

Analisis Segmentasi Transaksi Penjualan Sayuran menggunakan *Algoritma K-Means Clustering* untuk Optimalisasi Strategi Penjualan

Vegetable Sales Transaction Segmentation using K-Means Clustering for Sales Strategy Optimization

¹Dimas Wahyu Eko Prasetyo, ²Charitas Fibriani*

^{1,2}Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana, Salatiga, Jawa Tengah, Indonesia

*e-mail: charitas.fibriani@uksw.edu

(received: 29 June 2026, revised: 12 August 2026, accepted: 13 August 2026)

Abstrak

Pengelolaan persediaan sayuran segar memerlukan keputusan yang tepat karena memiliki masa simpan terbatas serta variasi harga satuan dan jumlah penjualan. Penelitian ini bertujuan melakukan segmentasi transaksi penjualan sayuran pada supplier sayur D2 Vegetables menggunakan K-Means Clustering berdasarkan variabel harga satuan dan jumlah penjualan, mengevaluasi kualitas kluster, serta menghasilkan rekomendasi operasional berdasarkan karakteristik setiap segmen. Data penelitian terdiri atas 1.016 transaksi penjualan sayuran periode 25 Juni 2024–23 Juni 2025. Tahapan analisis meliputi prapemrosesan data, standardisasi, penentuan jumlah kluster menggunakan Elbow Method, evaluasi menggunakan Silhouette Coefficient dan Davies-Bouldin Index, penerapan K-Means, serta perbandingan dengan Agglomerative Hierarchical Clustering. Evaluasi internal menunjukkan bahwa $K=2$ memperoleh nilai Silhouette tertinggi sebesar 0,6114 dan Davies-Bouldin Index terendah sebesar 0,5156. Namun, $K=3$ dipilih dengan mempertimbangkan pola Elbow, kualitas struktur kluster, dan interpretabilitas hasil menjadi tiga segmen transaksi. Pada $K=3$, K-Means memperoleh Silhouette Coefficient sebesar 0,5719 dan Davies-Bouldin Index sebesar 0,6029, lebih baik dibandingkan Agglomerative Clustering yang memperoleh 0,5031 dan 0,6595. Hasil segmentasi membentuk segmen Standar sebanyak 682 transaksi (67,13%), Massal sebanyak 235 transaksi (23,13%), dan Premium sebanyak 99 transaksi (9,74%). Segmen Massal memiliki total jumlah penjualan tertinggi, segmen Standar memiliki jumlah transaksi dan keragaman jenis sayur terbesar, sedangkan segmen Premium memiliki rata-rata harga satuan tertinggi dan jumlah penjualan relatif rendah. Hasil segmentasi dapat digunakan sebagai dasar pendukung keputusan pengadaan, pengelolaan stok, penataan penyimpanan, dan strategi promosi sesuai karakteristik masing-masing segmen transaksi.

Kata kunci: K-Means Clustering, segmentasi transaksi, sayuran segar, persediaan, evaluasi kluster

Abstract

Managing fresh vegetable inventory requires accurate decision-making because of its limited shelf life and variations in unit prices and sales volumes. This study aims to segment vegetable sales transactions at D2 Vegetables, a vegetable supplier, using K-Means Clustering based on unit price and sales volume; evaluate cluster quality; and develop operational recommendations based on the characteristics of each segment. The dataset consists of 1,016 vegetable sales transactions recorded from June 25, 2024, to June 23, 2025. The analytical procedures included data preprocessing, standardization, determination of the optimal number of clusters using the Elbow Method, evaluation using the Silhouette Coefficient and Davies-Bouldin Index, K-Means clustering, and comparison with Agglomerative Hierarchical Clustering. Internal evaluation indicated that $K = 2$ achieved the highest Silhouette Coefficient (0.6114) and the lowest Davies-Bouldin Index (0.5156). However, $K = 3$ was selected by considering the Elbow pattern, cluster structure quality, and interpretability of the resulting three transaction segments. At $K = 3$, K-Means achieved a Silhouette Coefficient of 0.5719 and a Davies-Bouldin Index of 0.6029, outperforming Agglomerative Clustering, which achieved

<http://sistemasi.ftik.unisi.ac.id>

scores of 0.5031 and 0.6595, respectively. The resulting segmentation comprised 682 Standard transactions (67.13%), 235 Bulk transactions (23.13%), and 99 Premium transactions (9.74%). The Bulk segment had the highest total sales volume, the Standard segment had the largest number of transactions and the greatest diversity of vegetable types, while the Premium segment had the highest average unit price and relatively low sales volume. The segmentation results can serve as a decision-support basis for procurement planning, inventory management, storage allocation, and segment-specific promotional strategies, enabling operational decisions to be aligned with the characteristics of each transaction segment.

Keywords: K-Means Clustering, transaction segmentation, fresh vegetables, inventory, cluster evaluation

1. Pendahuluan

Pengelolaan persediaan sayuran segar memerlukan ketepatan karena memiliki masa simpan terbatas, mudah mengalami penurunan kualitas, serta dipengaruhi oleh ketidakpastian permintaan dan kondisi penyimpanan. Persediaan berlebih dapat meningkatkan risiko kerusakan dan pemborosan, sedangkan kekurangan persediaan dapat menyebabkan kehilangan penjualan. Kondisi tersebut menunjukkan pentingnya pemanfaatan data dalam mendukung pengelolaan persediaan bahan pangan mudah rusak[1].

Data transaksi dapat dimanfaatkan untuk mengidentifikasi pola penjualan melalui pendekatan clustering. K-Means merupakan salah satu metode yang banyak digunakan karena relatif sederhana dan efisien dalam mengelompokkan data berdasarkan kemiripan karakteristik [2]. Namun, hasil clustering perlu dievaluasi karena kualitas pembentukan kluster dipengaruhi oleh jumlah kluster dan kondisi centroid awal. Penggunaan Elbow Method dan Silhouette Coefficient dapat membantu menentukan konfigurasi kluster yang layak dianalisis [3], sedangkan kajian segmentasi menunjukkan bahwa evaluasi menggunakan lebih dari satu metrik diperlukan untuk memperoleh hasil yang lebih kuat [4]. K-Means juga memiliki keterbatasan berupa sensitivitas terhadap centroid awal, kemungkinan konvergensi pada solusi lokal, dan kebutuhan menentukan jumlah kluster sebelum proses pengelompokan [5]. Oleh karena itu, interpretasi hasil perlu mempertimbangkan kualitas teknis sekaligus kegunaannya dalam konteks pengambilan keputusan [6].

Permasalahan tersebut relevan dengan supplier sayur D2 Vegetables, karena transaksi penjualan memiliki variasi harga satuan dan jumlah penjualan yang berbeda. Sebagian transaksi memiliki harga satuan tinggi dengan jumlah penjualan rendah, sebagian memiliki jumlah penjualan tinggi dengan harga satuan relatif rendah, sedangkan transaksi lainnya berada pada karakteristik menengah. Perbedaan tersebut menyebabkan penerapan kebijakan pengadaan, pengelolaan stok, penataan penyimpanan, dan promosi yang sama terhadap seluruh transaksi berpotensi kurang efisien.

Penelitian terdahulu telah menerapkan K-Means pada data pertanian, penjualan, dan persediaan, tetapi sebagian besar masih berfokus pada pembentukan kelompok produktivitas, pola penjualan, atau tingkat persediaan. Analisis yang secara bersamaan mengevaluasi kualitas kluster, menjelaskan jumlah dan proporsi anggota, mengidentifikasi jenis sayur dominan, serta menghubungkan karakteristik setiap segmen dengan kebutuhan operasional transaksi penjualan sayuran segar masih terbatas. Berdasarkan kesenjangan tersebut, pembaruan penelitian ini terletak pada integrasi K-Means dengan Elbow Method, Silhouette Coefficient, dan Davies-Bouldin Index serta analisis karakteristik setiap kluster untuk menghasilkan rekomendasi operasional.

