



Penerapan Algoritma K-Means Clustering dalam Analisis Pengelompokan Produk Toko Oleh-Oleh Berdasarkan Data Penjualan

Chairunnita*, Agung Handayanto, Febrian Murti Dewanto

Fakultas Teknik dan Informatika, Informatika, Universitas PGRI Semarang, Semarang

Jl. Sidodadi Timur Jalan Dokter Cipto No. 24, Karangtempel, Kec. Semarang Tim., Kota Semarang, Jawa Tengah, Indonesia

Email: ^{1,*}chairunnita01@gmail.com, ²agunghan@upgris.ac.id, ³febrianmd@upgris.ac.id

Email Penulis Korespondensi: chairunnita01@gmail.com

Submitted: 30/06/2025; Accepted: 11/07/2025; Published: 11/07/2025

Abstrak–Toko Oleh-Oleh Dury Weleri menghadapi tantangan dalam pengelolaan stok dan strategi promosi akibat keterbatasan sistem pencatatan data penjualan yang masih konvensional. Penelitian ini bertujuan untuk mengelompokkan produk berdasarkan performa penjualan menggunakan algoritma K-Means Clustering dengan mempertimbangkan atribut total penjualan, penjualan rata-rata, dan sisa stok. Penentuan jumlah cluster optimal dilakukan melalui kombinasi Elbow Method, Silhouette Score, dan Davies-Bouldin Index, yang menghasilkan empat cluster utama. Cluster 0 terdiri dari produk dengan penjualan rendah dan stok tinggi (potensi overstock), Cluster 1 mencakup produk dengan penjualan tinggi namun stok rendah (produk unggulan), Cluster 2 berisi produk dengan penjualan sedang dan stok cukup tinggi (perlu promosi ringan), dan Cluster 3 mencakup produk dengan penjualan rendah dan stok sangat rendah (potensi produk musiman). Hasil evaluasi menggunakan Silhouette Score sebesar 0,47336 dan DBI sebesar 0,72644 menunjukkan bahwa pengelompokan yang dilakukan memiliki kualitas yang cukup baik. Visualisasi interaktif dengan Streamlit memberikan insight strategis bagi pengambilan keputusan terkait restocking dan pengelolaan promosi. Temuan ini diharapkan dapat membantu manajemen dalam mengoptimalkan pengelolaan stok, meningkatkan efisiensi operasional, dan merancang strategi penjualan yang lebih tepat sasaran.

Kata Kunci: K-Means Clustering; Pengelompokan Produk; Elbow Method; Manajemen Stok; Streamlit

Abstract–Toko Oleh-Oleh Dury Weleri faces challenges in inventory management and promotional strategy due to the limitations of a conventional sales data recording system. This study aims to classify products based on sales performance using the K-Means Clustering algorithm by analyzing total sales, average sales, and remaining stock attributes. The optimal number of clusters was determined through a combination of the Elbow Method, Silhouette Score, and Davies-Bouldin Index, resulting in four main clusters. Cluster 0 consists of products with low sales and high stock (indicating potential overstock), Cluster 1 includes products with high sales but low stock (key products), Cluster 2 comprises products with moderate sales and relatively high stock (requiring light promotions), and Cluster 3 contains products with low sales and very low stock (likely seasonal or low-priority items). The clustering evaluation produced a Silhouette Score of 0.47336 and a DBI of 0.72644, indicating a reasonably good grouping quality. Interactive visualization via Streamlit provided strategic insights for decision-making regarding restocking and promotional planning. These findings are expected to support management in optimizing inventory control, improving operational efficiency, and developing more targeted sales strategies.

Keywords: K-Means Clustering; Product Segmentation; Elbow Method; Inventory Management; Streamlit

1. PENDAHULUAN

Toko Oleh-Oleh Dury Weleri adalah salah satu pusat oleh-oleh yang terletak di Weleri, menyediakan berbagai produk khas lokal untuk memenuhi kebutuhan wisatawan maupun masyarakat sekitar. Meski berkomitmen memberikan produk berkualitas, toko ini menghadapi kendala dalam mengoptimalkan data penjualan untuk pengambilan keputusan strategi. Sistem manajemen data yang masih konvensional menyulitkan identifikasi produk dengan performa penjualan tinggi maupun rendah secara efisien dan akurat [1]. Kondisi ini berdampak pada pengelolaan stok dan efektivitas strategi promosi.

Dalam menghadapi tantangan ini, pemanfaatan teknologi menjadi solusi yang tepat. Salah satu pendekatan yang dapat digunakan adalah data mining, yaitu proses otomatis untuk menemukan pola dan informasi bermakna dari kumpulan data berukuran besar. Proses ini memanfaatkan metode analisis statistik, perhitungan matematis, kecerdasan buatan, serta machine learning untuk menelusuri dan mengidentifikasi pengetahuan tersembunyi yang relevan dan bernilai dalam database [2]. Dengan demikian, manajemen dapat mengenali dan mengatasi permasalahan terkait pengelolaan stok produk yang belum optimal.

Pengelompokan produk berdasarkan jumlah stok, transaksi, dan harga, dapat membantu dalam strategi penjualan dan pengelolaan stok yang lebih efektif [3]. Pendekatan ini juga memungkinkan manajemen untuk memetakan performa produk di Dury Weleri berdasarkan jumlah transaksi, total penjualan dan rata-rata. Dengan mengelompokkan produk ke dalam kategori penjualan tinggi, sedang, dan rendah, manajemen toko dapat mempersiapkan stok barang yang disesuaikan dengan permintaan [4]. Selain itu, pendekatan ini memungkinkan manajemen untuk mengoptimalkan alokasi sumber daya dan meningkatkan rotasi stok secara keseluruhan.

Salah satu algoritma unsupervised learning yang populer dan relevan untuk proses pengelompokan data adalah K-Means Clustering. Algoritma ini dipilih karena sederhana, memiliki kecepatan komputasi yang tinggi, dan mampu mengelompokkan data numerik secara efisien dan mudah diinterpretasikan [5]. K-Means sangat cocok digunakan pada dataset berukuran kecil hingga menengah, yang umumnya dimiliki oleh pelaku usaha mikro, kecil,



serta pelaku skala menengah [6]. Selain itu, algoritma ini mampu membentuk struktur kluster yang jelas, sehingga memudahkan analisis serta pengambilan keputusan strategis dalam konteks bisnis [7]. Dalam pengelolaan stok barang, K-Means terbukti efektif dalam mengidentifikasi pola ketidakseimbangan persediaan, sebagaimana ditunjukkan pada studi kasus Toko Oleh-Oleh Dury Weleri [8].

