

Segmentasi Produk Fashion Berdasarkan Harga, Ukuran, dan Merek Menggunakan K-Means di Rapidminer

Ival Sanjaya, Nitami Evita Inonu, Muhammad Fahmi Fudholi, Adelia Pratiwi, Heni Sulistiani*

Fakultas Teknik dan Ilmu Komputer, Magister Ilmu Komputer, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Email: ¹ival_sanjaya@teknokrat.ac.id, ²nitami_evita_inonu@teknokrat.ac.id, ³muhammad_fahmi_fudholi@teknokrat.ac.id,

⁴adelia_pratiwi24@teknokrat.ac.id, ⁵*henisulistiani@teknokrat.ac.id

Email Penulis Korespondensi: henisulistiani@teknokrat.ac.id

Abstrak—Persaingan yang ketat serta keragaman produk menjadi ciri khas industri fashion, terutama dalam hal variasi harga, ukuran, dan merek. Untuk membantu proses pengambilan keputusan bisnis yang lebih akurat, segmentasi produk diperlukan guna mengidentifikasi karakteristik masing-masing kelompok. Studi ini memanfaatkan algoritma K-Means Clustering untuk mengelompokkan produk fashion berdasarkan atribut-atribut tersebut. Implementasi dilakukan menggunakan platform RapidMiner, diawali dengan tahapan normalisasi data dan transformasi atribut kategorikal ke dalam bentuk numerik. Jumlah kluster optimal ditentukan melalui pendekatan elbow method, yang memperlihatkan penurunan signifikan dalam rata-rata jarak antar data dalam kluster. Hasil klusterisasi menunjukkan pembentukan kelompok produk dengan karakteristik yang berbeda, yang dapat dimanfaatkan dalam perencanaan stok dan strategi pemasaran. Penelitian ini menegaskan bahwa algoritma K-Means efektif dalam menganalisis distribusi produk fashion berdasarkan atribut utama yang dimiliki.

Kata Kunci: Industri Fashion, Segmentasi Produk, Klusterisasi, K-Means, Elbow Method, RapidMiner

Abstract—Tight competition and product diversity are the hallmarks of the fashion industry, especially in terms of price variation, size, and brand. To help the process of making more accurate business decisions, product segmentation is needed to identify the characteristics of each group. This study utilizes the K-Means Clustering algorithm to group fashion products based on these attributes. The implementation is carried out using the RapidMiner platform, starting with the data normalization stage and the transformation of categorical attributes into numeric form. The optimal number of clusters is determined through the elbow method approach, which shows a significant decrease in the average distance between data in the cluster. The clustering results show the formation of product groups with different characteristics, which can be utilized in stock planning and marketing strategies. This study confirms that the K-Means algorithm is effective in analyzing the distribution of fashion products based on the main attributes they have.

Keywords: Fashion Industry, Product Segmentation, Clustering, K-Means, Elbow Method, RapidMiner

1. PENDAHULUAN

Industri fashion merupakan salah satu sektor yang berkembang sangat pesat dalam dekade terakhir, didorong oleh kemajuan teknologi digital dan peningkatan daya beli masyarakat, terutama di negara berkembang seperti Indonesia [1]. Fashion tidak hanya dilihat sebagai kebutuhan primer dalam berpakaian, melainkan juga menjadi simbol status sosial, gaya hidup, serta identitas personal seseorang yang terus berevolusi seiring tren [2]. Oleh karena itu, produsen dan pelaku bisnis fashion dituntut untuk mampu merespon perubahan tren dengan cepat serta menyediakan variasi produk yang beragam agar tetap relevan dan kompetitif di pasar global [3].

Dinamika industri fashion sangat kompleks karena melibatkan berbagai variabel yang memengaruhi preferensi konsumen, seperti harga, ukuran, dan merek [4]. Ketiga faktor ini merupakan komponen utama dalam proses pengambilan keputusan konsumen saat memilih produk fashion [5]. Harga menentukan keterjangkauan produk dan persepsi nilai (Liu & Xu, 2023), ukuran mencerminkan kenyamanan dan kesesuaian dengan tubuh konsumen yang bervariasi [6], sedangkan merek menciptakan persepsi kualitas dan prestige terhadap produk tersebut [7]. Variasi yang luas dari ketiga atribut ini membuat pengelolaan portofolio produk menjadi semakin menantang bagi pelaku usaha fashion.

Dalam konteks pengambilan keputusan bisnis, pemahaman mendalam terhadap karakteristik produk menjadi sangat penting. Strategi seperti penetapan harga, manajemen inventori, serta kampanye pemasaran sangat bergantung pada seberapa baik suatu perusahaan memahami struktur dan distribusi produknya. Oleh karena itu, penerapan metode analisis berbasis data, khususnya teknik segmentasi, menjadi sangat relevan untuk menyederhanakan kompleksitas data produk dan mengelompokkannya ke dalam kategori-kategori yang lebih terstruktur dan mudah dianalisis [8].

