

Sales Performance Analysis untuk Peningkatan Kualitas Produk Menggunakan K-Means dan Hierarchical Clustering

Umi Meganinditya Wulandari^{*1}, Muhammad Kholilurrahman², Arif Iman Anshori³

Program Studi Informatika, Fakultas Sains dan Teknologi,
Institut Teknologi dan Sains Nahdlatul Ulama Pekalongan

Email: ^{*1}nindityaw@gmail.com, ²ti15.0003@gmail.com, ³arifaan82@gmail.com

(Naskah masuk: 8 September 2025, diterima untuk diterbitkan: 30 Desember 2025)

Abstrak: Industri fashion, khususnya toko distro, menghadapi persaingan yang semakin ketat sehingga kualitas dan inovasi produk menjadi faktor utama dalam mempertahankan daya saing. Salah satu tantangan yang sering muncul adalah minimnya pemanfaatan data transaksi penjualan sebagai sumber informasi strategis dalam menilai performa produk. Penelitian ini membahas pengelompokan produk berdasarkan performa penjualan menggunakan teknik data mining serta membandingkan efektivitas dua metode clustering, yaitu K-Means dan Hierarchical Agglomerative Clustering. Data yang digunakan berasal dari transaksi penjualan Toko Distro Sextors, dengan tahapan meliputi praproses data, penerapan algoritma clustering, dan evaluasi hasil menggunakan metrik silhouette score. Tujuan penelitian adalah mengidentifikasi kelompok produk dengan performa tinggi, margin keuntungan optimal, serta tingkat penerimaan pasar yang baik, sekaligus menentukan metode clustering yang paling sesuai. Hasil analisis menunjukkan bahwa algoritma K-Means menghasilkan 4 kluster, sedangkan Hierarchical Agglomerative Clustering menghasilkan 3 kluster. Perbedaan jumlah dan karakteristik kluster dari kedua algoritma memberikan gambaran mengenai variasi pola penjualan produk. Temuan ini diharapkan dapat memberikan rekomendasi praktis bagi pelaku UMKM untuk mengambil keputusan berbasis data, sehingga strategi bisnis dapat lebih tepat sasaran dan meningkatkan daya saing produk di pasar.

Kata Kunci – Distro; Data mining; Clustering; K-Means; Hierarchical Clustering

Analyzing Sales Performance to Enhance Product Quality Using K-Means and Hierarchical Clustering

Abstract: The fashion industry, particularly local clothing stores (distro), is facing increasingly intense competition, making product quality and innovation essential for maintaining competitiveness. One key challenge is the limited use of sales transaction data as a strategic source of information to evaluate product performance. This study aims to cluster products based on sales performance using data mining techniques and to compare the effectiveness of two clustering methods, namely K-Means and Hierarchical Agglomerative Clustering. The dataset was obtained from the sales transactions of Sextors Distro Store, with stages including data preprocessing, algorithm implementation, and evaluation of clustering results using the silhouette score metric. The research objectives are to identify product groups with high performance, optimal profit margins, and strong market acceptance, while also determining the most effective clustering method. The results show that the K-Means algorithm produced four clusters, while Hierarchical Agglomerative Clustering produced three clusters. Differences in the number and characteristics of clusters between the two algorithms reveal variations in product sales patterns. These findings provide valuable insights for small and medium enterprises (SMEs) to support data-driven decision-making, enabling more precise business strategies and enhancing product competitiveness in the market.

Keywords – Distro; Data mining; Clustering; K-Means; Hierarchical Clustering

1. PENDAHULUAN

Perindustrian fashion khususnya distro sedang menghadapi persaingan yang sangat ketat. Distro (Distribution Store) merupakan salah satu jenis toko di Indonesia yang menjual aksesoris dan pakaian hasil produksi sendiri maupun orang lain yang bekerjasama dengan distro[1]. Persaingan dalam dunia industri khususnya fashion memiliki pengaruh yang signifikan dengan kualitas produk. Adanya peningkatan kualitas produk dapat menarik perhatian konsumen dan memberikan

nilai tambah yang membedakan produk tersebut dari pesaing di pasar. Kualitas produk secara parsial berpengaruh terhadap keunggulan bersaing, dengan kontribusi sebesar 90,2% terhadap variabel keunggulan bersaing secara simultan bersama inovasi produk[2]. Selain itu, perusahaan dituntut untuk mengembangkan strategi inovasi dan diferensiasi produk guna mempertahankan keunggulan kompetitif. Inovasi produk yang dimiliki oleh produsen berperan signifikan dalam meningkatkan kualitas dan daya tarik produk, sementara diferensiasi membantu perusahaan membedakan tawaran mereka di pasar[3].