Penelitian yang berkaitan dengan clustering dalam menentukan kelompok produk pernah dilakukan oleh Pambudi dan Witanti (2022) mengenai penggunaan algoritma K-Means dalam mengolah data penjualan di Toko Ayu Collection yang membuktikan efektivitas metode tersebut dalam mengelompokkan produk berdasarkan pola penjualannya. Hasil analisis berhasil mengidentifikasi sejumlah segmen produk, sehingga manajemen dapat membedakan antara produk yang laris dan kurang diminati. Temuan tersebut dapat dimanfaatkan sebagai dasar dalam menyusun strategi pemasaran yang lebih tepat sasaran serta upaya meningkatkan kepuasan pelanggan [9].

Penelitian serupa oleh Lestari et al (2024) yang mengkaji penggunaan algoritma K-Means Clustering terhadap data penjualan produk sepatu futsal menunjukkan bahwa algoritma ini berhasil dalam mengelompokkan item sesuai dengan pola penjualannya. Hasil penelitian mengidentifikasi kategori produk yang laris dan kurang diminati, untuk mendukung manajemen dalam menentukan kebijakan yang lebih tepat dalam pengadaan dan pemasaran. Temuan ini diharapkan dapat meningkatkan efisiensi penjualan dan kepuasan pelanggan [10].

Penelitian lainnya dilakukan oleh Tabianan et al (2022) mengenai pendekatan K-Means clustering untuk segmentasi pelanggan cerdas menggunakan data perilaku pembelian menunjukkan bahwa metode ini efektif dalam mengelompokkan pelanggan berdasarkan kebiasaan belanja mereka. Hasil penelitian mengidentifikasi beberapa segmen pelanggan, yang memungkinkan perusahaan untuk merumuskan strategi pemasaran yang lebih tepat sasaran. Temuan ini diharapkan dapat meningkatkan kepuasan pelanggan dan efisiensi penjualan [11].

Penelitian berikutnya oleh Rifqi dan Aldisa (2023) menerapkan algoritma K-Means clustering untuk mengidentifikasi pelanggan potensial pada koperasi simpan pinjam. Hasilnya menunjukkan bahwa metode ini efektif dalam mengelompokkan pelanggan berdasarkan karakteristik keuangan. Temuan ini dapat membantu koperasi merumuskan strategi pemasaran yang lebih tepat [12].

Penelitian selanjutnya oleh Ameliana et al (2024) menunjukkan efektivitas penggunaan metode K-Means dalam mengklasifikasikan pelaku UMKM di Jawa Barat dengan mempertimbangkan karakteristik tertentu. Hasil analisis mengidentifikasi dua cluster utama yang berbeda dalam kinerja dan potensi pertumbuhan, memberikan wawasan bagi pemerintah dan pelaku usaha untuk merumuskan strategi pengembangan yang lebih baik. Temuan ini diharapkan dapat meningkatkan efisiensi operasional dan daya saing UMKM di pasar [13].

Berbeda dengan penelitian sebelumnya, penelitian ini menggabungkan evaluasi multi-metrik yaitu Elbow Method, Silhouette Score, dan Davies Bouldin untuk menentukan jumlah cluster yang optimal secara lebih akurat. Hasil pengelompokan divisualisasikan secara interaktif menggunakan Streamlit untuk memudahkan interpretasi dan pengambilan keputusan, terutama bagi pelaku usaha non-teknis. Dengan mempertimbangkan permasalahan yang dihadapi oleh Toko Oleh-Oleh Dury Weleri, penelitian ini bertujuan untuk mengklasifikasikan data penjualan produk melalui penerapan algoritma K-Means Clustering [14]. Pendekatan ini diharapkan mampu menghasilkan segmentasi produk yang mendukung manajemen dalam menetapkan kebijakan restocking dan strategi promosi yang lebih tepat, serta berkontribusi pada peningkatan efisiensi operasional dan kepuasan pelanggan [15].

2. METODOLOGI PENELITIAN

2.1 Jenis Penelitian

Penelitian ini bersifat kuantitatif melalui pendekatan eksploratif dengan memanfaatkan metode data mining, yaitu algoritma K-Means Clustering guna mengelompokkan produk sesuai pola dari data penjualan pada Toko Oleh-Oleh Dury Weleri.

2.2 Tahapan Penelitian

Tahapan dalam sebuah penelitian merupakan rangkaian proses yang tersusun secara sistematis dan dijalankan sebagai panduan dalam pelaksanaan penelitian ini [16]. Tahapan ini menjelaskan prosedur yang direncanakan untuk dilaksanakan selama proses penelitian berlangsung. Langkah-langkah dari tahapan penelitian sebagai berikut:

- a. Pengumpulan data
Data yang digunakan dalam penelitian ini dikumpulkan dari Toko Oleh-Oleh Dury Weleri pada periode April - Mei 2025. Dataset terdiri dari 79 entri produk yang dikelompokkan dalam lima kategori yaitu: pie dan kue kering, olahan ikan, olahan pisang dan buah, oleh-oleh khas Kudus serta oleh-oleh khas Semarang. Setiap entri memuat atribut nama produk, kategori, total penjualan, penjualan rata-rata, dan sisa stok.
- b. Preprocessing Data
Atribut seperti total penjualan, penjualan rata-rata, dan sisa stok dipilih untuk dianalisis. Data telah dipastikan bebas dari missing value dan duplikasi, lalu dinormalisasi menggunakan Min-Max Scaling. Outlier tidak dihapus agar seluruh produk, termasuk yang ekstrem, tetap dianalisis untuk mendukung strategi pemasaran yang lebih tepat sasaran. Langkah ini memastikan bahwa model clustering mencerminkan kondisi riil dan mampu mengidentifikasi pola dari seluruh spektrum data yang tersedia.
- c. Penentuan Jumlah Cluster

Penentuan jumlah k-cluster dilakukan menggunakan kombinasi Elbow Method, Silhouette Score, dan Davies-Bouldin Index (DBI) untuk memperoleh nilai K optimal secara visual dan statistik, yang kemudian divisualisasikan dalam grafik. Ketiga metrik ini saling melengkapi dalam mengevaluasi efisiensi, pemisahan, dan kekompakan cluster, sehingga hasil klusterisasi menjadi lebih akurat dan dapat diandalkan.

