

Segmentasi Kategori Produk Berdasarkan Responsivitas Promosi pada Bisnis Fashion Retail Menggunakan Agglomerative Hierarchical Clustering

Wahiddin Ishak¹, Rusdah^{2*}

^{1,2}Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

E-mail: ¹2211601725@student.budiluhur.ac.id, ²rusdah@budiluhur.ac.id

(* : corresponding author)

Abstrak

Perusahaan *fashion retail* kerap mengulang jenis promosi yang sama tanpa mencapai penjualan yang sebanding, karena tidak tersedia pemetaan empiris mengenai kategori produk mana yang benar-benar responsif terhadap mekanisme promosi tertentu. Penelitian ini bertujuan untuk melakukan segmentasi kategori produk berdasarkan responsivitas promosi menggunakan Agglomerative Hierarchical Clustering dengan metode Ward's Linkage. Data transaksi penjualan tahun 2023 dari PT Matahari Department Store Tbk., setelah dilakukan preprocessing, menghasilkan 6.089 record yang mencakup 790 departemen, 14 nama promosi, dan enam periode promosi khas pasar Indonesia. Penentuan jumlah kluster optimal dilakukan melalui Silhouette Coefficient, Davies-Bouldin Index, dan penurunan SSE, serta divalidasi silang dengan lonjakan jarak penggabungan pada dendrogram. Ward's Linkage terbukti unggul atas Average Linkage pada seluruh nilai k yang diuji, dengan konfigurasi terbaik pada empat kluster (Silhouette 0,474; DBI 0,049). Keempat kluster menunjukkan responsivitas yang berbeda-beda: satu kluster beranggotakan 317 departemen menyumbang 99,11% dari kuantitas terjual dengan PromoHitRate 0,986, sementara kluster lain beranggotakan 316 departemen sama sekali tidak mencatat penjualan dari promosi. Distribusi tipe dan periode promosi yang hampir seragam antarkluster menunjukkan bahwa perbedaan kinerja tidak berasal dari bias paparan promosi, melainkan dari efektivitas internal masing-masing departemen. Hasil segmentasi ini dapat dijadikan dasar alokasi anggaran promosi yang lebih terarah serta pengelolaan inventaris yang lebih efisien.

Kata kunci: agglomerative hierarchical clustering, fashion retail, responsivitas promosi, segmentasi produk, Ward's linkage

Abstract

Fashion retail companies often repeat the same types of promotions without achieving comparable sales results, largely because they lack empirical mapping of which product categories respond to specific promotional mechanisms. This study segments product categories based on promotional responsiveness using Agglomerative Hierarchical Clustering with Ward's Linkage. Following preprocessing, 2023 sales transaction data from PT Matahari Department Store Tbk. yielded 6,089 records covering 790 departments, 14 promotion names, and six promotional periods characteristic of the Indonesian market. We determined the optimal number of clusters using the Silhouette Coefficient, Davies-Bouldin Index, and SSE reduction, and cross-validated it against jumps in merging distances in the dendrogram. Ward's Linkage outperformed Average Linkage across all tested values of k , with the best configuration achieved at four clusters (Silhouette: 0.474; DBI: 0.049). The four clusters showed varying levels of responsiveness: one cluster of 317 departments accounted for 99.11% of sales volume with a PromoHitRate of 0.986, whereas another cluster of 316 departments recorded no sales from promotions. The nearly uniform distribution of promotion types and periods across clusters suggests that performance differences stemmed not from exposure bias, but from the internal effectiveness of individual departments. These segmentation results can support more targeted promotional budget allocation and more efficient inventory management.