Segmentasi produk merupakan proses pengelompokan item atau barang berdasarkan atribut-atribut tertentu yang relevan, dengan tujuan untuk memahami karakteristik masing-masing kelompok secara lebih detail [9]. Dalam industri fashion, segmentasi berdasarkan harga, ukuran, dan merek memiliki nilai strategis karena dapat digunakan untuk mengidentifikasi segmen pasar yang potensial, merancang strategi promosi yang tepat sasaran, serta menyesuaikan distribusi stok sesuai permintaan [10]. Segmentasi yang dilakukan secara manual akan sangat memakan waktu dan berisiko tinggi terhadap kesalahan manusia, sehingga dibutuhkan pendekatan otomatisasi yang efisien dan akurat.

Salah satu pendekatan yang populer dalam segmentasi berbasis data adalah algoritma K-Means Clustering. K-Means merupakan metode klusterisasi yang termasuk dalam kategori unsupervised learning, di mana data yang tidak memiliki label akan dikelompokkan berdasarkan kesamaan nilai atributnya [11]. Keunggulan utama dari algoritma ini terletak pada kesederhanaannya, kecepatan dalam proses iterasi, serta kemampuannya dalam menangani data dalam jumlah besar [12]. Prinsip kerja K-Means adalah dengan menentukan pusat kluster (centroid) secara acak, kemudian

menghitung jarak antara setiap data ke centroid terdekat, dan mengelompokkan data berdasarkan jarak tersebut. Proses ini diulangi hingga centroid tidak mengalami perubahan signifikan atau hingga jumlah iterasi maksimum tercapai.

Namun demikian, pemilihan jumlah klaster (K) yang optimal merupakan tantangan tersendiri dalam penerapan K-Means. Jika jumlah klaster yang digunakan tidak sesuai, hasil segmentasi bisa menjadi tidak akurat atau tidak representatif. Untuk mengatasi hal ini, metode validasi seperti Elbow Method digunakan. Elbow Method menganalisis penurunan rata-rata jarak antar data dalam klaster (inertia) terhadap variasi jumlah klaster, dan titik di mana penurunan tersebut mulai melambat (titik "siku") menjadi indikasi jumlah klaster optimal.

Berbagai penelitian sebelumnya telah membuktikan efektivitas K-Means dalam segmentasi, baik untuk data pelanggan, transaksi ritel, hingga produk teknologi. Misalnya, [13] berhasil menggunakan K-Means untuk pengelompokan produk smartphone berdasarkan harga dan spesifikasi, sementara [14] menerapkan metode ini untuk segmentasi pelanggan e-commerce. Meski demikian, belum banyak studi yang secara spesifik menggabungkan atribut harga, ukuran, dan merek secara simultan dalam segmentasi produk fashion. Hal ini menjadi celah penelitian yang penting untuk dijawab, mengingat kebutuhan akan strategi segmentasi yang lebih komprehensif dalam industri fashion modern.

Untuk menjawab tantangan tersebut, penelitian ini menggunakan pendekatan praktis berbasis aplikasi visual, yaitu RapidMiner. RapidMiner adalah perangkat lunak data science berbasis antarmuka grafis (GUI) yang memungkinkan pengguna melakukan analisis data tanpa perlu menulis kode program [15]. Platform ini sangat cocok digunakan oleh pelaku usaha mikro, kecil, dan menengah (UMKM) yang mungkin belum memiliki kapasitas teknis dalam pemrograman, namun membutuhkan solusi data mining yang efektif dan efisien.

Dengan menggabungkan kekuatan algoritma K-Means dan kemudahan antarmuka RapidMiner, penelitian ini bertujuan untuk membangun model segmentasi produk fashion berdasarkan atribut harga, ukuran, dan merek. Penelitian ini diharapkan mampu memberikan kontribusi praktis bagi industri fashion, khususnya dalam perencanaan strategi pemasaran, pengelolaan stok, dan pengembangan sistem rekomendasi produk. Selain itu, dari sisi akademik, penelitian ini juga memperkaya literatur mengenai penerapan K-Means dalam segmentasi berbasis atribut campuran (numerik dan kategorikal), serta menunjukkan potensi penggunaan alat bantu visual seperti RapidMiner dalam riset data mining.

Secara keseluruhan, tujuan utama penelitian ini adalah untuk menjawab kebutuhan industri fashion terhadap pemodelan segmentasi produk yang akurat, praktis, dan aplikatif. Dengan demikian, hasil penelitian ini tidak hanya diharapkan bermanfaat bagi dunia akademik, tetapi juga dapat diimplementasikan langsung oleh pelaku usaha untuk mendukung pengambilan keputusan bisnis yang berbasis data.

2. METODOLOGI PENELITIAN

Penelitian ini bersifat kuantitatif dengan pendekatan eksploratif yang bertujuan untuk melakukan segmentasi produk fashion menggunakan algoritma K-Means Clustering berdasarkan tiga atribut utama, yaitu harga, ukuran, dan merek. Proses dilakukan secara menyeluruh menggunakan platform RapidMiner yang memudahkan tahapan pengolahan data tanpa memerlukan pemrograman.