d. Penerapan K-Means Clustering

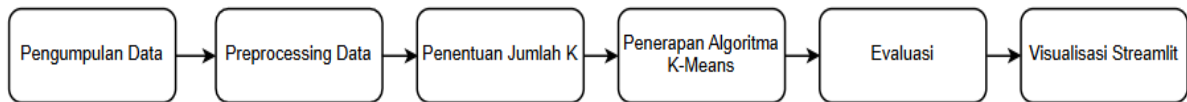
Setelah banyaknya cluster ditentukan, data yang sudah telah dinormalisasi dianalisis menggunakan algoritma K-Means. Tahapan ini dilakukan secara iteratif, mulai dari inialisasi centroid hingga konvergen, dengan menggunakan perhitungan jarak Euclidean untuk mengelompokkan data.

e. Evaluasi

Evaluasi hasil clustering dilakukan menggunakan Silhouette Score untuk mengukur koherensi dan keterpisahan cluster (dengan skor antara -1 hingga 1, semakin tinggi semakin baik), serta Davies-Bouldin Index (DBI) digunakan untuk mengevaluasi perbandingan antara jarak antara satu cluster dengan cluster lainnya dan kedekatan internal dalam cluster (nilai yang lebih rendah menunjukkan hasil yang lebih baik).

f. Visualisasi Streamlit

Tahap ini memanfaatkan Streamlit untuk membangun antarmuka visual dari hasil clustering. Visualisasi dilakukan menggunakan scatter plot dua dimensi yang menampilkan distribusi produk per cluster secara interaktif. Komponen ini dirancang untuk mendukung interpretasi hasil analisis dan memfasilitasi penyampaian insight secara real-time. Gambar 1 menunjukkan alur proses tahap penelitian.



Gambar 1. Alur Tahapan Penelitian

2.3 Algoritma K-Means

K-Means merupakan algoritma unsupervised learning yang dimanfaatkan untuk membagi data terbagi dalam kelompok (cluster) berdasarkan kemiripan antar data [17]. Algoritma K-means adalah algoritma yang memerlukan input berupa jumlah cluster (k) dan mengelompokkan n objek ke dalam k cluster dengan tujuan agar objek-objek dalam satu cluster memiliki tingkat kemiripan yang tinggi, sementara kemiripan antar objek dari cluster yang berbeda seminimal mungkin [18]. Langkah-langkah yang dilakukan dalam proses K-Means adalah sebagai berikut [19]:

- a. Menentukan nilai K, di mana nilai K merepresentasikan jumlah cluster.
- b. Menentukan titik secara acak sebagai centroid atau pusat cluster untuk inialisasi cluster awal.
- c. Menghitung jarak setiap titik dengan centroid. Rumus yang digunakan adalah Euclidean Distance.

$$D_{(ij)} = \sqrt{\sum_{i=0}^n (x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + (x_{3i} - x_{3j})^2} \tag{1}$$

- d. Menetapkan setiap titik ke cluster dengan jarak terdekat.
- e. Menghitung rata-rata dari setiap cluster untuk menjadi centroid baru menggunakan rumus:

$$C_j = \frac{1}{|s_j|} \sum_{x_i \in s_j} x_i \tag{2}$$

- f. Mengulangi langkah ketiga hingga langkah kelima untuk titik tengah yang baru dari setiap cluster.

2.4 Kombinasi Elbow Method, Silhouette Score, dan Davies-Bouldin Index

Penentuan jumlah cluster yang optimal dalam penelitian ini dilakukan dengan pendekatan multi-metrik menggunakan Elbow Method, Silhouette Score, dan Davies-Bouldin Index (DBI). Elbow Method menganalisis hubungan antara jumlah cluster dan nilai SSE untuk menemukan titik siku yang menunjukkan jumlah cluster terbaik [20]. Silhouette Score digunakan untuk mengukur sejauh mana suatu data sesuai dengan cluster-nya sendiri dibandingkan dengan cluster lain, dengan skor mendekati +1 menunjukkan pemisahan yang baik [21]. Sementara itu, DBI mengevaluasi kekompakan dan pemisahan antar cluster, di mana nilai DBI yang lebih rendah menunjukkan kualitas cluster yang lebih tinggi [22]. Pendekatan gabungan ini memungkinkan pemilihan jumlah cluster tidak hanya berdasarkan visualisasi, tetapi juga validasi statistik yang lebih kuat. Studi terdahulu juga menunjukkan bahwa kombinasi metrik ini memberikan hasil pengelompokan yang lebih akurat dan informatif dibandingkan hanya menggunakan satu indikator tunggal [23]. Pendekatan ini memberikan dasar yang lebih objektif dalam menentukan struktur cluster yang representatif terhadap karakteristik data. Dengan mempertimbangkan ketiga metrik secara bersamaan, keputusan jumlah cluster yang dipilih menjadi lebih andal untuk mendukung analisis lanjutan dan pengambilan keputusan bisnis yang tepat.

2.5 Alat dan Bahasa Pemrograman

Penelitian ini menggunakan:



- a. Google Colab, digunakan sebagai platform pemrosesan data berbasis cloud yang memungkinkan kolaborasi dan eksekusi kode Python secara interaktif.
- b. Python, digunakan sebagai bahasa pemrograman untuk mengimplementasikan algoritma K-Means Clustering.
- c. Library Python yang digunakan meliputi: pandas, numpy, sklearn, matplotlib, dan seaborn.
- d. Microsoft Excel, dimanfaatkan sebagai alat bantu untuk perhitungan manual dan verifikasi hasil clustering secara tabular.
- e. Streamlit, digunakan untuk membangun visualisasi interaktif dari hasil pengelompokan guna mempermudah interpretasi data oleh pengguna.

3. HASIL DAN PEMBAHASAN

3.1 Data Selection

Tahap ini bertujuan untuk memilih atribut-atribut yang relevan dari dataset awal untuk dianalisis menggunakan algoritma K-Means Clustering. Dataset penelitian terdiri dari 79 entri produk, masing-masing memiliki atribut seperti nama produk, kategori, total penjualan, penjualan rata-rata dan sisa stok. Hanya tiga atribut numerik seperti total penjualan, penjualan rata-rata, dan sisa stok yang digunakan dalam proses clustering karena K-Means bekerja berdasarkan perhitungan jarak (distance-based) yang membutuhkan nilai kuantitatif. Atribut nama produk dan kategori tetap disertakan sebagai informasi tambahan. Contoh sebagian data yang dipakai selama tahapan analisis tersaji pada Tabel 1.