Keywords: agglomerative hierarchical clustering, fashion retail, product segmentation, promotional responsiveness, Ward's linkage

1. PENDAHULUAN

Industri *fashion retail* di Indonesia menunjukkan dinamika yang tajam pasca pandemi COVID-19. PT Matahari Department Store, Tbk. mencatat kerugian Rp873 miliar dengan *Same-Store Sales Growth* (SSSG) -52,4% pada 2020, lalu pulih menjadi laba bersih Rp913 miliar pada 2021 dan Rp1,38 triliun pada 2022, sebelum kembali turun menjadi Rp675 miliar dengan SSSG -2,4% pada 2023 [1]. Fluktuasi ini menunjukkan bahwa keberhasilan penjualan tidak semata-mata

ditentukan oleh ekspansi gerai, melainkan oleh ketepatan strategi promosi yang disesuaikan dengan karakteristik produk. Dinamika tersebut berlangsung di tengah pasar *fashion* Indonesia yang perilaku konsumsinya terus bergeser, baik pada segmen premium yang makin mempertimbangkan nilai jangka panjang [2] maupun pada segmen umum yang keputusan belinya makin dipengaruhi kanal digital [3].

Persoalan praktis muncul ketika perusahaan mengulang jenis promosi yang sama secara berkala tanpa mempertimbangkan bahwa tidak semua kategori produk merespons dengan intensitas yang sama. Akibatnya, sebagian kategori berubah menjadi *slow-moving*, bahkan *aging stock*, sehingga menambah beban penyimpanan dan menurunkan efisiensi rantai pasok. Produk *slow-moving* memiliki tingkat penjualan yang jauh lebih rendah dibandingkan produk lain, sehingga memerlukan penanganan promosi yang lebih spesifik [4]. Di sisi lain, efektivitas promosi terbukti tidak seragam antarproduk. Analisis atas data penjualan harian menunjukkan bahwa respons terhadap mekanisme promosi yang sama sangat bervariasi antarkategori [5], sementara simulasi berbasis agen menegaskan bahwa promosi yang tidak disesuaikan dengan karakteristik produk cenderung menggerus margin tanpa menaikkan volume secara proporsional [6]. Kedua temuan tersebut mengarah pada kebutuhan yang sama, yaitu pemetaan empiris atas segmen produk berdasarkan responsivitasnya terhadap promosi.

Responsivitas dalam konteks penelitian ini didefinisikan sebagai tingkat perubahan kinerja penjualan suatu kategori produk selama periode promosi berlangsung. Definisi ini sejalan dengan konsep *marketing agility*, yaitu kemampuan perusahaan untuk beradaptasi dengan cepat terhadap perubahan pasar dan merespons umpan balik dari konsumen. Operasionalisasi responsivitas pada penelitian ini dilakukan melalui dua indikator, yaitu proporsi kesempatan promosi yang benar-benar menghasilkan penjualan dan intensitas volume penjualan yang terstandarisasi.

Sejumlah penelitian terdahulu telah memanfaatkan teknik klasterisasi untuk persoalan segmentasi di ranah ritel. Agussalim *et al.* [7] menggabungkan model *Recency, Frequency, Monetary* (RFM) dengan *K-Means* dan memperoleh tiga klaster produk. Namun, hasilnya sangat dipengaruhi oleh inisialisasi *centroid* sehingga tidak konsisten pada data yang tidak berdistribusi linear. Ufeli *et al.* [8] membandingkan *K-Means* dan *agglomerative clustering* pada data ritel bertipe campuran dan menemukan bahwa pendekatan aglomeratif menghasilkan *Silhouette Coefficient* yang lebih tinggi pada jumlah klaster optimal; kelebihan pendekatan ini terletak pada tidak perlunya asumsi jumlah klaster di awal. Wulandari *et al.* [9] membandingkan *DBSCAN*, *K-Means*, dan *X-Means* pada data tren belanja dan menemukan bahwa *DBSCAN* unggul dalam validitas klaster, meskipun algoritma tersebut tidak menyediakan visualisasi hierarki. Lubis *et al.* [10] turut menegaskan keunggulan pendekatan aglomeratif dalam memberikan struktur klaster yang lebih rinci pada data penjualan produk.