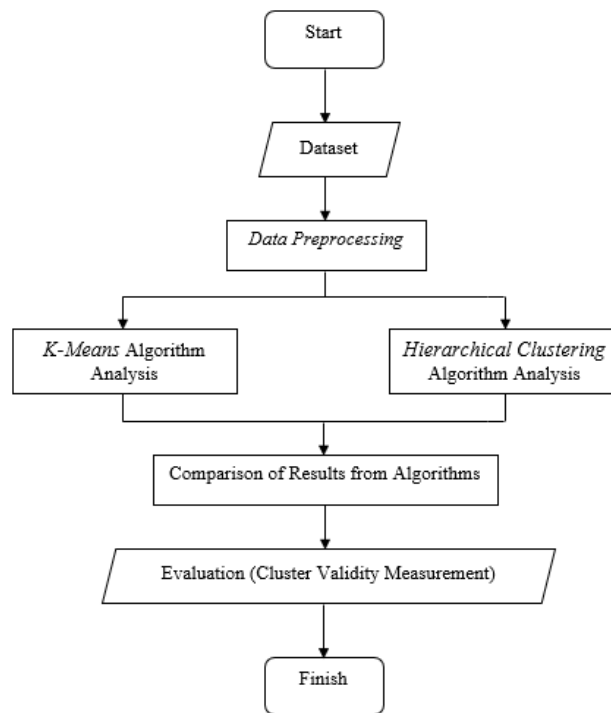
Data transaksi penjualan menyimpan informasi yang jika diolah akan menghasilkan informasi baru sangat berharga contohnya mengenai performa masing-masing produk, termasuk jumlah unit yang terjual, margin keuntungan, dan respons konsumen terhadap program diskon. Sayangnya, potensi strategis dari data penjualan ini sering kali tidak dimanfaatkan secara optimal oleh pelaku usaha[4], terutama di sektor UMKM dan toko distro. Keterbatasan dalam penggunaan alat analisis data serta kurangnya pemahaman tentang metode analitik yang tepat menjadi hambatan utama dalam menggali wawasan bisnis dari data yang sebenarnya sudah tersedia. Analisis terhadap data penjualan memungkinkan pelaku usaha untuk mengevaluasi kualitas produk dari perspektif pasar, bukan hanya dari aspek produksi. Kualitas produk tidak lagi cukup diukur dari bahan atau desain semata, tetapi juga dari seberapa baik produk tersebut diterima oleh konsumen dalam konteks persaingan nyata[5]. Penerapan teknik data mining[6] seperti *clustering*, pelaku usaha dapat mengelompokkan produk berdasarkan performa penjualannya, mengidentifikasi mana yang berkualitas tinggi secara pasar, dan mana yang perlu ditinjau ulang.

Clustering adalah proses pengelompokan data berdasarkan kemiripan karakteristik yang dimilikinya[7]. Penelitian ini mengusulkan menggunakan dua algoritma klustering. Dua algoritma klustering yang digunakan adalah K-Means dan Hierarchical Clustering. Algoritma K-Means adalah metode unsupervised learning yang digunakan untuk mengelompokkan data ke dalam sejumlah cluster berdasarkan kemiripan karakteristiknya[8]. Hierarchical Clustering merupakan metode unsupervised learning yang digunakan untuk mengelompokkan data berdasarkan kemiripan karakteristiknya, dengan membentuk struktur hierarki atau tingkatan kelompok[9]. Ada dua pendekatan pada algoritma ini, namun yang diusulkan pada penelitian ini adalah pendekatan Agglomerative (bottom-up).

Pada tahun 2023, Hakim telah melakukan penelitian menggunakan K-Means pada data penyewaan alat - alat event yang menghasilkan nilai akurasi 86% dan 15%[8]. Pada tahun 2022, Abdullah melakukan penelitian menggunakan algoritma yang sama pada data Coronavirus Disease 2019 (COVID-19) yang menghasilkan 3 kluster provinsi[10]. Penelitian menggunakan algoritma Hierarchical Clustering dilakukan oleh Liu pada tahun 2022 untuk beban dinamis terdistribusi (DDL) yang menunjukkan kelayakan pendekatan identifikasi DDL dan aplikasi yang menjanjikan dalam berbagai struktur dan kondisi pembebanan[11]. Pada tahun 2023, Viktorin melakukan penelitian menggunakan algoritma yang sama untuk menyederhanakan model alokasi tempat pembuangan limbah. Hasil dari penelitian Viktorin menunjukkan pendekatan pemilihan representatif menghasilkan distribusi kontainer yang lebih baik di dalam area dalam formulasi model tujuan tunggal[12].

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan memanfaatkan teknik data mining berbasis unsupervised learning. Algoritma yang digunakan adalah K-Means dan Hierarchical Agglomerative Clustering (HAC) untuk mengelompokkan produk berdasarkan performa penjualannya. Tahapan penelitian dalam studi ini dimulai dari pengumpulan data, preprocessing data, analisis dengan algoritma clustering, hingga evaluasi hasil klusterisasi. Alur penelitian secara keseluruhan ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1. Dataset (Pengumpulan Data)

Data yang digunakan dalam penelitian ini diperoleh dari transaksi penjualan Toko Distro Sextors selama tahun 2021. Dataset tersebut terdiri atas 4940 data transaksi dengan 30 fitur yang berkaitan dengan informasi produk, termasuk jumlah unit yang terjual, margin keuntungan, serta respons konsumen terhadap diskon. Penelitian ini difokuskan pada satu toko karena bersifat studi kasus, sehingga hasil analisis dapat memberikan pemahaman yang lebih mendalam mengenai pola penjualan pada konteks usaha yang spesifik, meskipun dengan keterbatasan generalisasi pada toko lain.

2.2. Preprocessing Data

Tahapan selanjutnya adalah preprocessing data yang bertujuan untuk memastikan kualitas data sebelum dilakukan analisis. Pada tahap ini dilakukan dua proses utama, yaitu data cleaning untuk mengatasi adanya *missing value* dan data yang tidak konsisten, serta data transformation untuk mengubah atribut yang berbentuk nominal maupun ordinal ke dalam bentuk numerik.