2.1. Tahapan Dasar Penelitian



Gambar 1. Kerangka Dasar Penelitian

Tahapan penelitian ini dapat dijelaskan sebagai berikut:

a. Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari dataset publik di Kaggle dengan nama *Fashion Products*

Dataset (Bhanupratap Biswas, 2022). Dataset ini mencakup atribut seperti Product ID, Product Name, Brand, Category, Price, Rating, Color, dan Size.

- b. Pra-pemrosesan Data
 1. Pemilihan Atribut: Dipilih tiga atribut utama yang digunakan dalam segmentasi, yaitu *Price*, *Size*, dan *Brand*.
 2. Encoding Data Kategorikal: Atribut *Size* dan *Brand* yang bersifat kategorikal diubah ke bentuk numerik menggunakan metode *Nominal to Numerical* agar dapat digunakan dalam algoritma K-Means.
 3. Normalisasi Data Numerik: Atribut *Price* dinormalisasi menggunakan metode Z-Score Normalization agar berada dalam skala yang sebanding dengan atribut lainnya.
- c. Penentuan Jumlah Kluster (K)
 1. Digunakan Elbow Method untuk menentukan nilai K optimal. Proses dilakukan dengan menjalankan iterasi K dari 2 hingga 6 dan mengamati penurunan rata-rata jarak antar data ke pusat kluster (inertia).
 2. Titik elbow yang menunjukkan penurunan signifikan namun mulai melambat diidentifikasi sebagai nilai optimal. Pada penelitian ini diperoleh nilai K = 3.
- d. Penerapan Algoritma K-Means
 1. Data yang telah diproses dimasukkan ke operator *K-Means* di RapidMiner dengan nilai K = 3.
 2. Sistem secara otomatis membentuk tiga kluster berdasarkan kesamaan atribut.
- e. Visualisasi dan Interpretasi Hasil
 1. Hasil kluster divisualisasikan dalam bentuk *scatter plot* dengan warna berbeda untuk tiap kluster.
 2. Dilakukan interpretasi terhadap masing-masing kluster berdasarkan nilai centroid dan karakteristik dominan yang muncul dari atribut-atribut yang digunakan.
- f. Penarikan Kesimpulan

Berdasarkan hasil segmentasi, ditarik kesimpulan mengenai karakteristik masing-masing kelompok produk dan aplikasinya dalam strategi bisnis seperti promosi, manajemen stok, dan rekomendasi produk.

2.2. Kajian Pustaka

Bagian ini membahas secara singkat teori atau konsep dasar dari objek-objek yang menjadi fokus penelitian, sebagaimana tercantum pada kata kunci dalam abstrak: industri fashion, segmentasi produk, klusterisasi, K-Means, Elbow Method, dan RapidMiner.

2.2.1. Industri Fashion

Industri fashion adalah sektor ekonomi yang bergerak dalam produksi dan distribusi pakaian, aksesoris, dan produk gaya hidup lainnya. Industri ini bersifat dinamis dan sangat dipengaruhi oleh tren, musim, budaya populer, serta perilaku konsumen [16]. Persaingan yang ketat dalam industri ini mendorong pentingnya analisis data untuk memahami pasar dan merancang strategi yang adaptif.

2.2.2. Segmentasi Produk

Segmentasi produk adalah proses membagi item atau produk ke dalam kelompok-kelompok berdasarkan karakteristik tertentu agar analisis dan strategi bisnis dapat dilakukan secara lebih terarah. Dalam konteks ini, atribut seperti harga, ukuran, dan merek sering digunakan karena berpengaruh langsung terhadap persepsi dan preferensi konsumen [17].

2.2.3. Klusterisasi

Klusterisasi (clustering) adalah salah satu metode dalam data mining yang bertujuan mengelompokkan data berdasarkan kemiripan antar data tersebut. Metode ini bersifat unsupervised, artinya tidak memerlukan label atau kategori sebelumnya, sehingga sangat cocok untuk eksplorasi data awal atau segmentasi produk [18].

2.2.4. Algoritma K-Means

K-Means adalah algoritma klusterisasi yang bekerja dengan membagi data ke dalam sejumlah kluster berdasarkan jarak terdekat terhadap pusat kluster (centroid). Proses dilakukan secara iteratif hingga centroid stabil. Kelebihannya meliputi kecepatan proses, kesederhanaan implementasi, dan efektivitas dalam menangani data besar [19].

2.2.5. Elbow Method

Elbow Method digunakan untuk menentukan jumlah kluster optimal (nilai K) dalam proses K-Means. Metode ini menganalisis grafik inertia terhadap jumlah kluster dan mencari titik "tekuk" (elbow) sebagai indikator bahwa penambahan kluster berikutnya tidak lagi signifikan mengurangi nilai inertia [20].