Berdasarkan permasalahan dan kesenjangan penelitian tersebut, penelitian ini diarahkan untuk menjawab tiga permasalahan utama, yaitu bagaimana K-Means Clustering diterapkan pada transaksi penjualan sayuran berdasarkan variabel harga satuan dan jumlah penjualan, bagaimana hasil pengelompokan tersebut membentuk segmen transaksi Premium, Massal, dan Standar serta memiliki kualitas kluster yang dapat dipertanggungjawabkan, dan bagaimana karakteristik setiap segmen dapat dimanfaatkan sebagai dasar pengambilan keputusan dalam pengadaan, pengelolaan stok, penataan penyimpanan, dan strategi promosi.

Penelitian menggunakan 1.016 transaksi penjualan sayuran periode 25 Juni 2024 sampai 23 Juni 2025. Harga satuan digunakan untuk merepresentasikan nilai ekonomi transaksi, sedangkan jumlah penjualan menggambarkan tingkat permintaan dalam setiap transaksi. Penelitian bertujuan membentuk dan mengevaluasi segmen transaksi Premium, Massal, dan Standar serta menghasilkan

<http://sistemasi.ftik.unisi.ac.id>

rekomendasi bagi pengadaan, pengelolaan stok, penataan penyimpanan, dan strategi promosi pada D2 Vegetables.

2. Tinjauan Literatur

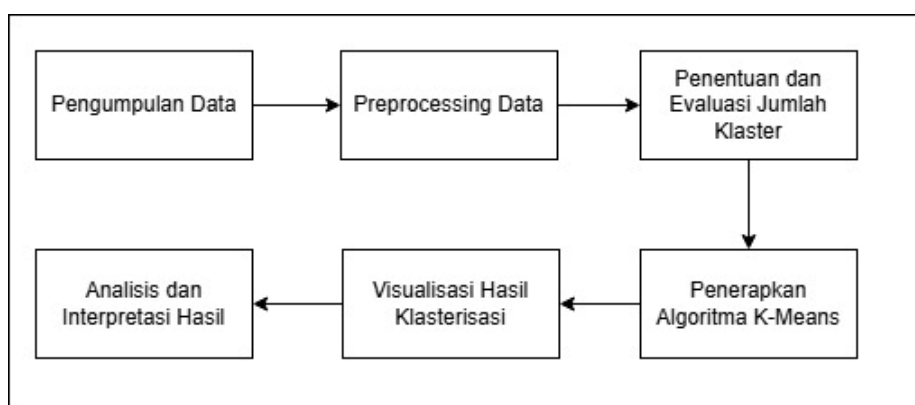
Penerapan K-Means pada bidang pertanian dan penjualan menunjukkan bahwa clustering dapat digunakan untuk mengelompokkan data dengan karakteristik yang beragam. Pada sektor pertanian, K-Means telah diterapkan untuk mengelompokkan sayuran unggulan [7]. Dalam konteks transaksi, metode yang sama digunakan untuk menentukan pola penjualan serta mengubah data historis menjadi kelompok penjualan yang lebih terstruktur [8], [9]. Penggunaan serupa pada pemetaan produktivitas sayuran menunjukkan bahwa K-Means tidak terbatas pada transaksi, tetapi juga dapat digunakan untuk mengidentifikasi perbedaan karakteristik data pertanian berdasarkan wilayah [10].

Pada konteks pangan dan pertanian, K-Means juga digunakan untuk membedakan tingkat perputaran penjualan frozen food [11], sedangkan pengelompokan penjualan sembako digunakan untuk mengidentifikasi kemiripan pola berdasarkan perilaku pembeli [12]. Penelitian lain menunjukkan bahwa K-Means dapat digunakan untuk membaca pola penjualan secara lebih terstruktur [13]. Pada bidang pertanian, metode yang sama juga diterapkan untuk mengelompokkan produksi bawang merah serta produktivitas padi [14], [15]. Temuan tersebut menunjukkan bahwa K-Means dapat diterapkan pada berbagai karakteristik data, tetapi sebagian penerapannya masih berorientasi pada pembentukan kelompok dan pemetaan kondisi data.

Pemanfaatan K-Means juga berkembang ke arah pengelolaan persediaan dan kegiatan operasional. Hasil clustering telah digunakan untuk menentukan prioritas persediaan barang [16], sedangkan pengelompokan data penjualan ritel dapat dimanfaatkan untuk mendukung pengelolaan stok berdasarkan karakteristik penjualan [17]. Pendekatan serupa digunakan dalam penentuan persediaan stok barang dan segmentasi penjualan minimarket [18], [19]. Pada bidang lain, K-Means digunakan dalam analisis pola penjualan obat [20], pengelompokan penjualan kue menggunakan kerangka CRISP-DM [21]. K-Means juga telah diintegrasikan ke dalam sistem berbasis web untuk menganalisis data penjualan toko pakaian [22]. Penelitian-penelitian tersebut menunjukkan bahwa clustering tidak hanya dapat digunakan untuk membaca pola data, tetapi juga dapat memberikan informasi bagi pengelolaan operasional.

Berdasarkan sintesis penelitian terdahulu, penerapan K-Means umumnya mencakup tiga konteks, yaitu pengelompokan data pertanian, identifikasi pola penjualan, dan dukungan terhadap pengelolaan persediaan. Namun, analisis yang menggabungkan evaluasi kualitas kluster, jumlah dan proporsi anggota, karakteristik harga satuan dan jumlah penjualan, serta jenis sayur dominan belum banyak dilakukan secara bersamaan pada transaksi penjualan sayuran segar. Penelitian ini mengisi kesenjangan tersebut dengan mengevaluasi hasil K-Means menggunakan Elbow Method, Silhouette Coefficient, dan Davies-Bouldin Index, kemudian menganalisis karakteristik setiap kluster dan menginterpretasikannya menjadi segmen transaksi Premium, Massal, dan Standar sebagai dasar rekomendasi operasional.

3. Metode Penelitian



Gambar 1 Tahapan penelitian

<http://sistemasi.ftik.unisi.ac.id>

Gambar 1 menunjukkan tahapan penelitian yang digunakan untuk melakukan segmentasi transaksi penjualan sayuran. Penelitian diawali dengan pengumpulan data transaksi dari supplier sayur D2 Vegetables, kemudian dilanjutkan dengan prapemrosesan yang mencakup pemeriksaan kualitas data, pemilihan variabel, dan standardisasi harga satuan serta jumlah penjualan agar memiliki skala yang sebanding. Penentuan jumlah kluster dilakukan menggunakan Elbow Method dan dievaluasi menggunakan Silhouette Coefficient serta Davies-Bouldin Index. Setelah jumlah kluster ditetapkan, K-Means Clustering diterapkan dan hasilnya dibandingkan dengan Agglomerative Hierarchical Clustering untuk menilai kualitas pengelompokan. Kluster yang terbentuk selanjutnya divisualisasikan dan dianalisis berdasarkan karakteristik transaksinya, kemudian diinterpretasikan menjadi segmen Premium, Massal, dan Standar sebagai dasar penyusunan rekomendasi operasional.

Data penelitian berasal dari data transaksi internal penjualan supplier sayur D2 Vegetables selama periode 25 Juni 2024 sampai 23 Juni 2025. Dataset yang digunakan dalam proses analisis terdiri atas 1.016 transaksi penjualan sayuran dengan atribut NO, TANGGAL, NAMA PEMBELI, NAMA SAYUR, JUMLAH, dan HARGA. Data transaksi digunakan karena merepresentasikan aktivitas penjualan aktual selama periode penelitian sehingga dapat digunakan untuk mengidentifikasi perbedaan karakteristik transaksi berdasarkan harga satuan dan jumlah penjualan.