2.3. Analisis dan Perbandingan Algoritma K-Means dan Hierarchical Clustering

Analisis klusterisasi dilakukan dengan menggunakan dua algoritma. Algoritma pertama adalah K-Means, yang bekerja dengan membagi data ke dalam sejumlah klaster berdasarkan tingkat kemiripan. Jumlah klaster optimal ditentukan menggunakan metode Elbow, yang mempertimbangkan seberapa baik data terbagi dalam kelompok yang berbeda. Algoritma kedua adalah Hierarchical Agglomerative Clustering (HAC) dengan pendekatan agglomerative (bottom-up). Pada metode ini, setiap data awalnya diperlakukan sebagai satu klaster terpisah, kemudian secara bertahap digabungkan dengan klaster lain yang paling mirip hingga terbentuk satu kelompok besar. Hasil pengelompokan dari HAC divisualisasikan dalam bentuk dendrogram yang menggambarkan hubungan antar produk.

2.4. Evaluasi

Tahap terakhir adalah evaluasi hasil klusterisasi untuk menilai kualitas pengelompokan. Evaluasi dilakukan dengan menggunakan dua metrik, yaitu Silhouette Coefficient (SC) dan Davies-

Bouldin Index (DBI). SC digunakan untuk mengukur kesesuaian data dalam klasternya, sedangkan DBI digunakan untuk menilai keseimbangan antara jarak antar klaster dan kedekatan data dalam klaster. Hasil dari kedua metrik ini digunakan untuk membandingkan performa K-Means dan HAC, sekaligus menentukan metode yang paling sesuai dalam menggambarkan pola performa produk berdasarkan data penjualan.

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Data penelitian diperoleh dari transaksi penjualan Toko Distro Sextors pada tahun 2021. Total terdapat 4940 transaksi dengan 30 atribut awal, di antaranya kode transaksi, kode barang, nama barang, jumlah unit (qty), harga, subtotal, dan biaya pokok penjualan (HPP).

Tabel 1. Data Penjualan

nourut	kode	kode_barang	kd_bar...	nama_barang	...	stok_by_ukuran_warna
1	R43-130421001	SE265	...	Anting	...	FALSE
1	R43-130421002	SE130	...	Sextor Cln Pendek Paspol	...	FALSE
...
1	R43-130421004	SE128	...	Sextor Celana Panjang Paspol	...	FALSE

Tabel 1 memperlihatkan format asli data yang masih mengandung banyak atribut. Tidak semua atribut relevan untuk analisis klasterisasi, sehingga diperlukan tahapan lanjutan berupa preprocessing.

3.2. Data Preprocessing

Data penjualan yang diperoleh kemudian melalui tahap preprocessing untuk memastikan kualitas dan konsistensinya. Proses ini meliputi pembersihan data dari atribut yang tidak relevan serta perbaikan nilai kosong, penambahan atribut margin laba per unit, dan normalisasi variabel numerik agar berada pada skala yang sama. Hasil preprocessing menghasilkan 11 atribut utama yang relevan untuk analisis klasterisasi, dengan fokus pada variabel harga dan margin sebagai indikator performa penjualan.

Tabel 2. Data Hasil Preprocessing

No	Atribut	Hitungan Bukan Nol	Tipe Data
1	nourut	4940 non-null	int64
2	kode	4940 non-null	object
3	kode_barang	4940 non-null	object
4	nama_barang	4940 non-null	object
5	satuan	4940 non-null	object
6	qty	4940 non-null	int64
7	harga	4940 non-null	int64
8	subtotal	4940 non-null	int64
9	hpp	4940 non-null	int64
10	qty_terima	4940 non-null	int64
11	detail_disc_global	4940 non-null	int64

Hasil proses ini ditunjukkan pada Tabel 2, yang memperlihatkan bahwa semua atribut utama sudah terisi penuh (*non-null*). Tabel tersebut memperlihatkan dataset yang telah bersih, konsisten, dan siap digunakan untuk tahap analisis berikutnya.

Setelah proses cleaning, dilakukan standarisasi penamaan kolom menggunakan format snake_case agar konsisten dan mudah diproses dalam bahasa pemrograman. Misalnya, kolom

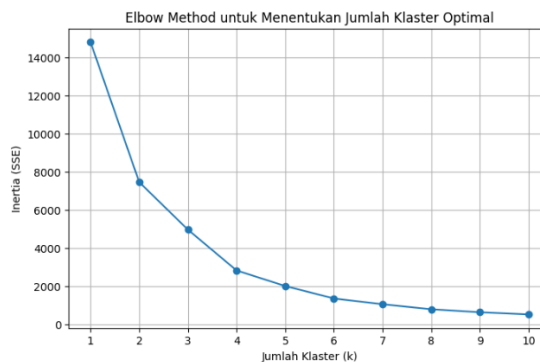
“Nama Barang” diubah menjadi “nama_barang”, sedangkan “Detail-Disc-Global” menjadi “detail_disc_global”.

Selain itu, ditambahkan atribut baru berupa margin laba per unit, yang dihitung dari selisih antara harga jual dan biaya pokok penjualan (HPP). Variabel margin ini penting karena menggambarkan tingkat keuntungan dari setiap produk, sehingga analisis klusterisasi tidak hanya didasarkan pada harga, tetapi juga profitabilitas. Produk dengan margin tinggi dapat diprioritaskan dalam strategi promosi, sementara produk dengan margin rendah perlu dievaluasi kembali dari sisi harga maupun biaya produksi.