Tabel 1. Data Selection

No	Nama	Total Penjualan	Penjualan Rata-Rata	Sisa Stok
1	Chocolate Ball	209	54	29
2	BDE SPR1	726	162	23
3	Mubarak Mika'15	248	62	79
4	Krasikan KCL	252	63	58
5	Intip Goreng	220	55	95
..
79	Gula Kacang	218	54	86

Tabel 1 menyajikan sebagian data yang dianalisis. Sebagai contoh, produk BDE SPR 1 memiliki total penjualan 726, penjualan rata-rata 162 dan sisa stok 23, yang mencerminkan performa penjualan tinggi. Atribut numerik dipilih karena relevan untuk mengevaluasi penjualan stok secara kuantitatif, serta mendukung analisis pola produk yang lebih akurat.

3.2 Preprocessing Data

Langkah awal dalam analisis adalah normalisasi data dengan metode Min-Max Scaling agar nilai setiap atribut berada pada rentang 0 hingga 1. Tujuannya adalah untuk menyetarakan skala antar fitur sehingga tidak ada atribut yang mendominasi perhitungan jarak pada proses K-Means Clustering. Normalisasi ini penting karena algoritma K-Means sangat sensitif terhadap skala data, di mana fitur dengan nilai besar dapat mempengaruhi hasil pengelompokan secara tidak proporsional. Proses ini dilakukan menggunakan library sklearn.preprocessing di Python untuk memastikan hasil yang konsisten dan efisien. Sebelum tahap normalisasi, data telah diperiksa dan dipastikan bebas dari missing value maupun data duplikat. Selain itu, data outlier tidak dihapus, agar seluruh produk including yang memiliki karakteristik ekstrem tetap dianalisis sebagai bagian dari strategi segmentasi pemasaran yang lebih menyeluruh dan relevan. Data yang telah dinormalisasi dapat dilihat pada Tabel 2.

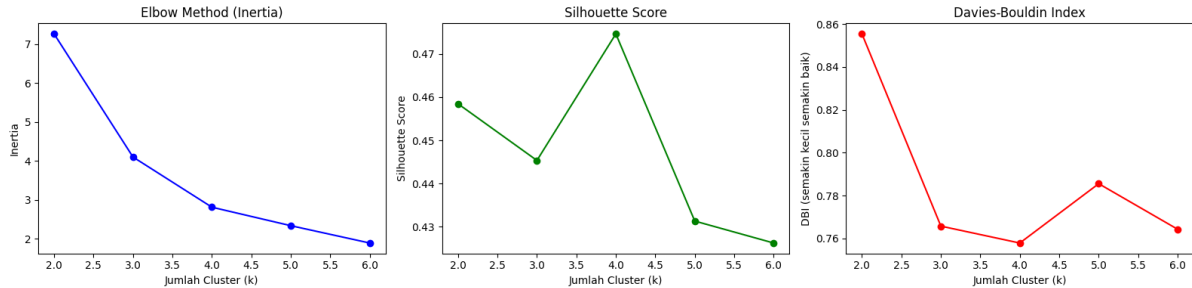
Tabel 2. Data Normalization

No	Nama	Total Penjualan	Penjualan Rata-Rata	Sisa Stok
1	Chocolate Ball	0.15533	0.19403	0.19048
2	BDE SPR1	1	1	0.14286
3	Mubarak Mika'15	0.21895	0.25373	4,078472
4	Krasikan KCL	0.22549	0.26119	0.42063
5	Intip Goreng	1,202778	0.20149	0.71429
..
79	Gula Kacang	0.16993	0.19403	0.64286

Tabel 2 pada produk BDE SPR 1 memiliki nilai 1 pada dua atribut, menandakan performa tertinggi dalam penjualan total dan penjualan rata-rata dibanding produk lainnya. Hasil normalisasi ini menjadi dasar input dalam proses K-Means Clustering.

3.3 Penentuan Jumlah Cluster

Berbeda dengan penelitian sebelumnya, penelitian ini dalam menentukan jumlah cluster yang optimal tidak hanya bergantung pada satu pendekatan, melainkan menggunakan kombinasi tiga metode, yaitu Elbow Method, Silhouette Score, dan Davies-Bouldin Index (DBI), guna memastikan segmentasi yang valid secara statistik dan relevan secara bisnis. Kombinasi tiga metode dapat dilihat pada Gambar 2.



Gambar 2. Metode Elbow

Grafik Elbow menunjukkan penurunan inerti yang signifikan hingga $K = 4$, kemudian perubahan nilai mulai melandai, sehingga $K = 4$ diidentifikasi sebagai titik siku (elbow point) dan dianggap sebagai jumlah cluster yang optimal. Selanjutnya, analisis melalui Silhouette Score memperkuat hasil tersebut, di mana skor tertinggi diperoleh pada $K = 4$, yang mengindikasikan bahwa struktur cluster cukup terpisah dan koheren. Hal ini menunjukkan bahwa mayoritas data lebih dekat dengan cluster-nya masing-masing daripada dengan cluster lainnya. Dukungan tambahan ditunjukkan oleh nilai Davies-Bouldin Index terendah yang juga muncul pada $K = 4$, menandakan bahwa antar cluster memiliki pemisahan yang baik dengan kekompakan internal yang memadai. Dengan demikian, berdasarkan keselarasan hasil ketiga metode ini, jumlah cluster $K = 4$ dianggap optimal karena mampu memberikan segmentasi yang informatif, mudah diinterpretasi, dan dapat dijadikan dasar dalam pengambilan keputusan strategis, khususnya dalam pengelolaan stok dan penyusunan strategi promosi yang lebih efektif.

3.4 Menghitung Jarak Objek ke Titik Pusat

Menentukan centroid awal sangat penting untuk menentukan seberapa jauh suatu objek dari pusat cluster. Penentuan centroid awal dilakukan secara acak. Dalam penelitian ini, centroid awal untuk C0 ditetapkan dari data ke-25, C1 dari data ke-2, C2 dari data ke-3, dan C3 dari data ke-12. Informasi lebih lanjut disajikan pada Tabel 3.