Tabel 1 Atribut data transaksi penjualan sayuran

Atribut	Jenis Data	Peran	Keterangan
NO	Integer/numerik	Identitas Data	Menunjukkan nomor urut transaksi
TANGGAL	Tanggal	Informasi Waktu	Menunjukkan waktu terjadinya transaksi
NAMA PEMBELI	String / kategorikal	Informasi Pendukung	Menunjukkan nama pelanggan yang melakukan pembelian
NAMA SAYUR	String / kategorikal	Identifikasi Objek	Menunjukkan jenis sayur yang dibeli; digunakan pada tahap interpretasi
JUMLAH	Integer/numerik	Variable Utama	Menunjukkan jumlah penjualan; digunakan untuk pembentukan kluster
HARGA	Integer/numerik	Variable Utama	Menunjukkan harga satuan; digunakan untuk pembentukan kluster

Tabel 1 menunjukkan bahwa data penelitian terdiri atas atribut identitas, waktu, informasi pendukung, dan variabel analisis. Variabel harga satuan dan jumlah penjualan digunakan sebagai variabel pembentuk kluster karena merepresentasikan dua dimensi utama transaksi. Harga satuan menggambarkan nilai ekonomi transaksi, sedangkan jumlah penjualan menggambarkan tingkat permintaan dalam setiap transaksi. Kombinasi kedua variabel memungkinkan identifikasi transaksi dengan karakteristik harga satuan tinggi dan jumlah penjualan rendah, jumlah penjualan tinggi dan harga satuan relatif rendah, serta karakteristik yang berada di antara keduanya. Sementara itu, atribut NAMA SAYUR tidak digunakan dalam perhitungan jarak, tetapi tetap dipertahankan untuk mengidentifikasi jenis sayur yang dominan pada setiap segmen setelah proses clustering selesai.

Tahap prapemrosesan dilakukan untuk memastikan data sesuai dengan kebutuhan analisis. Proses ini meliputi pemeriksaan nilai kosong, penyeragaman tipe data numerik, pemeriksaan nilai yang tidak sesuai, serta pemilihan variabel yang digunakan dalam pembentukan kluster. Transaksi dengan jumlah penjualan sebesar nol dikeluarkan karena tidak merepresentasikan penjualan aktual. Setelah tahap prapemrosesan, data yang digunakan dalam analisis berjumlah 1.016 transaksi penjualan sayuran.

Penggunaan harga satuan dan jumlah penjualan memberikan segmentasi yang mudah diinterpretasikan sesuai tujuan penelitian, tetapi belum merepresentasikan seluruh faktor yang dapat memengaruhi pola transaksi. Penelitian ini belum memasukkan variabel seperti frekuensi transaksi, waktu penjualan, margin keuntungan, karakteristik pelanggan, dan pola musiman. Oleh karena itu, hasil segmentasi diinterpretasikan secara khusus berdasarkan hubungan antara harga satuan dan

jumlah penjualan, sedangkan faktor lain tersebut menjadi keterbatasan dan peluang pengembangan pada penelitian selanjutnya.

Proses standardisasi dilakukan menggunakan z-score untuk menyetarakan skala variabel harga satuan dan jumlah penjualan. Standardisasi diperlukan karena K-Means merupakan metode berbasis jarak, sehingga perbedaan rentang nilai antarvariabel dapat menyebabkan variabel dengan skala lebih besar memberikan pengaruh yang lebih dominan terhadap pembentukan klaster. Proses standardisasi dinyatakan pada Rumus (1).

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

- x = nilai data awal
- μ = rata-rata atribut
- σ = simpangan baku atribut
- z = nilai hasil standardisasi

Penentuan dengan x merupakan nilai data awal, μ merupakan rata-rata variabel, σ merupakan simpangan baku, dan z merupakan nilai hasil standardisasi. Melalui proses ini, harga satuan dan jumlah penjualan berada pada skala yang sebanding sehingga memberikan kontribusi yang lebih proporsional terhadap perhitungan jarak. Pemilihan dan karakteristik variabel perlu diperhatikan dalam penerapan K-Means karena dapat memengaruhi struktur klaster yang terbentuk [4].

Penentuan jumlah klaster dilakukan menggunakan Elbow Method dan diperkuat dengan Silhouette Coefficient serta Davies-Bouldin Index. Penggunaan beberapa pendekatan dilakukan karena penentuan jumlah klaster merupakan salah satu aspek penting dalam partitional clustering dan tidak selalu dapat ditentukan secara akurat hanya melalui satu indikator [5]. Pada Elbow Method, nilai Within-Cluster Sum of Squares (WCSS) dihitung untuk nilai $K=1$ sampai $K=10$. WCSS menunjukkan jumlah kuadrat jarak setiap data terhadap centroid klasternya dinyatakan pada Rumus (2).

$$WCSS = \sum_{j=1}^k \sum_{x_i \in C_j} \|x_i - \mu_j\|^2 \quad (2)$$

- k = jumlah klaster
- x_i = data ke- i
- C_j = himpunan data pada klaster ke- j
- μ_j = centroid pada klaster ke- j

Nilai WCSS semakin menurun ketika jumlah klaster bertambah. Kandidat jumlah klaster ditentukan berdasarkan titik ketika penurunan WCSS mulai melandai atau membentuk pola siku (*elbow point*).

Evaluasi selanjutnya dilakukan menggunakan Silhouette Coefficient pada beberapa alternatif jumlah klaster. Silhouette mengukur kesesuaian suatu data terhadap klasternya sendiri dibandingkan dengan klaster terdekat dan dinyatakan pada Rumus (3).

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (3)$$

- $a(i)$ = rata-rata jarak data ke- terhadap anggota dalam klaster yang sama;
- $b(i)$ = rata-rata jarak minimum data ke- terhadap klaster lain;
- $s(i)$ = nilai Silhouette data ke- i .

Nilai Silhouette berada pada rentang -1 sampai 1. Nilai yang semakin mendekati 1 menunjukkan bahwa data memiliki kesesuaian yang lebih baik terhadap klasternya dan memiliki pemisahan yang lebih jelas dari klaster lainnya.

Evaluasi selanjutnya menggunakan Davies-Bouldin Index (DBI). DBI menilai kualitas klaster berdasarkan tingkat kekompakan anggota dalam klaster dan tingkat pemisahan antarklaster. Davies-

Bouldin dan Silhouette termasuk metrik internal yang dapat digunakan untuk mengevaluasi kualitas clustering[5]. Davies-Bouldin Index dinyatakan melalui Rumus (4).

$$DBI = \frac{1}{K} \sum_{i=1}^K \max_{j \neq i} \left(\frac{S_i + S_j}{M_{ij}} \right) \quad (4)$$

- K = jumlah klaster,
- S_i = rata-rata jarak anggota klaster i terhadap centroidnya,
- S_j = rata-rata jarak anggota klaster j terhadap centroidnya,
- M_{ij} = jarak antara centroid klaster i dan j .

Berbeda dengan Silhouette Coefficient, nilai DBI yang lebih kecil menunjukkan kualitas klaster yang lebih baik, karena menunjukkan bahwa klaster memiliki kekompakan internal tinggi dan pemisahan yang lebih jelas.

Dalam penelitian ini, Elbow Method digunakan untuk mengidentifikasi kandidat jumlah klaster berdasarkan perubahan WCSS, sedangkan Silhouette Coefficient dan Davies-Bouldin Index digunakan untuk mengevaluasi kualitas struktur klaster. Penetapan $K=3$ tidak hanya mempertimbangkan hasil metrik internal, tetapi juga pola Elbow dan interpretabilitas hasil sesuai konteks transaksi penjualan sayuran. Nilai masing-masing metrik dan dasar pemilihan tiga klaster disajikan pada bagian Hasil dan Pembahasan.