Tahap akhir preprocessing adalah normalisasi seluruh variabel numerik agar berada pada skala yang sama. Normalisasi ini mencegah dominasi salah satu variabel terhadap hasil pengelompokan, sehingga kluster yang dihasilkan lebih objektif. Dengan tahapan ini, dataset yang digunakan telah bersih, seragam, dan siap diproses menggunakan algoritma K-Means maupun Hierarchical Clustering.

3.3. Perhitungan Jumlah Kluster Optimal

Penentuan jumlah kluster optimal dilakukan menggunakan Elbow Method, yang didasarkan pada nilai Within-Cluster Sum of Squares (WCSS). Metode ini menghitung total variasi antara data dengan pusat klusternya untuk berbagai nilai k, kemudian diplot dalam grafik. Hasil perhitungan ditunjukkan pada Gambar 2.



Gambar 2. Hasil *Elbow Method*

Gambar 3 memperlihatkan bahwa nilai WCSS menurun seiring bertambahnya jumlah kluster, namun penurunannya mulai melandai pada $k = 4$. Titik siku (elbow) pada grafik tersebut menunjukkan bahwa jumlah kluster optimal adalah empat, karena memberikan keseimbangan antara kualitas pengelompokan dan efisiensi analisis.

3.4. Analisis ALgoritma K-Means

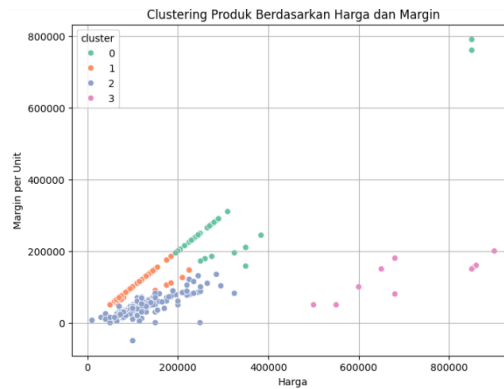
Berdasarkan hasil Elbow Method, jumlah kluster optimal ditetapkan sebanyak empat. Dengan $k = 4$, data dapat terbagi dengan baik tanpa kehilangan informasi penting sekaligus tetap menjaga efisiensi model. Jumlah kluster yang lebih sedikit tidak cukup merepresentasikan variasi data, sedangkan jumlah kluster yang lebih banyak tidak memberikan penurunan inerti yang signifikan. Distribusi jumlah produk pada masing-masing kluster ditunjukkan pada Tabel 3.

Tabel 3. Jumlah Produk per Kluster K-Means

Kluster	Jumlah Produk
0	158
1	833
2	3914
3	35

Tabel 3 menunjukkan bahwa mayoritas produk terkumpul pada kluster 2 dengan 3914 produk, sedangkan kluster 3 hanya berisi 35 produk. Hal ini mengindikasikan dominasi produk dengan harga rendah dan margin kecil dalam dataset. Selain melihat jumlah produk pada tiap kluster, penting juga untuk mengetahui bagaimana posisi produk tersebar dalam ruang analisis

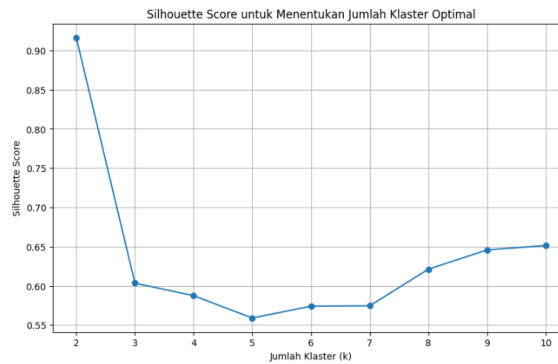
harga dan margin. Visualisasi hasil klastering K-Means berdasarkan kedua variabel tersebut ditampilkan pada Gambar 3.



Gambar 3. Hasil Klastering Algoritma K-Means

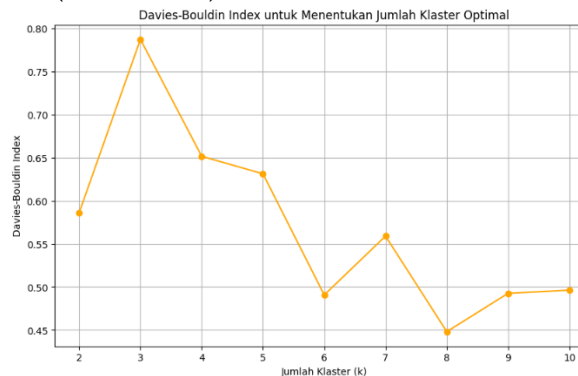
Gambar 4 menunjukkan empat klaster produk yang dibedakan dengan warna berbeda. Klaster 2 berisi produk dengan harga rendah dan margin kecil (produk massal), klaster 1 mencakup produk dengan harga menengah dan margin lebih tinggi (produk kasual), klaster 0 terdiri dari produk dengan harga lebih tinggi dan margin besar (produk premium), sedangkan klaster 3 berisi produk dengan harga sangat tinggi namun margin bervariasi, yang menunjukkan segmen produk eksklusif dengan strategi harga khusus. Klaster 0 didominasi oleh produk pakaian luar dengan harga dan margin menengah, klaster 1 berisi produk kasual dengan harga terjangkau, klaster 2 mencakup aksesoris dan pakaian sederhana dengan harga rendah, sedangkan klaster 3 berfokus pada produk eksklusif dengan harga tinggi.