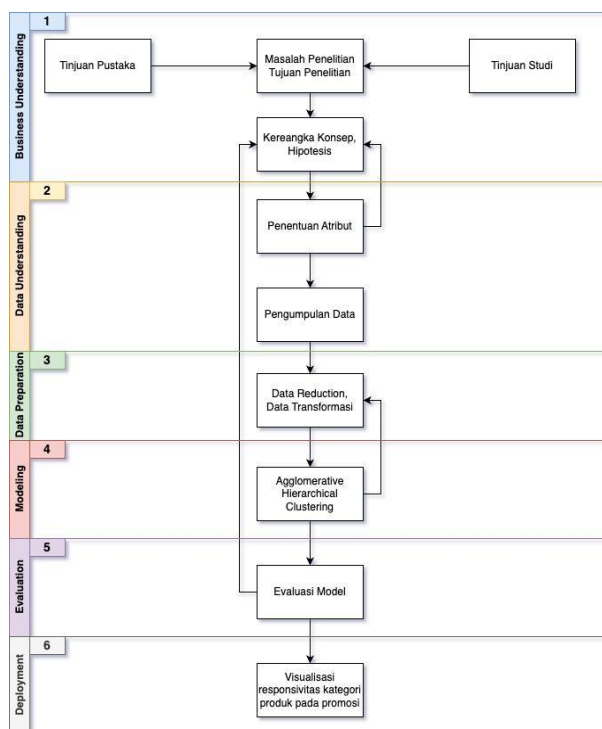
Pada konteks *retail* secara lebih luas, John *et al.* [11] mengeksplorasi beragam algoritma klasterisasi untuk segmentasi pelanggan di pasar *retail* Inggris. Febrianty *et al.* [12] serta Arvio *et al.* [13] memanfaatkan *K-Means* untuk mengoptimalkan strategi pemasaran dan memodelkan segmentasi transaksi. Adapun kajian yang secara khusus mengaitkan penambangan data dengan penyusunan strategi promosi telah dilakukan oleh Iswandaru dan Indriyanti [14], Sugianti *et al.* [15], serta Pii *et al.* [16]. Seluruh kajian tersebut berhenti pada segmentasi pelanggan atau segmentasi produk secara umum, dan belum ada yang menempatkan responsivitas promosi sebagai dimensi segmentasi dengan periode promosi yang disesuaikan dengan karakteristik musiman pasar di Indonesia.

Perbandingan karakteristik algoritma memperkuat pemilihan metode dalam penelitian ini. AHC tidak memerlukan penentuan jumlah klaster di awal, menghasilkan klaster yang kompak dan homogen, relatif *robust* terhadap *outlier*, serta menyediakan visualisasi hierarki melalui dendrogram. *K-Means* mensyaratkan jumlah klaster ditetapkan terlebih dahulu dan sangat terpengaruh oleh *outlier*, sedangkan *DBSCAN* mengabaikan *outlier* dan tidak menghasilkan struktur hierarki [8]. Struktur hierarki tersebut justru menjadi kebutuhan dalam penelitian ini karena hierarki produk di *department store* bersifat berjenjang, mulai dari *division*, *group*, *department*, *class*, hingga *subclass*. Sifat *order-preserving* pada klasterisasi aglomeratif juga memungkinkan struktur penggabungan ditelusuri secara bertahap [17].

Kontribusi penelitian ini adalah menghadirkan segmentasi kategori produk yang secara eksplisit dibangun di atas dimensi responsivitas promosi, dengan memasukkan periode promosi khas konsumen Indonesia sebagai atribut, serta memvalidasi jumlah kluster optimal melalui kombinasi tiga metrik internal dan analisis lonjakan jarak penggabungan. Penelitian ini menjawab dua pertanyaan, yaitu bagaimana mengidentifikasi mekanisme promosi yang paling efektif bagi setiap kelompok kategori produk berdasarkan pola penjualan historis, dan bagaimana menganalisis cakupan serta keberagaman departemen yang terpapar berbagai jenis dan periode promosi dalam kelompok yang terbentuk.