Setelah jumlah klaster ditetapkan, algoritma K-Means diterapkan pada data yang telah distandardisasi. K-Means merupakan metode partitional clustering yang menempatkan setiap data ke dalam satu klaster berdasarkan kedekatannya dengan centroid. Metode ini banyak digunakan karena memiliki proses yang relatif sederhana dan efisien pada berbagai ukuran data[5]. Jarak antara transaksi dan centroid dihitung menggunakan Euclidean Distance, sebagaimana dinyatakan pada Rumus (5).

$$d(x_i, \mu_k) = \sqrt{\sum_{j=1}^p (x_{ij} - \mu_{kj})^2} \quad (5)$$

- $d(x_i, \mu_k)$ = jarak data ke- i ke centroid ke- k
- p = jumlah variabel
- x_{ij} = nilai data ke- i pada variabel ke- j
- μ_{kl} = nilai centroid ke- k pada variabel ke- j

Setiap transaksi ditempatkan pada centroid terdekat. Setelah seluruh transaksi memperoleh keanggotaan klaster, centroid diperbarui berdasarkan rata-rata seluruh anggota pada klaster tersebut menggunakan Rumus (6).

$$\mu_k = \frac{1}{n_k} \sum_{x_i \in C_k} x_i \quad (6)$$

- μ_k = centroid klaster ke- k ,
- n_k = jumlah anggota klaster ke- k ,
- x_i = anggota pada klaster ke- k .

Parameter K-Means yang digunakan terdiri atas tiga klaster, inisialisasi k-means++, 10 kali proses inisialisasi, dan seed acak 42 untuk menjaga reproduktibilitas hasil. Inisialisasi k-means++ digunakan untuk memperoleh centroid awal yang lebih representatif karena pemilihan centroid awal dapat memengaruhi kestabilan hasil K-Means [2]. Untuk memperkuat dasar pemilihan metode, hasil K-Means dibandingkan dengan Agglomerative Hierarchical Clustering menggunakan Ward linkage pada jumlah klaster yang sama, yaitu $K = 3$. Kedua metode menggunakan data hasil standardisasi yang sama sehingga perbandingan dilakukan dalam kondisi yang setara. Kualitas pengelompokan

dibandingkan menggunakan Silhouette Coefficient dan Davies-Bouldin Index, dengan nilai Silhouette yang lebih tinggi dan Davies-Bouldin yang lebih rendah menunjukkan struktur kluster yang lebih baik.

Visualisasi hasil K-Means dilakukan menggunakan scatter plot untuk menggambarkan persebaran transaksi berdasarkan keanggotaan kluster. Variabel jumlah penjualan ditempatkan pada sumbu horizontal dan harga satuan pada sumbu vertikal, sedangkan centroid digunakan sebagai representasi pusat masing-masing kluster. Visualisasi digunakan untuk membantu mengamati pola persebaran dan perbedaan karakteristik antarkluster, sementara penilaian kualitas kluster tetap didasarkan pada Silhouette Coefficient dan Davies-Bouldin Index.

Setelah proses clustering selesai, karakteristik setiap kluster dianalisis secara deskriptif berdasarkan jumlah transaksi, proporsi anggota terhadap keseluruhan 1.016 transaksi, rata-rata, minimum, dan maksimum harga satuan, serta rata-rata, minimum, maksimum, dan total jumlah penjualan. Analisis juga mencakup jumlah jenis sayur dan lima jenis sayur yang paling dominan berdasarkan frekuensi transaksi pada setiap segmen. Proporsi anggota kluster dihitung menggunakan Rumus (7).

$$P_k = \frac{n_k}{N} \times 100\% \quad (7)$$

- P_k = proporsi kluster ke- k ,
- n_k = jumlah transaksi pada kluster ke k ,
- N = jumlah keseluruhan transaksi.

Analisis karakteristik dilakukan agar hasil clustering tidak hanya menunjukkan persebaran transaksi secara visual, tetapi juga memberikan informasi kuantitatif mengenai ukuran, pola harga satuan, jumlah penjualan, dan keragaman jenis sayur pada setiap kluster. Pendekatan tersebut memungkinkan perbedaan antarkluster diinterpretasikan dalam konteks pengelolaan transaksi penjualan sayuran.

Label numerik yang dihasilkan K-Means tidak secara langsung memiliki makna operasional. Oleh karena itu, penamaan segmen dilakukan setelah karakteristik setiap kluster dianalisis. Kluster dengan rata-rata harga satuan tertinggi diinterpretasikan sebagai segmen Premium, kluster dengan rata-rata jumlah penjualan tertinggi sebagai segmen Massal, sedangkan kluster dengan karakteristik di antara keduanya sebagai segmen Standar. Label tersebut merepresentasikan karakteristik transaksi, bukan klasifikasi permanen terhadap jenis sayur. Interpretasi juga mempertimbangkan proporsi anggota, rentang nilai, total jumlah penjualan, jumlah jenis sayur, dan jenis sayur dominan karena hasil segmentasi perlu dinilai tidak hanya dari kualitas teknis, tetapi juga dari kegunaannya bagi pengambilan keputusan [6].

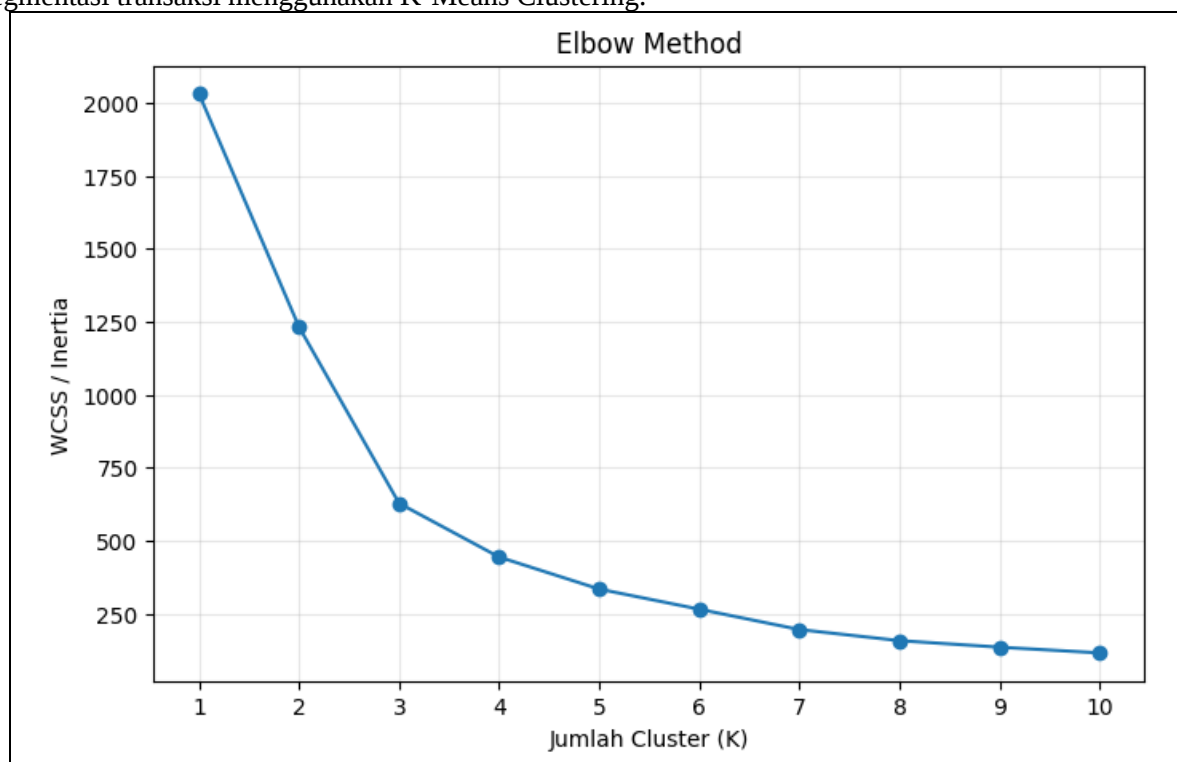
4. Hasil dan Pembahasan

Tahap pengumpulan data menghasilkan data transaksi penjualan sayuran yang berasal dari supplier sayur D2 Vegetables selama periode 25 Juni 2024 sampai 23 Juni 2025. Data yang digunakan merupakan data transaksi internal dan mencakup atribut nomor transaksi, tanggal transaksi, nama pembeli, nama sayur, jumlah penjualan, dan harga satuan. Dari atribut tersebut, harga satuan dan jumlah penjualan dipilih sebagai variabel pembentuk kluster karena merepresentasikan dua karakteristik utama transaksi, yaitu nilai ekonomi dan tingkat permintaan dalam setiap transaksi. Sementara itu, nama sayur tidak digunakan dalam perhitungan jarak K-Means, tetapi tetap dipertahankan sebagai informasi pendukung untuk mengidentifikasi jenis sayur dominan pada setiap segmen transaksi.