Proses selanjutnya adalah pengukuran validitas klaster menggunakan Silhouette Coefficient dan Davies-Bouldin Index. Hasil pengukuran ditunjukkan pada Gambar 5 dan Gambar 6.



Gambar 4. Hasil *Silhouette Coefficient*

Gambar 5 memperlihatkan nilai Silhouette tertinggi pada $k = 2$ dengan skor lebih dari 0,90. Hal ini menunjukkan bahwa dua klaster memberikan pemisahan data yang sangat baik. Namun, pemilihan $k = 4$ tetap dipertahankan karena memberikan segmentasi produk yang lebih detail meskipun skornya lebih rendah (sekitar 0,60).

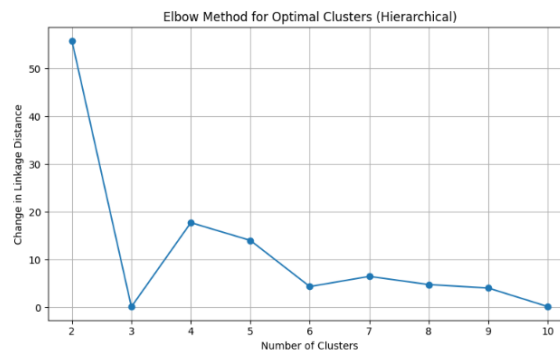


Gambar 5. Hasil *Davies-Bouldin Index*

Gambar 5 menunjukkan bahwa nilai DBI terendah terdapat pada $k = 8$ dengan skor sekitar 0,45. Akan tetapi, nilai DBI pada $k = 4$ masih cukup rendah (sekitar 0,65), sehingga kualitas kluster pada $k = 4$ dapat diterima dan tetap relevan untuk analisis segmentasi produk.

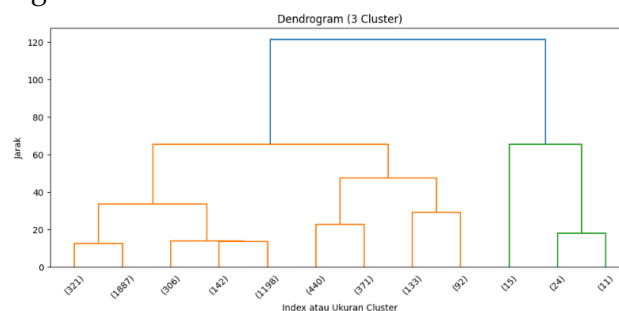
3.5. Analisis Algoritma Hierarchical Clustering

Analisis Hierarchical Clustering dengan pendekatan Ward linkage dilakukan untuk meminimalkan variasi di dalam kluster, sehingga setiap data dalam kluster memiliki kemiripan yang tinggi. Proses dimulai dari data yang sudah dinormalisasi kemudian diproses menggunakan fungsi linkage dengan parameter `method='ward'`, yang menghasilkan struktur hierarki berupa dendrogram. Untuk menentukan jumlah kluster optimal, dilakukan perhitungan nilai Within-Cluster Sum of Squares (WCSS) pada berbagai kemungkinan jumlah kluster. Nilai WCSS ini kemudian divisualisasikan dalam grafik Elbow Method, di mana sumbu-x merepresentasikan jumlah kluster dan sumbu-y menunjukkan nilai WCSS. Melalui grafik tersebut, titik siku (elbow) diidentifikasi sebagai jumlah kluster yang paling sesuai, yaitu saat penurunan nilai WCSS mulai melambat secara signifikan. Dengan cara ini, Elbow Method membantu memastikan pemilihan jumlah kluster yang seimbang, tidak terlalu sedikit sehingga variasi data diabaikan, maupun tidak terlalu banyak sehingga klusterisasi menjadi berlebihan. Oleh karena itu, hasil Hierarchical Clustering lebih mampu merepresentasikan pola alami yang ada pada data. Hasil ditunjukkan pada gambar berikut.



Gambar 6. Hasil *Linkage Ward*

Gambar 6 menunjukkan hubungan antara jumlah kluster dengan perubahan linkage distance. Terlihat bahwa ketika jumlah kluster berkurang dari 10 menjadi 2, nilai linkage distance mengalami penurunan signifikan. Titik siku (elbow) paling jelas tampak pada jumlah kluster tiga, karena setelah titik tersebut penurunan nilai menjadi relatif landai dan tidak signifikan. Dengan demikian, berdasarkan analisis Elbow Method, jumlah kluster optimal untuk data ini adalah tiga, sebagaimana divisualisasikan pada dendrogram berikut.