2. METODE PENELITIAN

Penelitian ini menggunakan metodologi *Cross-Industry Standard Process for Data Mining* (CRISP-DM) yang terdiri atas enam tahap, yaitu *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment* [18]. Pemilihan CRISP-DM didasarkan pada sifatnya yang iteratif dan berorientasi pada tujuan bisnis, sehingga sesuai untuk persoalan segmentasi yang hasilnya harus dapat diterjemahkan menjadi keputusan operasional. Alur tahapan penelitian ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian dengan Metodologi CRISP-DM

2.1. Sumber dan Pengumpulan Data

Data yang digunakan bersifat sekunder dan bersumber dari *Oracle Retail Merchandising System* (RMS), yaitu *core system* yang dioperasikan PT Matahari Department Store, Tbk. Ekstraksi dilakukan untuk seluruh transaksi penjualan sepanjang tahun 2023 pada jaringan toko fisik perusahaan, mencakup entitas *ItemHierarchy*, *Items*, *SalesHeader*, *SalesDetail*, *PromotionType*, *PromotionHeader*, dan *PromotionDetail*. Rentang satu tahun penuh dipilih agar seluruh siklus musiman belanja *fashion* terwakili.

Satu entitas tambahan dirancang khusus untuk penelitian ini, yaitu *PromotionPeriod*, yang mengelompokkan promosi ke dalam rentang waktu yang mencerminkan perilaku belanja konsumen Indonesia: *Peak Session Lebaran*, *Peak Session Nataru*, *Peak Session Imlek*, *Back to School*, *Independence Day*, dan *Partnership Campaign*. Entitas ini digabungkan ke dalam data penjualan melalui operasi *JOIN* berdasarkan kesesuaian tanggal transaksi dengan *StartDate* dan *EndDate* periode. Unit analisis ditetapkan pada level departemen, yaitu level hierarki tempat

inventaris dilacak dan dilaporkan, sehingga atribut *QuantitySold* diagregasi dari seluruh *SKU* pada departemen yang bersangkutan.

2.2. Persiapan Data

Pembersihan data dilakukan berdasarkan lima persoalan kualitas. Duplikasi sebanyak 14.736 baris yang timbul dari proses *re-sync* aplikasi *Point of Sales* saat jaringan toko terputus dideteksi melalui agregasi kombinasi *TrxDate*, *TrxNumber*, *SKU*, dan *Dept*, lalu dihapus. Nilai kosong pada *QuantitySold* sebanyak 6.402 baris diisi dengan median per departemen, sedangkan 2.997 baris *PromotionName* kosong akibat *hard-tag* manual yang diisi mengikuti kondisi harga pada baris tersebut; 184 baris dengan lebih dari 50% atribut kosong dihapus. Validasi format menemukan 8.541 baris bertanggal *default* 1900-01-01 yang dikoreksi berdasarkan data audit, serta 1.112 baris *QuantitySold* bernilai negatif akibat retur yang dihapus karena telah tercatat pada entitas lain. Variasi ejaan pada *PromotionName* dinormalisasi menjadi label tunggal, dan 12.236 baris di luar lingkup penelitian disaring. Ringkasan hasil pembersihan disajikan pada Tabel 1.

Tabel 1. Ringkasan Kondisi Data Sebelum dan Sesudah Pembersihan

Indikator	Sebelum	Sesudah
Total baris	103.953.201	103.905.233
Baris duplikat	14.736	0
Baris kosong (≥ 1 <i>key column</i>)	9.399	0
Baris bertanggal <i>default</i> 1900-01-01	8.541	0
Label promosi inkonsisten	5 variasi	1 variasi
Baris di luar lingkup penelitian	12.236	0

Pemilihan atribut mempertimbangkan keterkaitannya secara langsung dengan perilaku respons konsumen terhadap promosi. Lima atribut yang digunakan beserta tipe datanya disajikan pada Tabel 2.