Tahap prapemrosesan dilakukan untuk memastikan data yang digunakan telah sesuai dengan kebutuhan analisis clustering. Proses ini meliputi pemeriksaan kelengkapan data, kesesuaian tipe data, nilai yang tidak sesuai, serta transaksi dengan jumlah penjualan nol. Setelah proses tersebut, diperoleh 1.016 transaksi penjualan sayuran yang digunakan dalam analisis. Variabel harga satuan dan jumlah penjualan selanjutnya distandardisasi menggunakan z-score agar memiliki skala yang sebanding sehingga tidak ada variabel yang mendominasi perhitungan jarak pada K-Means. Data hasil

<http://sistemasi.ftik.unisi.ac.id>

prapemrosesan dan standardisasi kemudian digunakan pada tahap penentuan jumlah kluster melalui Elbow Method, Silhouette Coefficient, dan Davies-Bouldin Index sebelum dilakukan pembentukan segmentasi transaksi menggunakan K-Means Clustering.

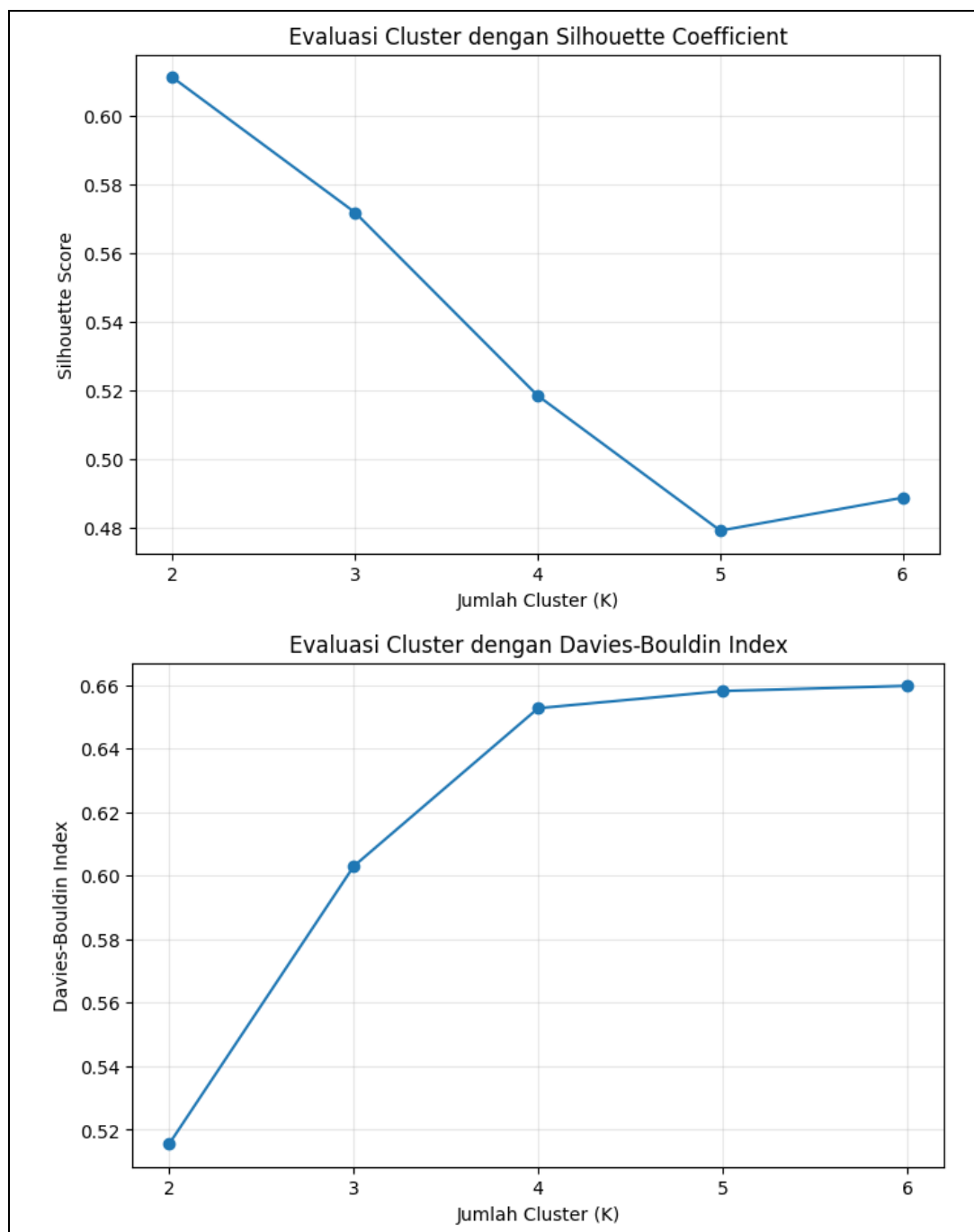


Gambar 3 Grafik metode elbow

Penentuan jumlah kluster dilakukan menggunakan Elbow Method dengan menguji K=1 sampai K=10. Gambar 3 menunjukkan penurunan nilai WCSS yang relatif besar hingga sekitar K=3, kemudian penurunannya mulai melandai. Pola tersebut menunjukkan bahwa K=3 merupakan salah satu kandidat jumlah kluster yang layak dipertimbangkan. Namun, karena penentuan titik siku dapat melibatkan interpretasi visual, evaluasi tidak hanya didasarkan pada Elbow Method, tetapi juga diperkuat menggunakan Silhouette Coefficient dan Davies-Bouldin Index [3]. Evaluasi Silhouette Coefficient dan Davies-Bouldin Index dilakukan pada K=2 sampai K=6. Nilai Silhouette digunakan untuk menilai kekompakan dan pemisahan kluster, dengan nilai yang lebih tinggi menunjukkan struktur kluster yang lebih baik, sedangkan Davies-Bouldin Index menggunakan kriteria sebaliknya, yaitu nilai yang lebih rendah menunjukkan kualitas kluster yang lebih baik.

Tabel 2 Evaluasi jumlah kluster

K	WCSS	Silhouette Coefficient	Davies-Bouldin Index
2	1.232,4032	0,6114	0,5156
3	627,5646	0,5719	0,6029
4	443,1694	0,5185	0,6527
5	333,4192	0,4791	0,6581
6	264,3891	0,4887	0,6597



Gambar 4 Evaluasi jumlah kluster menggunakan silhouette coefficient dan davies-bouldin index

Berdasarkan Tabel 2 dan Gambar 4, nilai Silhouette tertinggi diperoleh pada $K=2$, yaitu 0,6114, sedangkan nilai Davies-Bouldin terendah juga diperoleh pada $K=2$, yaitu 0,5156. Dengan demikian, apabila keputusan hanya didasarkan pada kedua metrik internal tersebut, dua kluster menghasilkan pemisahan yang lebih kuat. Namun, pada $K=3$, nilai Silhouette sebesar 0,5719 tetap positif dengan Davies-Bouldin sebesar 0,6029 dan WCSS menurun cukup besar menjadi 627,5646. Selain itu, grafik Elbow menunjukkan perubahan penurunan WCSS di sekitar $K=3$. Oleh karena itu, penelitian ini tidak menyatakan bahwa seluruh metrik menunjukkan $K=3$ sebagai konfigurasi terbaik, melainkan memilih $K=3$ sebagai kompromi antara kualitas struktur internal, pola Elbow, dan kebutuhan interpretasi transaksi ke dalam tiga karakteristik yang berbeda. Penentuan jumlah kluster memang merupakan

salah satu tantangan pada partitional clustering sehingga evaluasi perlu mempertimbangkan lebih dari satu indikator [5].