Gambar 7. Hasil Dendrogram 3 Kluster

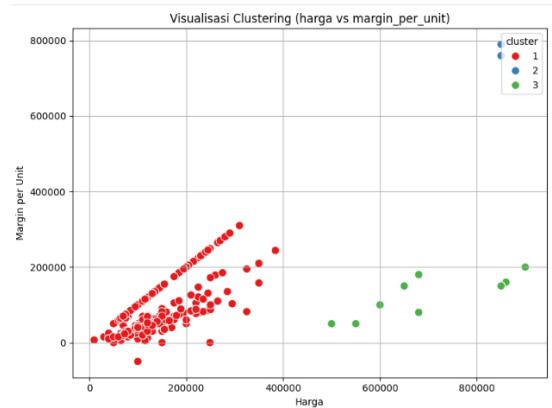
Dendrogram di atas menampilkan hasil Hierarchical Clustering dengan metode Ward linkage yang dipotong menjadi tiga kluster. Sumbu-x menunjukkan indeks data, sedangkan sumbu-y merepresentasikan tingkat kemiripan antar objek. Objek yang berdekatan digabung lebih dahulu, lalu bertahap membentuk kluster lebih besar hingga terbentuk tiga kluster utama. Garis horizontal pada dendrogram menjadi batas pemisah antar kluster, yang ditandai dengan warna berbeda.

Dengan demikian, dendrogram ini memperlihatkan struktur alami data yang terbagi ke dalam tiga kelompok, dengan jumlah produk pada masing-masing kluster ditunjukkan berikut ini.

Tabel 4 Jumlah Produk 3 Kluster

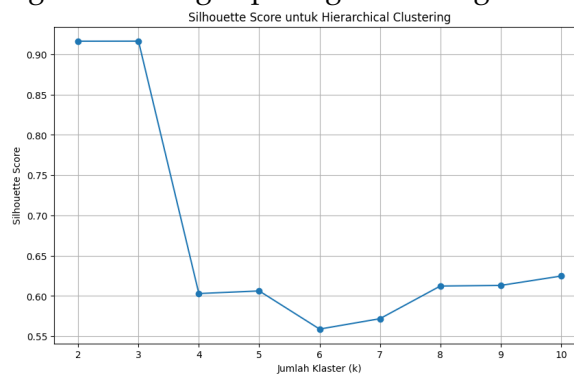
Kluster	Jumlah Produk
0.	4890
1.	35
2.	15

Dari Tabel 4.6 kemudian dibuat grafik berdasarkan garis x sebagai subtotal dan garis y sebagai margin per unit. Berikut gambar hasil klasteringnya.



Gambar 8. Hasil Klastering *Hierarchical Clustering*

Gambar 8 menampilkan hasil visualisasi clustering dengan variabel harga pada sumbu-x dan margin per unit pada sumbu-y. Data terbagi menjadi tiga kluster, yaitu: kluster pertama (merah) berisi mayoritas produk dengan harga rendah hingga menengah dan margin relatif rendah; kluster kedua (biru) terdiri atas produk berharga menengah hingga tinggi dalam jumlah terbatas; dan kluster ketiga (hijau) merepresentasikan produk dengan harga sangat tinggi serta margin besar. Pola ini menunjukkan bahwa harga memiliki pengaruh dominan terhadap pembentukan kluster, di mana kenaikan harga cenderung diikuti dengan peningkatan margin.



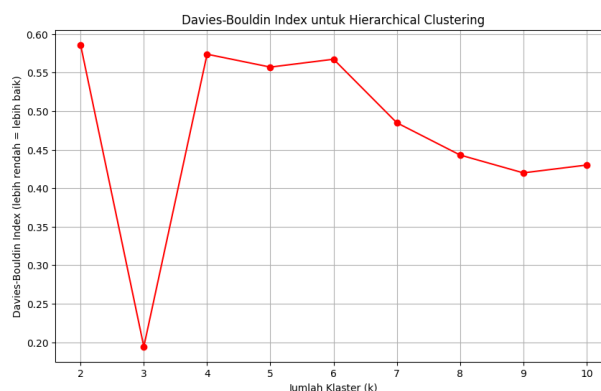
Gambar 9. Hasil *Silhouette Coefficient Hierarchical Clustering*

Visualisasi ini mempermudah identifikasi segmentasi produk berdasarkan harga dan margin keuntungan, serta dapat dijadikan dasar strategi penetapan harga dan pengelolaan produk. Kluster 0 berisi produk-produk dengan variasi umum seperti anting, celana pendek paspol, hingga kaos PROSHOP, yang umumnya memiliki harga dan margin lebih rendah serta dijual dalam satuan. Kluster 1 didominasi oleh produk celana panjang paspol dan turunannya, termasuk variasi dengan kantong bermotif, sehingga merepresentasikan kelompok pakaian dasar dengan desain serupa. Kluster 2 berisi produk dengan kuantitas penjualan berbentuk lusinan, seperti celana panjang tartan dan jins sobek, yang menunjukkan segmentasi produk grosir dengan margin lebih tinggi. Dengan demikian, hasil klasterisasi ini menggambarkan segmentasi produk menjadi tiga kelompok utama: produk eceran berharga rendah, pakaian dengan harga menengah, dan produk grosir bernilai

transaksi lebih besar. Selanjutnya, validitas kluster diuji menggunakan Silhouette Coefficient dan Davies-Bouldin Index untuk menilai kualitas pengelompokan, sebagaimana ditunjukkan pada gambar berikut.