Tabel 3. Centroid Awal

Centroid Awal	Total Penjualan	Penjualan Rata-Rata	Sisa Stok
C0	0.09967	0,829167	1
C1	1	1	0.14286
C2	0.21895	0.25373	4,078472
C3	0.47222	0.54478	0.07937

Setelah menetapkan centroid awal, langkah selanjutnya yakni menggunakan rumus Euclidean Distance untuk menghitung jarak antara setiap data ke centroid dengan excel. Dibawah ini akan dilakukan perhitungan dengan excel pada contoh data ke-2 seperti yang ditunjukkan di bawah ini:

$$D_{(ij)} = \sqrt{\sum_{i=0}^n (x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + (x_{3i} - x_{3j})^2}$$

Menghitung jarak dari Titik Data 2 ke Centroid 0 dengan nilai atribut (0.09967; 0.1194; 1):

$$C0 = \sqrt{(1 - 0,09967)^2 + (1 - 0,1194)^2 + (0,14286 - 1)^2}$$

$$C0 = 1,523397338$$

Menghitung jarak dari Titik Data 2 ke Centroid 1 dengan nilai atribut (1; 1; 0.14286):

$$C1 = \sqrt{(1 - 1)^2 + (1 - 1)^2 + (0,14286 - 0,14286)^2}$$

$$C1 = 0$$

Menghitung jarak dari Titik Data 2 ke Centroid 2 dengan nilai atribut (0.21895; 0.25373; 0.5873):

$$C2 = \sqrt{(1 - 0,21895)^2 + (1 - 0,25373)^2 + (0,14286 - 0,5873)^2}$$

$$C2 = 1,168111694$$

Menghitung jarak dari Titik Data 2 ke Centroid 3 dengan nilai atribut (0.47222; 0.54478; 0.07937):

$$C3 = \sqrt{(1 - 0,47222)^2 + (1 - 0,54478)^2 + (0,14286 - 0,07937)^2}$$

$$C3 = 0,699862813$$

Hasil perhitungan jarak Euclidean dapat dilihat pada iterasi-1 pada Tabel 4.

Tabel 4. Jarak setiap Data Pada Iterasi-1

Data ke-i	C0	C1	C2	C3	Cluster
1	0.814849177	1.168.542.536	0.40631321	0.485647665	2
2	1.523.397.338	0	1.168.111.694	0.699862813	1
3	0.450103941	1.168.111.694	0	0.63784691	2
4	0.609593859	1.105.830.968	0.166965003	0.507698108	2
5	0.306228074	1.283.646.643	0.144736105	0.78127101	2
..
79	0.371557646	1.260.398.289	0.09515232	0.729332425	2

Setelah dilakukan perhitungan rata-rata nilai centroid, pada iterasi 2,3,4,5,6,7,8 dan 9 nilai cluster masih berubah, baru pada iterasi-10 nilai nilai cluster tidak berubah menandakan bahwa proses clustering berhenti pada iterasi-10. Berikut adalah hasil dari jarak setiap data pada iterasi-10, bisa dilihat pada Tabel 5.

Tabel 5. Jarak Setiap Iterasi-10

Data ke-i	C0	C1	C2	C3	Cluster
1	0.48433272	0.848063175	0.595563514	0.135554993	3
2	134.702.233	0.358288845	0.892162949	1.036.141.113	1
3	0.18002271	0.816129351	0.328665526	0.377948675	0
4	0.30122645	0.759372586	0.373879788	0.21159883	3
5	0.10152958	0.932723057	0.408045903	0.515292097	0
..
79	0.08720265	0.908805152	0.405134762	0.448076645	0

Pada Tabel 5 dari iterasi ke-10, menandakan bahwa proses konvergensi telah tercapai dan algoritma K-Means berhenti secara otomatis. Hal ini menunjukkan bahwa model telah menemukan posisi centroid yang stabil untuk membagi data ke dalam kelompok yang optimal. Hasil akhir dari nilai centroid pada Iterasi-10 disajikan pada Tabel 6.

Tabel 6. Nilai Centroid Akhir Iterasi-10

Cluster	Total Penjualan	Penjualan Rata-Rata	Sisa Stok	Jumlah
0	0.114189231	0.132318846	0.669108846	26
1	0.725671111	0.81509222	0.28042333	9
2	0.435018462	0.49942462	0.618438462	13
3	0.24984129	0.2886371	0.212236774	31

Pada Tabel 6, merupakan nilai centroid yang merepresentasikan nilai rata-rata dari setiap atribut total penjualan, penjualan rata-rata, dan sisa stok dalam masing-masing cluster dan menjadi dasar interpretasi karakteristik tiap kelompok produk.

3.5 Implementasi K-Means dengan Python

a. Penyesuaian Centroid Awal

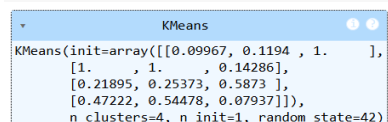
Untuk memastikan hasil clustering yang konsisten dan dapat dibandingkan secara adil antara analisis manual menggunakan Excel dan otomatis menggunakan Python, maka dilakukan penyesuaian terhadap pemilihan centroid awal. Dalam implementasi K-Means di Python, empat titik centroid awal ditetapkan secara manual dengan merujuk pada data ke-25, ke-2, ke-3, dan ke-12, sesuai dengan titik awal yang digunakan dalam proses clustering manual sebelumnya. Langkah ini dilakukan untuk menghilangkan variasi akibat inisialisasi acak dan memungkinkan evaluasi yang lebih objektif terhadap hasil pengelompokan. Visualisasi penempatan centroid awal tersebut disajikan pada Gambar 3.

```

initial_centroids = np.array([
    [0.09967, 0.1194, 1.0], # C0
    [1.0, 1.0, 0.14286], # C1
    [0.21895, 0.25373, 0.5873], # C2
    [0.47222, 0.54478, 0.07937] # C3
])

# Jalankan k-means dengan centroid awal
km = KMeans(n_clusters=4, init=initial_centroids, n_init=1, random_state=42)
labels = km.fit_predict(X_scaled)
km