Tabel 2. Atribut yang Digunakan dalam Pemodelan

Atribut	Deskripsi	Tipe Data
<i>PromotionName</i>	Nama skema promosi yang diterapkan	Kategorikal
<i>PromotionType</i>	Klasifikasi <i>Simple</i> atau <i>Complex Promotion</i>	Kategorikal
<i>Department</i>	Kategori atau <i>brand</i> produk sebagai unit analisis	Kategorikal
<i>PromotionPeriod</i>	Periode musiman saat promosi berlangsung	Kategorikal
<i>QuantitySold</i>	Total kuantitas produk terjual per departemen	Numerik

Transformasi data dilakukan dalam dua bentuk. Atribut kategorikal dikodekan menggunakan *label encoding* sehingga kompatibel dengan perhitungan jarak, menghasilkan 2 kode untuk tipe promosi, 14 kode untuk nama promosi, dan 6 kode untuk periode promosi. Atribut numerik *QuantitySold* diagregasi ke level departemen lalu dinormalisasi ke rentang [0,1] menggunakan *Min-Max Scaling* agar perbedaan skala antardepartemen tidak mendominasi perhitungan jarak. Dataset akhir yang dihasilkan berisi 6.089 *record* dengan 5 atribut, mencakup 790 departemen, dengan 11 dari 14 nama promosi yang benar-benar muncul dalam data.

2.3. Pemodelan dengan Agglomerative Hierarchical Clustering

AHC bekerja secara *bottom-up*: setiap objek diawali sebagai kluster tunggal, lalu dua kluster terdekat digabungkan secara iteratif hingga seluruh objek menyatu dalam satu kluster [8]. Penelitian ini menggunakan *Ward's Linkage*, yang menggabungkan pasangan kluster dengan kenaikan terkecil pada total varians intrakluster atau *sum of squared errors*. Pendekatan ini dipilih karena menghasilkan kluster yang lebih kompak dibandingkan *linkage* lain dan lebih sensitif terhadap perubahan kecil dalam distribusi penjualan, yang justru menjadi inti pengukuran responsivitas.

Proses penggabungan menghasilkan matriks *linkage* yang mencatat pasangan klaster yang digabungkan, jarak penggabungan, jumlah observasi dalam klaster baru, serta anggota klaster pada setiap langkah. Matriks inilah yang menjadi dasar untuk konstruksi dendrogram dan penentuan jumlah klaster yang optimal. Sebagai pembandingan, *Average Linkage* yang menghitung rata-rata jarak seluruh pasangan titik antara dua klaster turut diuji pada rentang k yang sama [8].

2.4. Evaluasi Model

Karena AHC bersifat *unsupervised* dan tidak tersedia label acuan, evaluasi dilakukan menggunakan metrik internal yang mengukur kohesi dalam klaster dan separasi antarklaster. Metrik utama adalah *Silhouette Coefficient*, yang menilai kesesuaian penempatan suatu objek pada klasternya dibanding klaster terdekat lain dengan rentang nilai -1 hingga 1 [19]. Metrik pelengkap adalah *Davies-Bouldin Index* (DBI) yang mengukur rasio antara dispersi dalam klaster dan jarak antarklaster, di mana nilai yang lebih rendah menandakan pemisahan yang lebih baik [20]. Kedua metrik ini dilaporkan sebagai indeks internal yang paling informatif dibandingkan alternatifnya [19]. Penurunan *Sum of Squared Errors* (ΔSSE) secara *post hoc* digunakan untuk memantau apakah penambahan jumlah klaster masih memberikan perbaikan yang berarti atau justru menimbulkan fragmentasi berlebihan.

Selain ketiga metrik tersebut, penentuan jumlah klaster divalidasi secara silang melalui analisis lonjakan jarak penggabungan pada dendrogram. Pendekatan berlapis ini dipilih agar keputusan jumlah klaster tidak bergantung pada satu indikator tunggal, dan agar kriteria statistik dapat dipertemukan dengan pertimbangan keterbacaan bisnis. Sebagai tahap *deployment*, hasil klasterisasi diimplementasikan ke dalam prototipe aplikasi web berbasis *Python* dengan *framework* Flask, yang menyajikan ringkasan klaster, dendrogram interaktif dengan garis ambang batas, dan panel evaluasi metrik.