Gambar 9 menunjukkan hasil evaluasi Silhouette Score untuk metode *Hierarchical Clustering* dengan jumlah kluster $k=2$ hingga $k=10$. Nilai Silhouette Score mencerminkan kualitas pemisahan antar kluster, di mana skor mendekati 1 menunjukkan kluster yang semakin jelas dan terpisah. Hasil analisis memperlihatkan bahwa skor tertinggi, sekitar 0,92, dicapai pada $k=2$ dan $k=3$, sehingga kedua pilihan ini memberikan kualitas pengelompokan terbaik. Setelah $k=3$, skor menurun tajam ke kisaran 0,55-0,60, yang menandakan semakin banyak kluster justru menurunkan kualitas pemisahan. Dengan demikian, evaluasi ini menguatkan hasil sebelumnya bahwa jumlah kluster optimal berada pada 2 atau 3 kluster.

Hasil pengukuran validasi kluster menggunakan metode lain, yaitu *Davies-Bouldin Index* ditunjukkan pada gambar berikut.



Gambar 10. Hasil *Davies-Bouldin Index Hierarchical Clustering*

Gambar 10 menyajikan hasil evaluasi Davies-Bouldin Index (DBI) untuk metode *Hierarchical Clustering* dengan jumlah kluster $k=2$ hingga $k=10$. Nilai DBI yang lebih rendah menunjukkan kualitas kluster yang lebih baik karena mengindikasikan kelompok data yang kompak dan terpisah jelas. Hasil analisis menunjukkan nilai terendah, sekitar 0,20, terdapat pada $k=3$, sehingga tiga kluster memberikan pemisahan terbaik dibandingkan jumlah kluster lainnya. Sebaliknya, pada $k=2$ maupun $k \geq 4$, nilai DBI lebih tinggi (0,55-0,60), yang menandakan adanya tumpang tindih antar kelompok. Dengan demikian, evaluasi ini semakin menguatkan hasil Elbow Method bahwa jumlah kluster optimal yang paling representatif untuk data adalah tiga kluster.

3.6. Perbandingan Hasil Kedua Algoritma

Hasil penelitian menunjukkan adanya perbedaan jumlah kluster yang dihasilkan oleh algoritma K-Means dan Hierarchical Clustering. Perbedaan ini terjadi karena perbedaan pendekatan dalam membentuk pengelompokan. K-Means, yang berbasis centroid, membagi data ke dalam empat kluster ($k=4$) sebagaimana ditentukan melalui Elbow Method. Algoritma ini menempatkan data sedekat mungkin dengan pusat kluster sehingga pembagian menjadi lebih granular dan variasi kecil antar kelompok tetap dipisahkan. Sebaliknya, Hierarchical Clustering dengan metode Ward linkage cenderung lebih konservatif, mengikuti struktur alami data, dan menghasilkan tiga kluster ($k=3$) yang lebih kompak serta terpisah jelas.

Perbedaan ini terkait dengan dua hal utama:

1. Pendekatan algoritma, di mana K-Means berbasis geometri sedangkan Hierarchical berbasis jarak antar objek.
2. Tingkat granularitas, di mana K-Means cenderung memecah data lebih detail, sementara Hierarchical menyederhanakan pola alami data.

Evaluasi menggunakan **Silhouette Score** dan **Davies-Bouldin Index** mendukung hasil tersebut. Pada K-Means, $k=4$ menunjukkan Silhouette Score yang tinggi dan Davies-Bouldin Index yang rendah, menandakan pemisahan yang baik dengan kluster yang tetap kompak. Pada Hierarchical, $k=3$ menjadi jumlah kluster optimal dengan Silhouette Score tertinggi dan Davies-Bouldin Index terendah.

Dengan demikian, kombinasi Elbow Method, Silhouette Score, dan Davies-Bouldin Index memberikan hasil konsisten bahwa jumlah kluster optimal berbeda pada tiap algoritma: K-Means optimal pada 4 kluster, sedangkan Hierarchical optimal pada 3 kluster, sesuai dengan karakteristik data dan pendekatan masing-masing metode.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat ditarik beberapa kesimpulan, yaitu Algoritma K-Means menghasilkan 4 kluster optimal berdasarkan Elbow Method, Silhouette Score, dan Davies-Bouldin Index. Hasil ini menunjukkan bahwa metode berbasis centroid mampu memisahkan variasi data secara lebih detail. Algoritma Hierarchical Clustering dengan metode Ward Linkage menghasilkan 3 kluster optimal, yang terbukti lebih representatif secara alami karena membentuk kluster yang lebih kompak dan terpisah jelas. Evaluasi menggunakan Silhouette Score dan Davies-Bouldin Index memperkuat hasil Elbow Method, sehingga jumlah kluster optimal dapat dipastikan sebanyak 4 kluster untuk K-Means dan 3 kluster untuk Hierarchical Clustering. Dalam konteks Sales Performance Analysis, kluster hasil K-Means yang lebih detail bermanfaat untuk mengidentifikasi segmen produk berdasarkan variasi harga, margin, dan pola transaksi, sedangkan kluster hasil Hierarchical Clustering yang lebih sederhana memudahkan manajemen dalam menentukan prioritas peningkatan kualitas produk unggulan dan efisiensi distribusi. Perbedaan hasil antara kedua algoritma memberikan wawasan penting yang dapat saling melengkapi, sehingga mendukung pengambilan keputusan strategis serta peningkatan kualitas produk sesuai kebutuhan pasar. Penelitian selanjutnya dapat dikembangkan dengan membandingkan algoritma lain seperti DBSCAN atau Gaussian Mixture Model (GMM), menggunakan dataset yang lebih besar dan bervariasi, serta menambahkan dimensi waktu untuk menganalisis tren penjualan secara dinamis.