Pemilihan tiga klaster juga mempertimbangkan kegunaan hasil segmentasi dalam konteks pengelolaan transaksi penjualan sayuran. Dua klaster akan menghasilkan pembagian yang lebih sederhana, tetapi tidak dapat memisahkan secara khusus kelompok transaksi dengan karakteristik menengah dari dua pola yang lebih ekstrem. Pada $K=3$, transaksi dapat dibedakan menjadi kelompok dengan harga satuan tinggi dan jumlah penjualan rendah, kelompok dengan jumlah penjualan tinggi dan harga satuan relatif rendah, serta kelompok dengan karakteristik di antara keduanya. Pertimbangan konteks tersebut penting karena evaluasi segmentasi tidak hanya berkaitan dengan nilai metrik, tetapi juga dengan kemampuan hasil pengelompokan untuk memberikan informasi yang dapat digunakan dalam pengambilan keputusan [6].

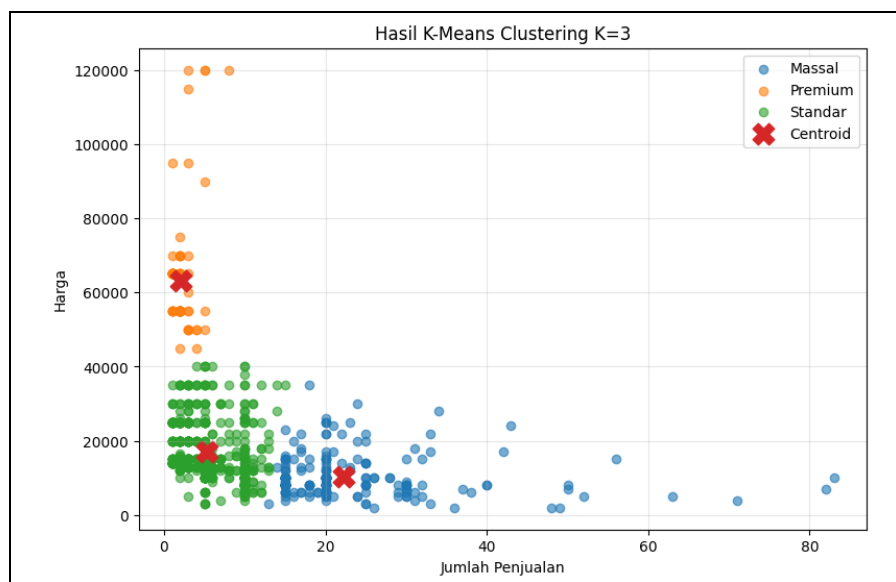
Untuk memperkuat alasan pemilihan K-Means, hasil pengelompokan pada $K=3$ dibandingkan dengan Agglomerative Hierarchical Clustering menggunakan Ward linkage. Kedua metode menggunakan data hasil standardisasi dan jumlah klaster yang sama.

Tabel 3 Perbandingan kualitas klaster K-Means dan agglomerative clustering

Metode	K	Silhouette Coefficient	Davies-Bouldin Index
K-Means	3	0,5719	0,6029
Agglomerative (Ward)	3	0,5031	0,6595

Tabel 3 menunjukkan bahwa pada konfigurasi $K=3$, K-Means menghasilkan kualitas klaster yang lebih baik dibandingkan Agglomerative Clustering. K-Means memperoleh Silhouette Coefficient sebesar 0,5719, lebih tinggi daripada Agglomerative sebesar 0,5031, serta Davies-Bouldin Index sebesar 0,6029, lebih rendah daripada 0,6595. Hasil tersebut menunjukkan bahwa K-Means membentuk struktur tiga klaster yang lebih kompak dan memiliki pemisahan yang lebih baik, sehingga digunakan sebagai metode utama dalam segmentasi transaksi penjualan sayuran.

Hasil K-Means pada $K=3$ divisualisasikan pada Gambar 5 berdasarkan jumlah penjualan pada sumbu horizontal dan harga satuan pada sumbu vertikal. Visualisasi menunjukkan tiga pola transaksi yang berbeda, yaitu segmen Standar yang terkonsentrasi pada harga satuan dan jumlah penjualan rendah hingga menengah, segmen Massal yang memiliki jumlah penjualan relatif tinggi dengan harga satuan lebih rendah, serta segmen Premium yang memiliki harga satuan tinggi dengan jumlah penjualan relatif rendah. Pola tersebut menunjukkan bahwa kombinasi harga satuan dan jumlah penjualan mampu membedakan karakteristik transaksi secara lebih terstruktur.



Gambar 5 Visualisasi hasil K-Means Clustering K=3

Gambar 5 digunakan sebagai pendukung interpretasi dan bukan sebagai satu-satunya dasar penilaian kualitas kluster, karena evaluasi telah diperkuat melalui Silhouette Coefficient dan Davies-Bouldin Index. Setelah pola persebaran transaksi teridentifikasi, setiap kluster dianalisis lebih lanjut berdasarkan jumlah dan proporsi transaksi, rata-rata serta rentang harga satuan, rata-rata dan total jumlah penjualan, serta jumlah jenis sayur. Analisis ini digunakan untuk menjelaskan karakteristik masing-masing segmen secara kuantitatif dan menjadi dasar penyusunan rekomendasi operasional.

Tabel 4 Karakteristik segmen transaksi

Segmen	Transaksi, Proporsi (%)	Jumlah Penjualan, Rata-rata (Min–Maks)	Total Jumlah Penjualan	Harga Satuan (Rp), Rata-rata (Min–Maks)	Jumlah Jenis Sayur
Standar	682(67,13)	5,28(1-15)	3.601	17.123,90 (3.000–40.000)	32
Massal	235(23,13)	22,29(13-83)	5.237	10.408,51 (2.000–35.000)	15
Premium	99(9,74)	2,06(1-8)	204	63.232,32 (45.000–120.000)	5

Tabel 4 menunjukkan bahwa segmen Standar merupakan kelompok terbesar dengan 682 transaksi atau 67,13% dari keseluruhan data, memiliki rata-rata jumlah penjualan 5,28, rata-rata harga satuan Rp17.123,90, serta mencakup 32 jenis sayur. Karakteristik tersebut menunjukkan bahwa segmen Standar merepresentasikan pola transaksi yang paling umum dan memiliki keragaman jenis sayur paling tinggi pada D2 Vegetables. Rentang jumlah penjualan 1–15 juga memperlihatkan bahwa transaksi pada segmen ini cenderung berada pada tingkat rendah hingga menengah. Kondisi tersebut mengindikasikan bahwa segmen Standar memerlukan pengelolaan stok secara rutin dan relatif stabil karena mencakup sebagian besar transaksi sekaligus variasi jenis sayur yang paling luas.

Segmen Massal terdiri atas 235 transaksi atau 23,13% dari keseluruhan data, tetapi menghasilkan total jumlah penjualan tertinggi sebesar 5.237 dengan rata-rata 22,29 per transaksi dan rata-rata harga satuan Rp10.408,51. Kondisi ini menunjukkan bahwa proporsi transaksi yang lebih kecil tidak selalu menghasilkan total jumlah penjualan yang lebih rendah karena setiap transaksi pada segmen Massal cenderung melibatkan jumlah penjualan yang lebih besar. Dibandingkan dengan segmen Standar yang memiliki jumlah transaksi jauh lebih banyak tetapi total jumlah penjualan sebesar 3.601, segmen Massal memperlihatkan pola transaksi dengan intensitas pembelian yang lebih tinggi. Pola tersebut

sejalan dengan penelitian sebelumnya yang menunjukkan bahwa clustering dapat membedakan kelompok berdasarkan tingkat perputaran penjualan [11].

Implikasi operasional segmen Massal adalah perlunya perhatian lebih terhadap ketersediaan stok dan frekuensi pengisian kembali. Tingginya jumlah penjualan per transaksi meningkatkan konsekuensi apabila persediaan tidak mencukupi ketika permintaan terjadi. Oleh karena itu, transaksi pada segmen Massal dapat menjadi indikator bagi pengelola untuk memprioritaskan ketersediaan jenis sayur yang sering muncul pada segmen tersebut. Temuan ini sejalan dengan penggunaan K-Means dalam pengelolaan persediaan yang menunjukkan bahwa karakteristik kelompok penjualan dapat digunakan sebagai dasar menentukan prioritas pengadaan [16].