3. HASIL DAN PEMBAHASAN

3.1. Perbandingan Ward's Linkage dan Average Linkage

Pengujian terhadap dua pendekatan *linkage* dilakukan pada rentang $k = 2$ hingga 6 dengan dua metrik sekaligus. Hasil yang disajikan pada Tabel 3 menunjukkan bahwa *Ward's Linkage* secara konsisten menghasilkan *Silhouette Coefficient* yang lebih tinggi dan DBI yang lebih rendah dibandingkan dengan *Average Linkage* pada seluruh nilai k yang diuji, tanpa pengecualian.

Tabel 3. Perbandingan Kualitas Klaster antara Ward's Linkage dan Average Linkage

k	<i>Silhouette</i> (Ward)	DBI (Ward)	<i>Silhouette</i> (Average)	DBI (Average)
2	0,468	0,062	0,412	0,089
3	0,372	0,053	0,348	0,078
4	0,474	0,049	0,429	0,071
5	0,455	0,055	0,410	0,075
6	0,426	0,058	0,395	0,080

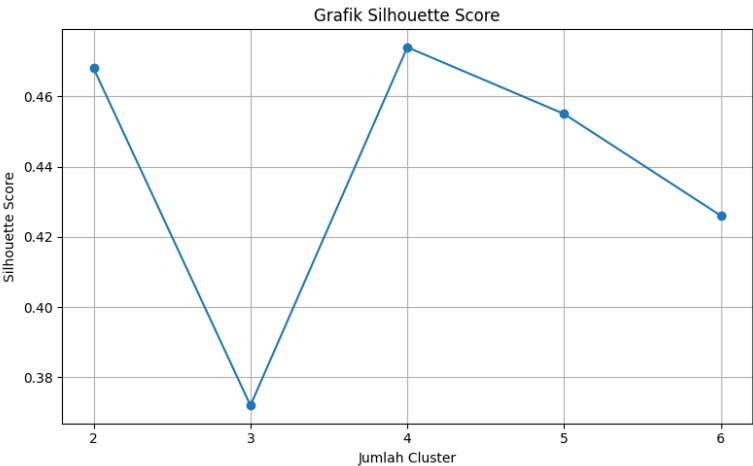
Konsistensi keunggulan ini menandakan bahwa struktur data promosi dan penjualan *fashion retail* memiliki sebaran varians yang besar, kondisi yang secara teoretis memang ditangani lebih baik oleh kriteria minimasi varians pada *Ward's Linkage* dibandingkan dengan kriteria rata-rata jarak. Temuan ini memperkuat hasil Ufeli *et al.* [8] dan Lubis *et al.* [10] yang juga menemukan bahwa pendekatan aglomeratif mampu menghasilkan struktur klaster yang lebih rinci pada data transaksi penjualan. Atas dasar itu, *Ward's Linkage* ditetapkan sebagai pendekatan utama untuk seluruh analisis selanjutnya.

3.2. Penentuan Jumlah Klaster Optimal

Penentuan jumlah klaster dilakukan melalui dua jalur yang saling melengkapi. Jalur pertama adalah evaluasi metrik internal pada rentang $k = 2$ hingga 6, sebagaimana disajikan pada Tabel 4 dan divisualisasikan pada Gambar 2.

Tabel 4. Rangkuman Metrik Evaluasi Internal Per Jumlah Kluster

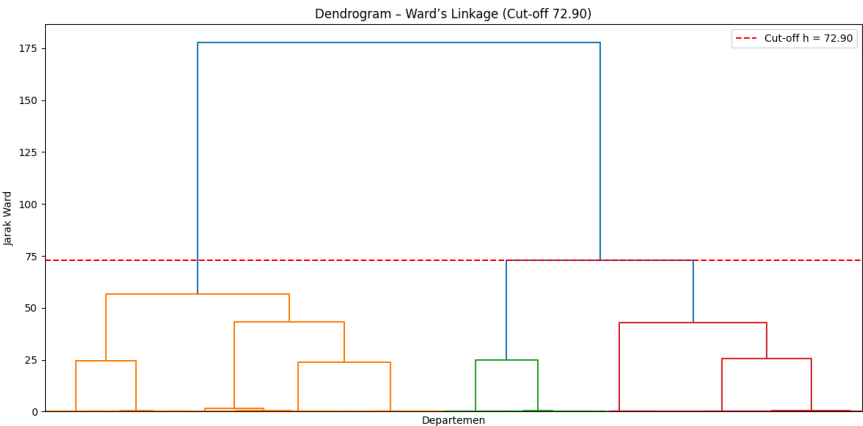
<i>k</i>	<i>Silhouette</i> (avg)	<i>Davies-Bouldin Index</i>	Δ SSE	Interpretasi
2	0,218	0,062	–	Pemisahan kasar
3	0,272	0,053	-15%	Struktur membaik
4	0,292	0,049	-22%	Optimal
5	0,258	0,055	-10%	Terfragmentasi
6	0,241	0,058	-7%	Tidak stabil