UCAPAN TERIMA KASIH

Ucapan terima kasih yang ditujukan kepada pihak-pihak yang telah berkontribusi dalam penelitian, selain dari author.

DAFTAR PUSTAKA

- [1] A. N. Safitri and R. Basiya, "YUME: Journal of Management Pengaruh Brand Image, Lifestyle, dan Promosi Media Sosial Terhadap Keputusan Pembelian di Toko 3Second," *YUME : Journal of Management*, vol. 5, no. 2, pp. 450–458, 2022, doi: [10.37531/yume.vxix.3463](https://doi.org/10.37531/yume.vxix.3463).
- [2] Supriatun and N. M. I. Pratiwi, "Neraca Pengaruh Inovasi Dan Kualitas Produk Terhadap Keunggulan Bersaing Pada Produk Ekspor Fashion Mahasiswa Program Studi Independen Vokasi Digital Ekspor Fesyen Lacorre The Influence Of Innovation And Product Quality On Competitive Advantage In Fashion Export Products Of Lacorre Digital Vocational Fashion Export Program Students," 2024. [Online]. Available: <http://jurnal.kolibi.org/index.php/neraca>
- [3] F. F. C. Waruwu, H. B. Sidauruk, I. A. Nainggolan, and K. A. Matondang, "INOVASI DAN DIFERENSIASI PRODUK DALAM PASAR PERSAINGAN TIDAK SEMPURNA: IMPLIKASI BAGI KEBIJAKAN PEMASARAN," *MUSYTARI, Neraca Manajemen, Ekonomi*, vol. 11, no. 6, 2024, doi: [10.8734/mnmae.v1i2.359](https://doi.org/10.8734/mnmae.v1i2.359).
- [4] U. M. Wulandari, A. T. Suseno, and M. Kholilurrahman, "Market Basket Analysis Using FP-Growth and Apriori on Distro Store Sales Transaction," *MATICS: Jurnal Ilmu Komputer dan Teknologi Informasi (Journal of Computer Science and Information Technology)*, vol. 17, no. 1, pp. 12–18, Mar. 2025, doi: [10.18860/mat.v17i1.28820](https://doi.org/10.18860/mat.v17i1.28820).
- [5] A. R. Efrat, R. Gernowo, and Farikhin, "Consumer purchase patterns based on market basket analysis using apriori algorithms," *J Phys Conf Ser*, vol. 1524, no. 1, 2020, doi: [10.1088/1742-6596/1524/1/012109](https://doi.org/10.1088/1742-6596/1524/1/012109).

- [6] L. Samboteng, Rulinawaty, M. R. Kasmad, M. Basit, and R. Rahim, "Market Basket Analysis of Administrative Patterns Data of Consumer Purchases Using Data Mining Technology," *Journal of Applied Engineering Science*, vol. 20, no. 2, pp. 339–345, 2022, doi: [10.5937/jaes0-32019](https://doi.org/10.5937/jaes0-32019).
- [7] T. M. Ghazal et al., "Performances of k-means clustering algorithm with different distance metrics," *Intelligent Automation and Soft Computing*, vol. 30, no. 2, pp. 735–742, 2021, doi: [10.32604/iasc.2021.019067](https://doi.org/10.32604/iasc.2021.019067).
- [8] M. A. Hakim, A. B. Prasetijo, and D. Eridani, "Penerapan Data Mining Dengan Algoritma K-Means Clustering Penyewaan Alat-Alat Event Pada Studi Kasus Cv. Dipo Rental Creativindo Implementation Of Data Mining Using The K-Means Clustering Event Rental Algorithm In A Case Study Of Cv. Dipo Rental Creativindo," *Jurnal Teknik Komputer*, vol. 1, no. 4, pp. 148–155, 2023, doi: [10.14710/jtk.v1i4.37011](https://doi.org/10.14710/jtk.v1i4.37011).
- [9] N. Boyko and O. Tkachyk, "Hierarchical clustering algorithm for dendrogram construction and cluster counting," *Informatics and mathematical methods in simulation*, vol. 13, pp. 5–15, May 2023, doi: [10.15276/imms.v13.no1-2.5](https://doi.org/10.15276/imms.v13.no1-2.5).
- [10] D. Abdullah, S. Susilo, A. S. Ahmar, R. Rusli, and R. Hidayat, "The application of K-means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data," *Qual Quant*, vol. 56, no. 3, pp. 1283–1291, Jun. 2022, doi: [10.1007/s11135-021-01176-w](https://doi.org/10.1007/s11135-021-01176-w).
- [11] Y. Liu, L. Wang, M. Li, and Z. Wu, "A distributed dynamic load identification method based on the hierarchical-clustering-oriented radial basis function framework using acceleration signals under convex-fuzzy hybrid uncertainties," *Mech Syst Signal Process*, vol. 172, p. 108935, 2022, doi: <https://doi.org/10.1016/j.ymssp.2022.108935>.
- [12] A. Viktorin, D. Hrabec, V. Nevrlý, R. Šomplák, and R. Šenkeřík, "Hierarchical clustering-based algorithms for optimal waste collection point locations in large-scale problems: A framework development and case study," *Comput Ind Eng*, vol. 178, p. 109142, 2023, doi: <https://doi.org/10.1016/j.cie.2023.109142>.