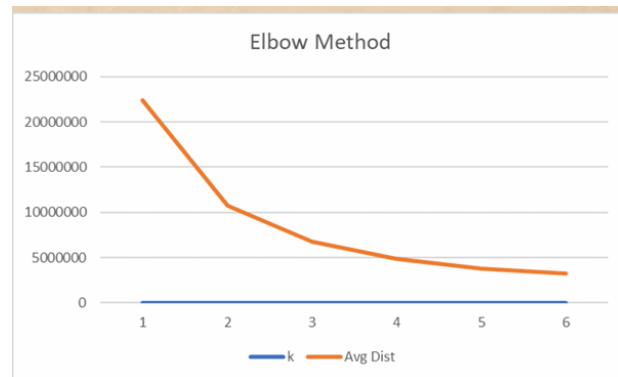
2.2.6. RapidMiner

RapidMiner adalah perangkat lunak data mining berbasis GUI (Graphical User Interface) yang menyediakan berbagai operator analisis data tanpa perlu pengkodean. Platform ini sangat cocok digunakan oleh pemula atau pengguna non-teknis karena proses analisis dapat dilakukan secara visual dan intuitif.

3. HASIL DAN PEMBAHASAN

3.1. Penentuan Jumlah Kluster optimal

Langkah awal dalam penerapan algoritma K-Means adalah menentukan jumlah kluster optimal (nilai K). Hal ini dilakukan untuk memastikan bahwa proses klusterisasi tidak menghasilkan pengelompokan yang terlalu sedikit (under-clustering) atau terlalu banyak (over-clustering). Salah satu metode yang digunakan dalam penelitian ini adalah Elbow Method, yang memvisualisasikan nilai inerti (rata-rata jarak data ke centroid) terhadap jumlah kluster.



Gambar 2. Grafik Elbow Method untuk Penentuan Jumlah Kluster

Gambar 2. Menampilkan grafik Elbow Method dari hasil eksperimen pada rentang K = 2 hingga K = 6. Dari grafik tersebut, terlihat penurunan nilai inerti yang signifikan dari K=2 ke K=3, kemudian penurunan mulai melambat setelahnya. Hal ini menunjukkan bahwa nilai K = 3 adalah titik elbow yang ideal untuk pengelompokan data produk fashion dalam penelitian ini.

3.2. Hasil Clustering Produk Fashion

Setelah nilai K ditentukan, algoritma K-Means dijalankan dengan K = 3. Hasil segmentasi data fashion terbagi ke dalam tiga kluster yang masing-masing memiliki karakteristik unik berdasarkan kombinasi atribut harga, ukuran, dan merek. Proses ini dilakukan secara otomatis oleh RapidMiner, yang kemudian menghasilkan tabel centroid, scatter plot visual, dan hasil penetapan kluster untuk tiap item produk.



Gambar 3. Scatter Plot Visualisasi Hasil Klusterisasi Produk

Gambar 3. Menampilkan visualisasi scatter plot dari hasil klusterisasi yang mengelompokkan produk berdasarkan sumbu X (harga) dan indeks data pada sumbu Y. Warna berbeda digunakan untuk menandai setiap kluster agar interpretasi visual menjadi lebih mudah. Data menunjukkan bahwa kluster dengan warna biru dominan berisi produk dengan harga rendah, kluster hijau mencerminkan harga menengah, dan kluster oranye berisi produk berharga tinggi.

Gambar 3. Menunjukkan Interpretasi Visual, Sumbu X: Price, Menunjukkan rentang harga produk fashion dari sekitar 10 hingga 100. Sumbu Y: Product ID, Digunakan sebagai indeks visualisasi, bukan fitur clustering utama. Ini hanya

membantu memisahkan titik data secara visual. Warna Cluster: Cluster 2 (Biru): Produk dengan harga lebih rendah. Cluster 0 (Hijau): Produk dengan harga menengah. Cluster 1 (Oranye): Produk dengan harga tinggi. Ukuran Bubble : Kemungkinan mencerminkan atribut lain seperti Rating atau Ukuran, tergantung pada ukuran.

3.3. Tabel Centroid Setiap Klaster

Sebagai hasil dari proses data K-Means, diperoleh centroid dari masing-masing klaster yang mewakili karakteristik umum dari tiap kelompok. Centroid ini dihitung berdasarkan rata-rata nilai dari setiap atribut dalam klaster, dan menjadi suatu acuan untuk memahami profil demografis produk yang ada di dalam tabel tersebut yang membuatnya menjadi suatu hasil.

Tabel 1. Centroid Tiap Klaster Berdasarkan Atribut Penentu

Klaster	Rata-Rata Harga	Rata-Rata Ukuran	Rata-Rata Merek
0 (Hijau)	Sedang	Campuran	Lokal/Semi-Premium
1 (Orange)	Tinggi	Besar (L, XL)	Premium/International
2 (Biru)	Rendah	Kecil (S, M)	Lokal

Tabel 1. Berikut menyajikan nilai rata-rata (centroid) dari atribut harga, ukuran (yang telah diencoding), dan merek (yang telah diencoding) untuk masing-masing klaster.