```



Gambar 3. Penyesuaian Centroid Awal

b. Centroid Akhir

Setelah algoritma K-Means dijalankan dengan $K = 4$ dan centroid awal yang telah ditentukan secara manual, diperoleh hasil centroid akhir untuk masing-masing cluster. Nilai ini mencerminkan posisi pusat kelompok berdasarkan hasil iterasi algoritma hingga mencapai konvergensi. Centroid akhir ini kemudian dibandingkan dengan hasil manual yang dihitung menggunakan Excel sebagai bentuk validasi. Nilai centroid akhir ditampilkan pada Gambar 4.

```
# Simpan centroid ke dalam DataFrame dengan nama kolom
centroids = km.cluster_centers_
dataCentroid = pd.DataFrame(centroids, columns=cols)
print("Centroid Akhir:")
print(dataCentroid)
```

Centroid Akhir:

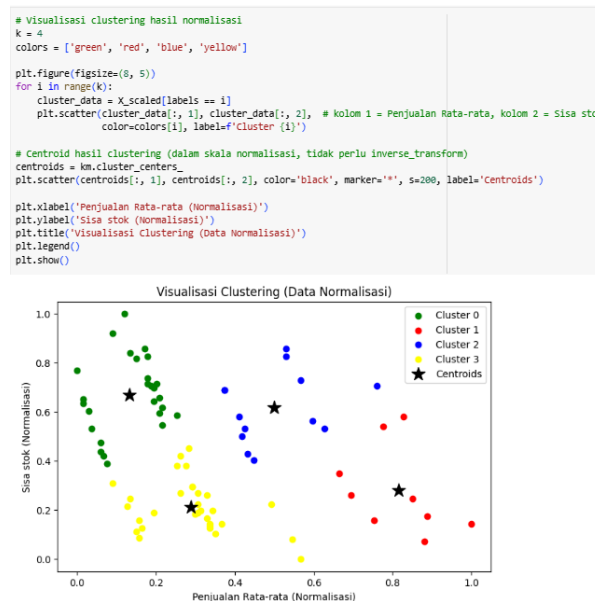
Total Penjualan	Penjualan Rata-rata	Sisa stok
0	0.114191	0.132319
1	0.725672	0.815091
2	0.435018	0.499426
3	0.249842	0.288637

Gambar 4. Nilai Centroid Akhir dari Python

Meskipun terdapat sedikit perbedaan pada digit desimal antara hasil centroid dari perhitungan manual di Excel dan hasil otomatis menggunakan Python, nilai tersebut secara prinsip identik. Perbedaan tersebut terjadi akibat tingkat presisi dan metode pembulatan bawaan masing-masing perangkat lunak, namun tidak memengaruhi struktur maupun akurasi hasil clustering secara keseluruhan.

c. Visualisasi Clustering (Scatter Plot)

Untuk melihat distribusi data hasil pengelompokan, digunakan scatter plot dengan pewarnaan berbeda untuk setiap cluster. Titik-titik pada plot mewakili produk, sedangkan centroid divisualisasikan sebagai titik pusat dari masing-masing cluster. Visualisasi ini mempermudah interpretasi hasil pengelompokan secara visual dan membantu dalam mengidentifikasi pola antar kelompok produk. Hasil scatter plot ditampilkan pada Gambar 5.

**Gambar 5.** Visualisasi Clustering K-Means dengan 4 Cluster

Grafik dari visualisasi di atas menunjukkan pengelompokan produk berdasarkan pola penjualan dan stok menggunakan pengelompokan K-Means. Analisis menunjukkan empat cluster utama. Cluster C0 berisi 26 produk dengan penjualan rendah tetapi stok tinggi, yang meandakan potensi overstock yang perlu dikelola dengan promosi atau pengurangan pasokan. Cluster C1 berisi 9 produk dengan penjualan tinggi tetapi stok rendah, yang menunjukkan produk dalam cluster perlu strategi restock cepat dan berkala. Cluster C2 berisi 13 produk dengan penjualan sedang tetapi stok cukup tinggi, yang menunjukkan perlu dilakukan pemantauan penjualan atau diberikan dorongan seperti bundling dan promosi ringan agar stok tetap terkontrol. Cluster C3 berisi 31 produk dengan penjualan rendah dan stok sangat rendah, yang menunjukkan perlu dilakukan evaluasi keberlanjutan produk yang mungkin menunjukkan bahwa produk ini musiman atau tidak prioritas pelanggan.

3.6 Evaluasi

Berdasarkan hasil perhitungan menggunakan metode Elbow, diperoleh bahwa nilai K optimal adalah 4. Untuk memastikan bahwa pemilihan $K = 4$ memang memberikan hasil pengelompokan yang baik, dilakukan evaluasi

menggunakan metrik validasi, yaitu Silhouette Score dan Davies-Bouldin Index (DBI). Nilai kedua metrik tersebut berfungsi untuk mengukur seberapa baik pemisahan dan kekompakan antar cluster. Hasil evaluasi terhadap nilai $K = 4$ ditampilkan pada Gambar 6.

```
# Evaluasi clustering
dbi = davies_bouldin_score(X_scaled, labels)
silh = silhouette_score(X_scaled, labels)
print("Davies-Bouldin Index:", round(dbi, 5))
print("Silhouette Score:", round(silh, 5))

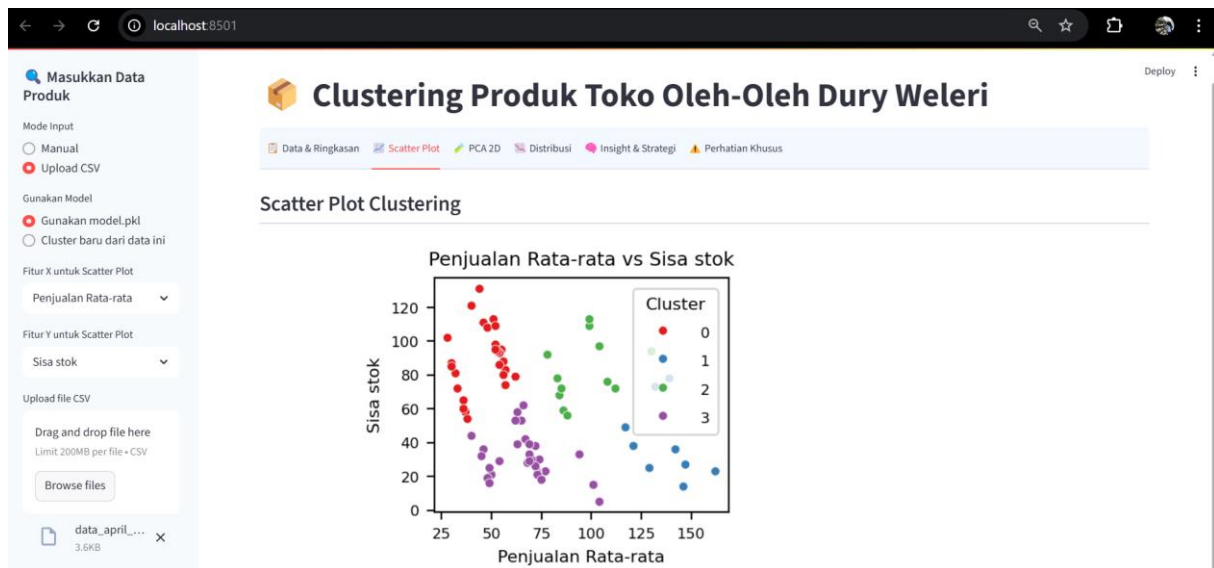
Davies-Bouldin Index: 0.72644
Silhouette Score: 0.47336
```