Segmen Premium merupakan kelompok terkecil dengan 99 transaksi atau 9,74%, memiliki rata-rata harga satuan tertinggi sebesar Rp63.232,32, rata-rata jumlah penjualan hanya 2,06, dan total jumlah penjualan sebesar 204. Karakteristik harga satuan tinggi dan jumlah penjualan rendah menunjukkan bahwa transaksi pada segmen Premium memiliki pola yang berbeda secara jelas dibandingkan segmen Standar dan Massal. Dalam konteks pengelolaan sayuran segar, kondisi tersebut mengindikasikan perlunya pengadaan yang lebih terkendali agar persediaan tidak melebihi kebutuhan transaksi, terutama karena sayuran segar memiliki masa simpan terbatas. Temuan tersebut memperluas penggunaan clustering untuk pengelolaan stok ritel yang sebelumnya memanfaatkan karakteristik penjualan sebagai dasar penyusunan kebijakan persediaan[17].

Secara keseluruhan, perbedaan ketiga segmen menunjukkan bahwa kebijakan pengelolaan tidak sebaiknya diterapkan secara seragam pada seluruh transaksi. Segmen Standar lebih relevan untuk pengelolaan stok rutin karena memiliki frekuensi transaksi dan keragaman jenis sayur tertinggi, segmen Massal lebih membutuhkan perhatian terhadap ketersediaan stok karena memiliki jumlah penjualan per transaksi dan total jumlah penjualan tertinggi, sedangkan segmen Premium memerlukan pengadaan yang lebih selektif karena memiliki harga satuan tinggi dengan jumlah penjualan relatif rendah. Pada penelitian ini, interpretasi tidak hanya didasarkan pada tinggi atau rendahnya jumlah penjualan, tetapi juga mempertimbangkan proporsi transaksi, harga satuan, total jumlah penjualan, rentang nilai, dan keragaman jenis sayur sehingga setiap segmen dapat memberikan dasar rekomendasi operasional yang lebih spesifik.

Tabel 5 Lima jenis sayur dominan pada setiap segmen

Segmen	Peringkat	Jenis Sayur	Jumlah Transaksi	Proporsi dalam Segmen (%)
Premium	1	Basil	40	40,40
Premium	2	Thyme	25	25,25
Premium	3	Ketumbar	23	23,23
Premium	4	Selada Merah	6	6,06
Premium	5	Parsley	5	5,05
Massal	1	Lotus	74	31,49
Massal	2	Spinach	51	21,70
Massal	3	Lobak	40	17,02
Massal	4	Zukini	17	7,23
Massal	5	Selada Merah	14	5,96
Standar	1	Kucai	83	12,17
Standar	2	Ketumbar	72	10,56
Standar	3	Spinach	72	10,56
Standar	4	Pasley	56	8,21
Standar	5	Rosemari	52	7,62

Tabel 5 menunjukkan perbedaan konsentrasi jenis sayur pada setiap segmen. Seluruh 99 transaksi segmen Premium berasal dari lima jenis sayur, dengan Basil sebagai jenis yang paling dominan sebesar 40,40%, sedangkan segmen Standar memiliki 32 jenis sayur dengan lima jenis teratas mencakup sekitar 49% transaksi. Hal ini menunjukkan bahwa segmen Premium lebih

terkonsentrasi pada jenis sayur tertentu, sedangkan segmen Standar memiliki keragaman yang lebih tinggi. Pada segmen Massal, Lotus menjadi jenis sayur paling dominan sebesar 31,49%, diikuti Spinach 21,70% dan Lobak 17,02%. Lima jenis sayur teratas mencakup lebih dari 80% transaksi, sehingga ketersediaannya perlu mendapat perhatian lebih dalam pengelolaan stok. Pemanfaatan clustering untuk mendukung keputusan persediaan juga telah diterapkan pada penelitian sebelumnya [18].

Label Premium, Massal, dan Standar pada penelitian ini merepresentasikan karakteristik transaksi penjualan, bukan klasifikasi permanen terhadap jenis sayur. Hal ini terlihat dari beberapa jenis sayur yang muncul pada lebih dari satu segmen karena setiap transaksi dikelompokkan berdasarkan kombinasi harga satuan dan jumlah penjualan. Dengan demikian, jenis sayur dominan pada suatu segmen menunjukkan frekuensi kemunculannya dalam transaksi dengan karakteristik tertentu, bukan menunjukkan bahwa jenis sayur tersebut secara tetap termasuk dalam kategori Premium, Massal, atau Standar.

Secara operasional, ketiga segmen memerlukan pendekatan pengelolaan yang berbeda. Segmen Massal dapat menjadi prioritas ketersediaan stok karena memiliki rata-rata dan total jumlah penjualan tertinggi. Segmen Standar memerlukan pengelolaan stok rutin karena mencakup sebagian besar transaksi dan memiliki keragaman jenis sayur paling tinggi. Sementara itu, segmen Premium memerlukan pengadaan yang lebih terkendali karena memiliki jumlah penjualan relatif rendah dengan harga satuan tinggi. Pemanfaatan hasil clustering sebagai pendukung pengelolaan persediaan juga ditemukan pada penelitian sebelumnya [20], sedangkan penelitian ini menambahkan evaluasi kualitas klaster dan analisis karakteristik transaksi sayuran segar secara lebih rinci. Hasil penelitian ini tidak hanya membentuk tiga klaster, tetapi juga menunjukkan bahwa evaluasi kualitas klaster, karakteristik kuantitatif, jenis sayur dominan, dan interpretasi operasional dapat digunakan secara bersama untuk menghasilkan segmentasi transaksi yang lebih informatif.

5. Kesimpulan

Penelitian ini berhasil menerapkan K-Means Clustering pada 1.016 transaksi penjualan sayuran di supplier sayur D2 Vegetables berdasarkan variabel harga satuan dan jumlah penjualan. Penentuan jumlah klaster menggunakan Elbow Method, Silhouette Coefficient, dan Davies-Bouldin Index menunjukkan bahwa $K=2$ memiliki nilai evaluasi internal terbaik, tetapi $K=3$ dipilih dengan mempertimbangkan pola Elbow dan interpretabilitas hasil dalam konteks operasional. Pada $K=3$, K-Means memperoleh Silhouette Coefficient sebesar 0,5719 dan Davies-Bouldin Index sebesar 0,6029, lebih baik dibandingkan Agglomerative Hierarchical Clustering yang memperoleh 0,5031 dan 0,6595, sehingga K-Means lebih sesuai untuk membentuk tiga klaster pada data penelitian ini.

Hasil segmentasi membentuk tiga segmen transaksi, yaitu Standar sebanyak 682 transaksi (67,13%), Massal sebanyak 235 transaksi (23,13%), dan Premium sebanyak 99 transaksi (9,74%). Segmen Standar memiliki jumlah transaksi dan keragaman jenis sayur tertinggi, segmen Massal memiliki rata-rata dan total jumlah penjualan tertinggi, sedangkan segmen Premium memiliki rata-rata harga satuan tertinggi dengan jumlah penjualan relatif rendah. Secara operasional, segmen Massal dapat diprioritaskan dalam menjaga ketersediaan stok, segmen Standar memerlukan pengelolaan stok rutin, sedangkan segmen Premium memerlukan pengadaan yang lebih terkendali. Dengan demikian, kontribusi penelitian ini terletak pada penggabungan evaluasi kualitas klaster, analisis karakteristik transaksi, identifikasi jenis sayur dominan, dan penerjemahan hasil segmentasi menjadi rekomendasi operasional.

Penelitian ini masih terbatas pada dua variabel, yaitu harga satuan dan jumlah penjualan, serta menggunakan data dari satu supplier dalam satu periode penelitian. Faktor seperti pola musiman, frekuensi transaksi, lead time pengadaan, margin keuntungan, dan perilaku pelanggan belum dianalisis. Penelitian selanjutnya dapat memperluas variabel, periode, dan sumber data serta membandingkan metode clustering lain untuk menghasilkan segmentasi transaksi yang lebih komprehensif.