3.4. Interpretasi dan Karakteristik Klaster

Berdasarkan visualisasi dan data centroid, masing-masing klaster dapat dijelaskan sebagai berikut:

- Klaster 0: Produk dengan harga menengah, ukuran beragam (S hingga XL), dan merek campuran antara lokal dan semi-premium. Klaster ini mencerminkan segmen pasar menengah yang cenderung fleksibel dari sisi ukuran dan memilih merek berdasarkan kebutuhan fungsional dan harga terjangkau.
- Klaster 1: Produk dengan harga tinggi, dominan pada ukuran besar (L dan XL), serta didominasi oleh merek internasional atau premium. Produk dalam klaster ini mengarah pada segmen konsumen kelas atas yang mengutamakan kualitas merek dan kenyamanan ukuran besar.
- Klaster 2: Produk dengan harga rendah, ukuran kecil (S dan M), dan berasal dari merek lokal. Klaster ini merepresentasikan konsumen yang mencari produk murah, ringan, dan kemungkinan berusia muda atau remaja.

3.5. Kelebihan dan Implikasi Praktis

Penggunaan RapidMiner memberikan sejumlah kelebihan, terutama dalam kemudahan pengoperasian bagi pengguna non-teknis. Proses encoding, normalisasi, pemilihan parameter, hingga visualisasi dapat dilakukan secara drag-and-drop melalui antarmuka grafis. Hal ini sangat mendukung implementasi algoritma K-Means di lingkungan bisnis seperti UMKM fashion, tanpa memerlukan keahlian pemrograman atau statistik tingkat lanjut. Implikasi praktis dari hasil segmentasi ini mencakup:

- Strategi Promosi: Produk klaster 2 dapat ditawarkan dengan diskon atau program bundling, sedangkan klaster 1 difokuskan pada strategi brand premium.
- Manajemen Stok: Toko yang berada di area dengan dominasi ukuran kecil bisa menekankan stok produk dari klaster 2, sedangkan pusat perbelanjaan menengah ke atas dapat difokuskan pada klaster 1.
- Sistem Rekomendasi: Hasil klaster dapat dijadikan dasar untuk menyarankan produk serupa kepada pelanggan berdasarkan pembelian sebelumnya.

3.6. K-Means di RAPIDMINER

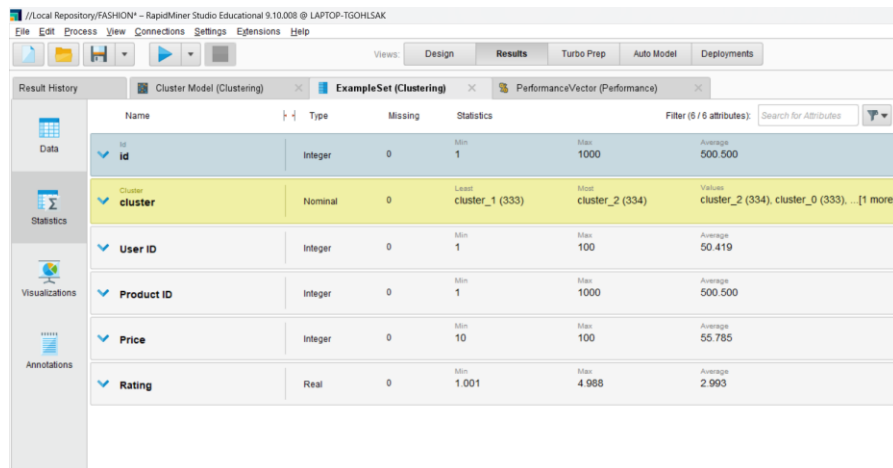
Setelah data produk fashion melalui tahap pra-pemrosesan, algoritma K-Means Clustering diterapkan menggunakan platform RapidMiner. Proses ini mencakup penerapan operator K-Means, konfigurasi jumlah klaster ($K = 3$), serta pengolahan hasil dalam bentuk tabel dan visualisasi grafis. RapidMiner secara otomatis menampilkan hasil pengelompokan dalam berbagai tampilan, yang memudahkan pengguna untuk meninjau hasil klaster secara intuitif.

Gambar dibawah menampilkan tampilan antarmuka dari proses clustering di RapidMiner. Tampilan ini mencakup beberapa panel penting, seperti Assignment Output (hasil penempatan data ke dalam klaster), Centroid Table (nilai rata-rata atribut di tiap klaster), serta Graph Tree dan Scatter Plot yang menggambarkan hubungan antar atribut dalam visualisasi klaster.

Cluster Model

```
Cluster 0: 333 items
Cluster 1: 333 items
Cluster 2: 334 items
Total number of items: 1000
```

Gambar 4. Tampilan Keluaran Clustering K-Means pada RapidMiner



Name	Type	Missing	Statistics	Filter (5 / 6 attributes)
id	Integer	0	Min 1, Max 1000, Average 500.500	
cluster	Nominal	0	Least cluster_1 (333), Most cluster_2 (334), Values cluster_2 (334), cluster_0 (333), ... [1 more]	
User ID	Integer	0	Min 1, Max 100, Average 50.419	
Product ID	Integer	0	Min 1, Max 1000, Average 500.500	
Price	Integer	0	Min 10, Max 100, Average 55.785	
Rating	Real	0	Min 1.001, Max 4.988, Average 2.993	

Gambar 5. Hasil Tampilan Statistics Cluster 0,1,2 yang sudah di proses Rapidminer.