Gambar 6. Hasil evaluasi dari Silhouette Score dan DBI

Berdasarkan hasil evaluasi terhadap pengelompokan K-Means dengan $K = 4$, diperoleh nilai Davies-Bouldin Index (DBI) sebesar 0.72644 dan Silhouette Score sebesar 0.47336. Nilai DBI yang berada di bawah 1 menunjukkan bahwa antar cluster memiliki pemisahan yang cukup baik, dengan tingkat keseragaman internal yang juga memadai. Semakin rendah nilai DBI, semakin baik kualitas pengelompokan, sehingga nilai 0.72644 dapat dikategorikan sebagai cukup baik. Sementara itu, nilai Silhouette Score yang diperoleh mendekati 0.5 menunjukkan bahwa struktur pengelompokan cukup jelas dan dapat diterima. Skor ini mengindikasikan bahwa sebagian besar data lebih dekat ke cluster-nya masing-masing dibandingkan ke cluster lain, meskipun belum mencapai kategori sangat baik. Dengan demikian, berdasarkan kedua metrik ini, dapat disimpulkan bahwa penggunaan $K = 4$ dalam proses clustering memberikan hasil pengelompokan yang layak secara statistik dan relevan secara bisnis.

3.7 Implementasi Visualisasi dengan Streamlit

Visualisasi hasil clustering K-Means diimplementasikan menggunakan Streamlit, sebuah framework Python yang sederhana untuk membuat aplikasi web interaktif. Berikut visualisasi streamlit pada Gambar 7.



Gambar 7. Visualisasi dengan Streamlit

Gambar 7 merupakan data hasil clustering yang ditampilkan dalam bentuk scatter plot dua dimensi, di mana setiap cluster diberi warna yang berbeda berdasarkan penjualan rata-rata dan sisa stok. Selain visualisasi, tampilan ini juga menyajikan data dan ringkasan, PCA 2D, distribusi, insight dan strategi serta produk dengan perhatian khusus. Tampilan data dan ringkasan mencakup keseluruhan data berada pada posisi cluster ke berapa. Tampilan PCA 2D merupakan visualisasi dari hubungan atribut total penjualan, penjualan rata-rata dan sisa stok. Pada menu distribusi merupakan seberapa persen distribusi dari masing-masing cluster. Insight dan strategi merupakan strategi yang bisa diambil oleh pihak toko berdasarkan pola data yang terbentuk, guna membantu pemilik toko dalam pengambilan keputusan bisnis yang lebih tepat. Dan pada menu perhatian khusus merupakan beberapa produk yang jumlah stoknya tinggi serta menunjukkan penjualan yang rendah. Visualisasi ini dirancang agar pengguna non-teknis pun dapat dengan mudah memahami hasil pengelompokan tanpa perlu membaca data mentah secara langsung. Dengan demikian, Streamlit tidak hanya berperan sebagai alat presentasi data, tetapi juga sebagai media eksplorasi yang mempercepat pemahaman dan penerapan hasil analisis dalam konteks bisnis.



4. . KESIMPULAN

Penerapan algoritma K-Means berhasil mengelompokkan 79 produk pada Toko Oleh-Oleh Dury Weleri ke dalam empat cluster berdasarkan atribut total penjualan, penjualan rata-rata, dan sisa stok. Penentuan jumlah cluster dilakukan menggunakan kombinasi Elbow Method, Silhouette Score, dan Davies-Bouldin Index, yang menunjukkan bahwa $K = 4$ merupakan jumlah cluster optimal dengan kualitas pengelompokan yang cukup baik. Cluster 0 terdiri dari 26 produk dengan penjualan rendah dan stok tinggi, menunjukkan potensi overstock yang membutuhkan strategi promosi atau pengurangan pasokan. Cluster 1 mencakup 9 produk dengan penjualan tinggi namun stok rendah, yang merupakan produk unggulan dan membutuhkan strategi restocking cepat serta berkala agar tidak terjadi kekosongan stok. Cluster 2 berisi 13 produk dengan penjualan sedang dan stok cukup tinggi, menandakan perlunya pemantauan berkala serta strategi promosi ringan seperti bundling untuk menjaga rotasi barang. Sementara itu, Cluster 3 berisi 31 produk dengan penjualan rendah dan stok sangat rendah, yang mengindikasikan kemungkinan produk musiman atau kurang diminati, sehingga memerlukan evaluasi keberlanjutan. Visualisasi menggunakan Streamlit memberikan gambaran interaktif atas distribusi produk dalam setiap cluster serta menyertakan insight dan saran strategis. Hasil ini menunjukkan bahwa pendekatan clustering dapat memberikan dasar pembuatan keputusan yang lebih akurat dalam manajemen stok serta strategi promosi. Sebagai arah pengembangan penelitian selanjutnya, disarankan melakukan eksplorasi algoritma clustering yang lain seperti DBSCAN atau Hierarchical Clustering sebagai perbandingan, guna meningkatkan akurasi segmentasi dan relevansi strategi bisnis.

