.35870/ijsecs.v5i1.4090.

[8] D. Ayu Pradina, Y. Kurniawati, A. Syawaldi Afwan, and J. Heikal, “RFM Segmentation and K–Means Clustering of Skincare Product (Case study Scarlett),” *Jurnal Sains dan Teknologi*, vol. 6, no. 2, pp. 213–216, 2024, doi: 10.55338/saintek.v6i2.3644.

[9] N. Agrawal, A. Roy, N. K. Saxena, and P. Jain, “Predictive inventory management using data mining in the retail sector,” *Academy of Marketing Studies Journal*, vol. 29, no. 5, pp. 1–13, 2025.

;