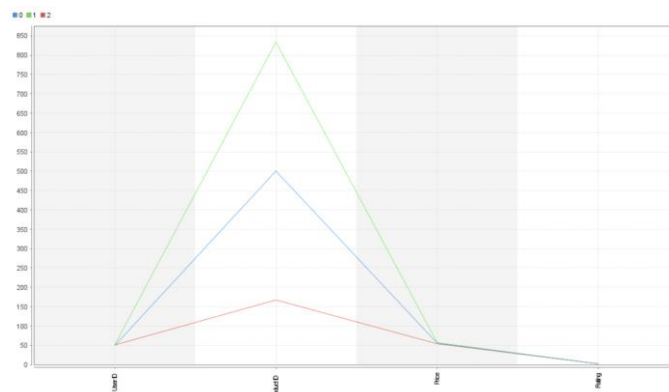
Gambar 6. Hasil Tampilan Graph Tree Cluster 0,1,2 yang sudah di Proses Rapidminer.

Graph Tree & Visualisasi 2D/3D: Digunakan untuk menunjukkan hubungan antar atribut dan memudahkan pemahaman terhadap pola distribusi data.

Attribute	cluster_0	cluster_1	cluster_2
User ID	48.895	51.093	51.266
Product ID	501	834	167.500
Price	56.276	56.652	54.431
Rating	2.930	3.058	2.992

Gambar 7. Hasil Tampilan dari Centroid Table yang sudah di Proses Rapidminer.

Centroid Table: Menampilkan nilai rata-rata setiap atribut dalam masing-masing kluster. Ini membantu dalam mengidentifikasi ciri khas dari masing-masing kelompok.



Gambar 8. Hasil Tampilan Scatter Plot Data pada Rapidminer.

Color Encoding: Warna kluster berfungsi untuk memperjelas pemisahan antar kelompok — sebagai contoh:

- a. Biru (Cluster 2): Produk dengan harga rendah, ukuran kecil, merek lokal.
- b. Hijau (Cluster 0): Produk dengan harga menengah, size campuran, merek semi-premium.
- c. Oranye (Cluster 1): Produk premium, harga tinggi, ukuran besar.

Keunggulan utama dari visualisasi ini adalah kemampuannya dalam menyampaikan hasil klasterisasi secara visual dan interaktif, yang sangat berguna dalam konteks pengambilan keputusan oleh pelaku bisnis non-teknis. Tanpa perlu memahami rumus statistik yang kompleks, pengguna dapat melihat perbedaan antar kelompok secara langsung melalui bentuk dan warna yang tampil di layar.

Dengan menggunakan RapidMiner, seluruh proses — mulai dari import data, transformasi atribut, eksekusi algoritma K-Means, hingga visualisasi hasil — dapat dilakukan hanya melalui antarmuka drag-and-drop, menjadikannya alat bantu yang sangat praktis untuk eksplorasi dan pemahaman data dalam industri fashion maupun sektor bisnis lainnya.

3.7. Pembahasan

Hasil segmentasi dengan K-Means clustering menghasilkan tiga klaster yang memiliki karakteristik berbeda berdasarkan atribut harga, ukuran, dan merek. Klaster pertama didominasi oleh produk dengan harga rendah, ukuran kecil (S dan M), dan merek lokal. Klaster kedua mencakup produk harga menengah dengan ukuran bervariasi dan merek campuran (lokal dan semi-premium). Sementara itu, klaster ketiga terdiri dari produk dengan harga tinggi, ukuran besar (L dan XL), serta berasal dari merek internasional atau premium.

Temuan ini hamper sama dengan sebuah penelitian oleh Hasibuan et al. (2021) yang menginformasikan penerapan K-Means untuk pengelompokan produk smartphone berdasarkan harga dan spesifikasi, yang menghasilkan pola distribusi konsumen terhadap produk berdasarkan preferensi nilai dan fitur. Dalam konteks fashion, atribut harga dan ukuran sering kali berkorelasi dengan positioning merek, sebagaimana ditunjukkan oleh Maresti et al. (2025) dalam segmentasi produk ritel berdasarkan data transaksi.

Secara metodologis, pemanfaatan Elbow Method berhasil mengidentifikasi nilai optimal $K=3$ sebagai jumlah klaster paling representatif, yang dibuktikan dengan penurunan tajam pada nilai inertia dari $K=2$ ke $K=3$, sebelum stabil pada K selanjutnya. Ini sesuai dengan pendekatan validasi yang diterapkan oleh Ardana et al. (2024) dalam segmentasi pelanggan online, yang juga menggunakan Elbow Method untuk validasi klasterisasi berbasis atribut campuran.