REFERENCES

- [1] R. I. Manarung, E. Widodo, and A. M. Rifai, "Sales Data Clustering Using the K-Means Algorithm to Determine Retail Product Needs," *Int. J. Softw. Eng. Comput. Sci.*, vol. 5, no. 1, pp. 226–234, 2025, doi: 10.35870/ijsecs.v5i1.4090.
- [2] O. N. Nugraha and G. Dwilestari, "Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Yana Sport," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 2, pp. 849–855, 2022, doi: 10.36040/jati.v6i2.5755.
- [3] H. Al Rasyid, B. F. K. Soebari, and D. S. Y. Kartika, "Implementasi algoritma K-Means clustering untuk pengelompokan penjualan produk pada online shop Toko Gizi," in *Prosiding Seminar Nasional Teknologi dan Sistem Informasi*, 2022, pp. 242–248.
- [4] S. H. Widiastuti and R. Jumardi, "Pengelompokan Daerah Rawan Demam Berdarah dengan Metode K-Means Clustering," *J. Inf. dan Teknol.*, pp. 185–190, 2022.
- [5] L. Adhitama, S. Murniati, and C. S. Pramudyo, "Minimasi Jarak Pengiriman Roti CV. Twin Setia dengan Metode K-Means Clustering dan Simulated Annealing," *J. Tek. SILITEK*, vol. 3, no. 2, pp. 68–75, 2023.
- [6] A. N. B. Prasetyo, M. Maimunah, and P. Sukmasetya, "K-Means Clustering Method for Determining Waste Transportation Routes to Landfill," *J. Ris. Inform.*, vol. 5, no. 3, pp. 277–284, 2023.
- [7] J. Jabbar, "Sistem Informasi Stok Barang Menggunakan Metode Clustering Kmeans (Studi Kasus Rmd Store)," *INFOTECH J.*, vol. 8, no. 1, pp. 70–75, 2022.
- [8] K. Hidayat, M. R. Adytama, H. A. Darmawan, Y. Arnando, and A. Mukarim, "Analisis Data Penjualan Menggunakan Algoritma K-Means Clustering pada Toko Superindo," *JoDMApps (Journal Data Sci. Methods Appl.)*, vol. 1, no. 1, pp. 1–6, 2025, [Online]. Available: <https://journal.darmajaya.ac.id/index.php/JoDMApps/article/view/904>
- [9] W. T. Pambudi and A. Witanti, "Penerapan Algoritma K-Means Untuk Menganalisis Data Penjualan Pada Toko Ayu Collection Berbasis Web," *J. Sist. Inf. dan Bisnis Cerdas*, vol. 15, no. 1, pp. 1–9, 2022.
- [10] S. Lestari, M. Muhatri, A. R. Fachrezi, M. A. Sutrisno, and M. Geovany, "Implementasi Algoritma K-Means Clustering pada Penjualan Sepatu Futsal Merk Specs," *Polyg. J. Ilmu Komput. dan Ilmu Pengetah. Alam*, vol. 2, no. 3, pp. 50–61, 2024.
- [11] K. Tabianan, S. Velu, and V. Ravi, "K-means clustering approach for intelligent customer segmentation using customer purchase behavior data," *Sustainability*, vol. 14, no. 12, p. 7243, 2022.
- [12] A. Rifqi and R. T. Aldisa, "Penerapan Data Mining dalam Implementasi Algoritma K-Means Clustering untuk Pelanggan Potensial pada Koperasi Simpan Pinjam," *Build. Informatics, Technol. Sci.*, vol. 5, no. 2, pp. 486–497, 2023, doi: 10.47065/bits.v5i2.4278.
- [13] N. Ameliana, N. Suarna, and W. Prihartono, "Analisis Data Mining Pengelompokan Umkm Menggunakan Algoritma K-Means Clustering Di Provinsi Jawa Barat," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 3, pp. 3261–3268, 2024, doi: 10.36040/jati.v8i3.9655.
- [14] R. S. Lisudatu, M. Marchelin, and L. K. Wibisono, "Klasifikasi UMKM Berdasarkan Kinerja Keuangan Menggunakan Algoritma K-Means Clustering: Studi Kasus UMKM Sektor Industri di Kecamatan Rantepao," *J. Bisnis Mhs.*, vol. 5, no. 2, pp. 891–900, 2025.
- [15] M. Miranda and S. Sriani, "Implementation of K-Means Clustering in Grouping Sales Data at Zura Mart," *JAIC*, vol. 9, no. 2, pp. 547–555, 2025.
- [16] F. Fauji and L. Farokhah, "Penerapan Data Mining Menggunakan K-Means Clustering Dalam Mengelompokkan Tingkat Kesulitan Mata Pelajaran," *J. Inf. Syst. Res.*, vol. 6, no. 3, pp. 1705–1714, 2025, doi: 10.47065/josh.v6i3.6959.
- [17] F. Pangestu, N. Yasin, R. C. Hasugian, and Yunita, "Penerapan Algoritma K-Means Untuk Mengklasifikasi Data Obat," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 12, no. 1, pp. 53–62, 2023, doi: 10.32736/sisfokom.v12i1.1461.
- [18] F. Yunita, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru (Studi Kasus: Universitas Islam Indragiri)," *J. Sist.*, vol. 7, no. 3, pp. 238–249, 2018.
- [19] R. R. Muhima, M. Kurniawan, S. Kom, M. Kom, and A. Yudhana, *Kupas tuntas algoritma clustering: Konsep, perhitungan manual, dan program*. Penerbit Andi, 2022.
- [20] I. F. Ashari, E. D. Nugroho, R. Baraku, I. N. Yanda, and R. Liwardana, "Analysis of elbow, silhouette, Davies-Bouldin,



- Calinski-Harabasz, and rand-index evaluation on k-means algorithm for classifying flood-affected areas in Jakarta,” *J. Appl. Informatics Comput.*, vol. 7, no. 1, pp. 95–103, 2023.
- [21] M. R. Syahkur, D. Hartama, and S. Solikhun, “Evaluasi Jumlah Cluster pada Algoritma K-Means++ Menggunakan Silhouette dan Elbow dengan Validasi Nilai DBI dalam Mengelompokkan Gizi Balita,” *JST (Jurnal Sains dan Teknol.*, vol. 13, no. 3, 2024.
- [22] H. Mulyani, R. A. Setiawan, and H. Fathi, “Optimization of K value in clustering using silhouette score (case study: Mall customers data),” *J. Inf. Technol. Its Util.*, vol. 6, no. 2, pp. 45–50, 2023.
- [23] I. T. Utami, F. Suryaningrum, and D. Ispriyanti, “K-Means Cluster Count Optimization With Silhouette Index Validation And Davies Bouldin Index (Case Study: Coverage Of Pregnant Women, Childbirth, And Postpartum Health Services In Indonesia In 2020),” *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 17, no. 2, pp. 707–716, 2023.