Referensi

- [1] D. G. Gioia, L. K. Felizardo, dan P. Brandimarte, “Simulation-based Inventory Management of Perishable Products Via Linear Discrete Choice Models,” *Comput. Oper. Res.*, Vol. 157, Sep 2023, DOI: 10.1016/j.cor.2023.106270.
- [2] L. Xiao dan J. Zhang, “A New Customer Segmentation Framework based on the Gaussian Mixture Model and ASSUMPTION-FREE K-MC2,” *Electronics (Switzerland)*, Vol. 13, No. 17, Sep 2024, DOI: 10.3390/electronics13173523.
- [3] M. S. E. Kasem, M. Hamada, dan I. Taj-Eddin, “Customer Profiling, Segmentation, and Sales Prediction using AI in Direct Marketing,” *Neural Comput. Appl.*, Vol. 36, No. 9, hlm. 4995–5005, Mar 2024, DOI: 10.1007/s00521-023-09339-6.
- [4] M. Alves Gomes dan T. Meisen, “A Review on Customer Segmentation Methods for Personalized Customer Targeting in E-Commerce Use Cases,” *Information Systems and e-Business Management*, Vol. 21, No. 3, hlm. 527–570, Sep 2023, DOI: 10.1007/s10257-023-00640-4.
- [5] A. Kaur, Y. Kumar, dan J. Sidhu, “Exploring Meta-Heuristics for Partitional Clustering: Methods, Metrics, Datasets, and Challenges,” *Artif. Intell. Rev.*, Vol. 57, No. 10, Okt 2024, DOI: 10.1007/s10462-024-10920-1.
- [6] J. Salminen, M. Mustak, M. Sufyan, dan B. J. Jansen, “How can Algorithms Help in Segmenting Users and Customers? A Systematic Review and Research Agenda for Algorithmic Customer Segmentation,” *Journal of Marketing Analytics*, Vol. 11, No. 4, hlm. 677–692, Des 2023, DOI: 10.1057/s41270-023-00235-5.
- [7] L. M. Harahap, W. Fuadi, L. Rosnita, E. Darnila, dan R. Meiyanti, “Klastering Sayuran Unggulan menggunakan Algoritma K-Means,” *Jurnal Teknik Informatika dan Sistem Informasi*, Vol. 8, No. 3, hlm. 567–579, Des 2022, DOI: 10.28932/jutisi.v8i3.5277.
- [8] S. Pujiono, R. Astuti, dan F. Muhamad Basysyar, “Implementasi Data Mining untuk menentukan Pola Penjualan Produk menggunakan Algoritma K-Means Clustering,” *Jurnal Mahasiswa Teknik Informatika*, Vol. 8, No. 1, hlm. 615–620, Feb 2024.
- [9] T. T. Alifa, R. Astuti, dan F. M. Basysyar, “Implementasi Data Mining menggunakan Algoritma K-means Clustering dalam Analisis Penjualan Produk,” *Jurnal Mahasiswa Teknik Informatika*, Vol. 8, No. 1, hlm. 602–607, Feb 2024.
- [10] E. Lianita, A. Pratama, dan A. F. Ulva, “Penerapan Metode K-Means Clustering untuk Pemetaan Produktivitas Sayur-Sayuran berbasis Sistem Informasi Geografis di Provinsi Sumatera Utara,” *Jurnal Sistem dan Teknologi Informasi (JustIN)*, Vol. 12, No. 2, hlm. 232–238, Apr 2024, DOI: 10.26418/justin.v12i2.72934.
- [11] T. Amalina, D. B. A. Pramana, dan B. N. Sari, “Metode K-Means Clustering dalam Pengelompokan Penjualan Produk Frozen Food,” *Jurnal Ilmiah Wahana Pendidikan*, Vol. 8, No. 15, hlm. 574–583, Sep 2022, DOI: 10.5281/zenodo.7052276.
- [12] M. H. Adlun, R. Astuti, dan F. M. Basysyar, “Pengelompokan Data Penjualan Sembako berdasarkan Perilaku Pembeli menggunakan Algoritma K-Means Clustering,” *Jurnal Mahasiswa Teknik Informatika*, Vol. 8, No. 4, hlm. 7699–7703, Agu 2024.
- [13] A. Yahya dan R. Kurniawan, “Implementasi Algoritma K-Means untuk Pengelompokan Data Penjualan berdasarkan Pola Penjualan,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, Vol. 5, No. 1, hlm. 350–358, Jan 2025, DOI: 10.57152/malcom.v5i1.1773.
- [14] J. Pahrudin dan S. Mulyeni, “Implementasi Algoritma K-Means Clustering dengan Python untuk Analisis Produksi Bawang Merah di Indonesia,” *SOSIAL: Jurnal Ilmiah Pendidikan IPS*, Vol. 3, No. 4, hlm. 01–15, Okt 2025, DOI: 10.62383/sosial.v3i4.1228.
- [15] E. Wahyudin, R. A. Rudin, Kaslani, dan S. E. Permana, “Penerapan Data Mining Pengelompokan Produktivitas Padi menggunakan Algoritma K-means pada Provinsi Jawa Barat,” *Jurnal Mahasiswa Teknik Informatika*, Vol. 8, No. 1, hlm. 522–528, Feb 2024, [Daring]. Tersedia pada: <https://opendata.jabarprov.go.id/id/dataset/produktivitas>
- [16] A. P. Rizkyandri, Jasmir, dan Y. Arvita, “Implementasi Metode K-Means Clustering untuk menentukan Persediaan Barang pada Toko SS BabyShop,” *Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM)*, Vol. 3, No. 2, hlm. 712–718, Sep 2023, DOI: 10.33998/jakakom.v3i2.

<http://sistemasi.ftik.unisi.ac.id>

- [17] M. C. Gumilang dan I. Tahyudin, “Optimasi Manajemen Stok melalui Pengelompokan Produk Ritel berdasarkan Data Penjualan dengan *K-Means*,” *Jurnal Algoritma*, Vol. 22, No. 2, hlm. 644–655, Nov 2025, DOI: 10.33364/algoritma/v.22-2.2526.
- [18] D. Indriani, B. Irawan, dan A. Bahtiar, “Penerapan *Algoritma K-means Clustering* untuk menentukan Persediaan Stok Barang,” *Jurnal Mahasiswa Teknik Informatika*, Vol. 8, No. 1, hlm. 182–187, Feb 2024.
- [19] I. Frandika, S. Hud Bayor, dan W. Handoko, “Penerapan *K-Means Clustering* untuk Segmentasi Penjualan di Minimarket Mardi dengan Pendakatan Machine Learning,” *Journal Of Computer Science And Technology (JOCSTEC)*, Vol. 3, No. 3, hlm. 174–182, Sep 2025, DOI: 10.59435/jocstec.v3i3.574.
- [20] M. Fajar, N. Rahaningsih, dan R. D. Dana, “Analisis Pola Penjualan Obat di Apotek An-Naafi menggunakan Metode *K-means Clustering*,” *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 1, hlm. 486–492, Feb 2024.
- [21] M. R. Muttaqin, T. I. Hermanto, dan M. A. Sunandar, “Penerapan *K-Means Clustering* dan *Cross-Industry Standard Process For Data Mining (CRISP-DM)* untuk mengelompokan Penjualan Kue,” *KOMPUTASI: Jurnal Ilmiah Ilmu Komputer dan Matematika*, Vol. 19, No. 1, hlm. 38–53, Jan 2022, [Daring]. Tersedia pada: <https://journal.unpak.ac.id/index.php/komputasi>
- [22] W. T. Pambudi dan A. Witanti, “Penerapan *Algoritma K-means* untuk Menganalisis Data Penjualan pada Toko Ayu Collection berbasis WEB,” *Jurnal Sistem Informasi dan Bisnis Cerdas (SIBC)*, Vol. 15, No. 1, hlm. 1–9, Feb 2022.