Penggunaan RapidMiner sebagai alat bantu visual dalam pemodelan memperkuat implementasi praktis metode K-Means, karena memungkinkan peneliti melakukan encoding atribut kategorikal dan normalisasi variabel numerik dengan lebih mudah, tanpa pemrograman. Hal ini sangat membantu dalam konteks penerapan industri seperti fashion retail, di mana pengguna non-teknis membutuhkan alat analisis yang intuitif namun tetap kuat secara metodologis.

Secara keseluruhan, hasil segmentasi dapat dimanfaatkan dalam berbagai strategi bisnis. Sebagai contoh:

- a. Strategi harga dan promosi: Produk dalam klaster harga rendah dapat ditargetkan dengan promosi diskon atau bundling, sedangkan produk premium dalam klaster 3 dapat difokuskan pada branding eksklusif.
- b. Perencanaan stok: Toko di area dengan dominasi ukuran kecil dapat difokuskan pada stok klaster 1, sedangkan toko di pusat perbelanjaan menengah-atas difokuskan pada klaster 3.
- c. Rekomendasi produk: Dengan menggunakan hasil klaster sebagai dasar sistem rekomendasi, sistem dapat menyarankan produk dari klaster yang sama berdasarkan preferensi konsumen sebelumnya.

4. KESIMPULAN

Penelitian ini membuktikan bahwa algoritma K-Means Clustering efektif digunakan dalam melakukan segmentasi produk fashion berdasarkan atribut harga, ukuran, dan merek. Melalui tahapan pra-pemrosesan seperti encoding data kategorikal dan normalisasi atribut numerik, serta penentuan jumlah klaster optimal menggunakan metode Elbow, diperoleh hasil segmentasi yang akurat dan representatif. Jumlah klaster optimal yang ditemukan adalah sebanyak tiga kelompok, di mana masing-masing klaster memiliki karakteristik unik yang dapat dijadikan dasar dalam pengambilan keputusan bisnis. Klaster pertama berisi produk dengan harga rendah, ukuran kecil, dan berasal dari merek lokal; klaster kedua mencakup produk dengan harga menengah, ukuran beragam, serta merek campuran; sedangkan klaster ketiga terdiri dari produk dengan harga tinggi, ukuran besar, dan didominasi oleh merek premium atau internasional. Segmentasi ini tidak hanya memberikan pemahaman mendalam terhadap karakteristik produk yang dimiliki, tetapi juga membantu dalam menyusun strategi pemasaran yang lebih terarah, pengelolaan stok yang efisien, serta sistem rekomendasi produk yang lebih relevan bagi konsumen. Penggunaan RapidMiner sebagai platform analisis data terbukti mempermudah proses implementasi K-Means, terutama bagi pelaku usaha non-teknis seperti UMKM, karena seluruh proses dapat dilakukan secara visual tanpa pemrograman. Dengan demikian, penelitian ini memberikan kontribusi praktis dalam mendukung digitalisasi pengelolaan produk di industri fashion, serta menjadi acuan bagi penelitian selanjutnya dalam pengembangan sistem segmentasi berbasis data mining.

REFERENCES

- [1] N. Lee and S. Suh, "How Does Digital Technology Inspire Global Fashion Design Trends? Big Data Analysis on Design Elements," *Applied Sciences (Switzerland)*, vol. 14, no. 13, 2024, doi: 10.3390/app14135693.

- [2] V. Schiaroli, L. Fraccascia, and R. M. Dangelico, "How can consumers behave sustainably in the fashion industry? A systematic literature review of determinants, drivers, and barriers across the consumption phases," *J Clean Prod*, vol. 483, no. February 2023, p. 144232, 2024, doi: 10.1016/j.jclepro.2024.144232.
- [3] P. Centobelli, S. Abbate, S. P. Nadeem, and J. A. Garza-Reyes, "Slowing the fast fashion industry: An all-round perspective," *Curr Opin Green Sustain Chem*, vol. 38, p. 100684, 2022, doi: 10.1016/j.cogsc.2022.100684.
- [4] A. Dwi Gitania, K. Adji Kusuma, M. Hariasih, J. Lebo No, K. Sidoarjo, and J. Timur, "Pengaruh Kualitas Produk, E-Commerce, Brand Image dan Harga Terhadap Keputusan Pembelian Melalui Minat Beli," *Journal of Business and Economics Research (JBE)*, vol. 6, no. 2, pp. 393–407, 2025, doi: 10.47065/jbe.v6i2.7140.
- [5] G. E. Putri, "Faktor-Faktor Mempengaruhi Keputusan Pembelian Produk Fashion Secara Online Melalui E-Commerce," *Jurnal Universitas Negeri Yogyakarta*, vol. 16, no. 1, pp. 1–8, 2021.
- [6] M. Mathew and R. Spinelli, "Decoding sustainable drivers: A systematic literature review on sustainability-induced consumer behaviour in the fast fashion industry," *Sustain Prod Consum*, vol. 55, no. February, pp. 132–145, 2025, doi: 10.1016/j.spc.2025.02.011.
- [7] A. F. Rahmawati, N. Farida, and Ngatno, "Pengaruh Brand Prestige dan Perceived Quality Terhadap Purchase Intention Melalui Brand Attitude," *Jurnal Ilmu Administrasi Bisnis*, vol. 12, no. 1, pp. 131–139, 2023, [Online]. Available: <https://ejournal3.undip.ac.id/index.php/jiab%0A%7C>
- [8] R. Jorgie, L. Batu, N. Simatupang, and A. H. Lubis, "Segmentasi Pelanggan Belanja Daring berdasarkan Click-ad dengan Algoritma K-means," vol. 3, no. 1, pp. 304–312, 2025.
- [9] D. Nopriyani, H. Rohayani, and Z. Akbar, "Penerapan Algoritma K-Means untuk Mengidentifikasi Pola Penjualan Frozen Food yang Paling Populer," *Bulletin of Computer Science Research*, vol. 5, no. 1, pp. 98–104, 2024, doi: 10.47065/bulletincsr.v5i1.445.
- [10] Mediana Rahmatika, Muharia, Tasya Salsabila, and Vicky F Sanjaya, "Analisis Swot Dalam Pengembangan Strategi Pemasaran Pada Produk Fashion Toko Zelora Lampung," *Jurnal Akuntansi dan Manajemen*, vol. 35, no. 3, pp. 213–234, 2024, doi: 10.53916/jam.v35i3.146.
- [11] R. F. Ramadhan, S. Hadi Wijoyo, and M. C. Saputra, "Penerapan Metode K-Means Clustering pada Ulasan Perumahan PT XYZ di Google Maps untuk Formulasi Strategi Bisnis dengan Analisis SWOT," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 6, pp. 2879–2888, 2023, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [12] M. Agustriya, M. Ula, and K. -, "Analisis Kinerja Algoritma Klasifikasi Naïve Bayes Menggunakan Genetic Algorithm dan Bagging untuk Data Publik Risiko Transaksi Kartu Kredit," *Jurnal Sistem dan Teknologi Informasi (JustIN)*, vol. 12, no. 3, p. 584, 2024, doi: 10.26418/justin.v12i3.80136.
- [13] N. Wakhidah, "Penerapan Data Mining K-Means Clustering Untuk," *Jurnal Transformatika*, vol. 8, no. 1, pp. 45–52, 2010.
- [14] A. L. M. Tampubolon, T. M. E. Y. Butar Butar, and S. Rochimah, "Segmentasi Pelanggan Majalah pada Situs Web E-Commerce dengan K-Means++ dan Metode RFM," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 6, pp. 1243–1252, 2024, doi: 10.25126/jtiik.2024118208.
- [15] S. R. Agustin, I. Purnamasari, and B. N. Sari, "Implementasi K-Means Untuk Pengelompokan Kategori Penjualan Barang Berbasis Web," *Journal of Informatics Management and Information Technology*, vol. 5, no. 3, pp. 167–176, 2025, doi: 10.47065/jimat.v5i3.610.
- [16] E. O. A. B. Nasution and I. Imsar, "Pengaruh Faktor Budaya Dan Penggunaan Media Sosial Terhadap Perilaku Konsumen Dalam Konsumsi Produk Halal Roti Ketawa Sambo Cap Ayam Roket Di Kota Pematangsiantar," *EKOMBIS REVIEW: Jurnal Ilmiah Ekonomi dan Bisnis*, vol. 11, no. 2, pp. 1981–1996, 2023, doi: 10.37676/ekombis.v11i2.6810.
- [17] W. N. Annisyak and Ali, "Pengaruh Citra Merek, Strategi Pemasaran Media Sosial, dan Minat Beli Terhadap Keputusan Pembelian," *Journal of Business and Economics Research (JBE)*, vol. 6, no. 1, pp. 225–233, 2025, doi: 10.47065/jbe.v6i1.6766.
- [18] R. S. Nurhalizah, R. Ardianto, and P. Purwono, "Analisis Supervised dan Unsupervised Learning pada Machine Learning: Systematic Literature Review," *Jurnal Ilmu Komputer dan Informatika*, vol. 4, no. 1, pp. 61–72, 2024, doi: 10.54082/jiki.168.
- [19] W. Alfian, K. -, and T. Hidayat, "Analisis Clustering Pegawai Berdasarkan Tingkat Kedisiplinan Menggunakan Algoritma K-Means dan Davies-Bouldin Index," *Journal of Electrical Engineering and Computer (JEECOM)*, vol. 6, no. 2, pp. 437–448, 2024, doi: 10.33650/jeeecom.v6i2.9556.
- [20] A. Rajsya and A. Rachman, "Rancang Bangun Penerapan Metode Elbow Pada K-Means Untuk Clustering Data Persediaan Barang," *Literatur Informatika & Komputer*, vol. 1, no. 4, pp. 395–403, 2024.