Gambar 2. Grafik Silhouette Score pada jumlah kluster *k* = 2 hingga 6

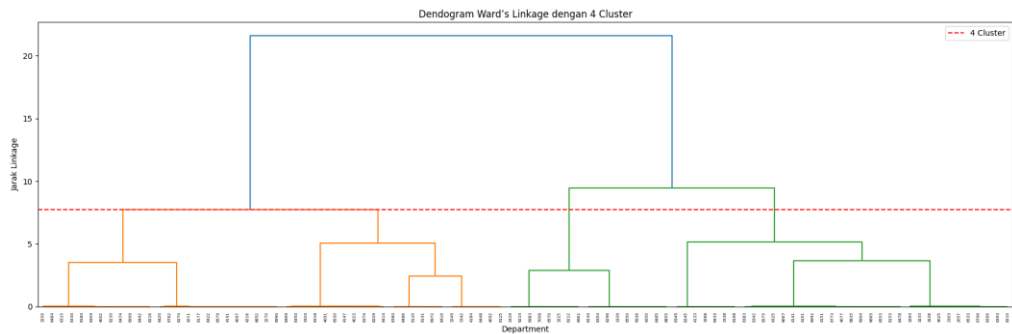
Ketiga metrik menunjuk pada titik yang sama. *Silhouette Coefficient* mencapai puncaknya pada *k* = 4, DBI mencapai nilai terendahnya pada *k* = 4, dan penurunan SSE terbesar sebesar 22% juga terjadi pada transisi menuju *k* = 4. Setelah titik tersebut, penambahan kluster justru menurunkan *Silhouette* dan menaikkan DBI, sekaligus memberikan penurunan SSE yang semakin kecil, yang merupakan tanda fragmentasi tanpa perbaikan struktur.

Jalur kedua adalah analisis lonjakan jarak penggabungan pada dendrogram. Lonjakan terbesar terjadi pada tahap penggabungan ke-6086, di mana jarak *Ward* meningkat tajam dari sekitar 72,90 menjadi sekitar 177,57. Lonjakan sebesar ini menandakan bahwa penggabungan pada tahap tersebut menyatukan dua kelompok dengan perbedaan karakteristik yang sangat besar. Garis potong karenanya ditempatkan sedikit di bawah lonjakan, yaitu pada *h* ≈ 72,90, sebagaimana ditunjukkan pada Gambar 3*.



Gambar 3. Dendrogram Ward's Linkage dengan Garis Ambang Batas *h** = 72,90

Konvergensi antara kriteria statistik dan kriteria struktural ini memperkuat keandalan keputusan. Empat klaster juga merupakan konfigurasi yang secara praktis paling mudah diterjemahkan menjadi kerangka pengambilan keputusan promosi, berbeda dengan $k = 2$ yang terlalu kasar untuk keperluan operasional maupun $k = 6$ yang mulai memunculkan klaster berukuran kecil dan tidak seimbang. Struktur dendrogram pada pemotongan $k = 4$ disajikan pada Gambar 4, memperlihatkan cabang-cabang yang relatif seimbang secara visual.



Gambar 4. Dendrogram Ward's Linkage pada Pemotongan $k = 4$

3.3. Karakteristik Responsivitas Antarklaster

Karakteristik keempat klaster diukur melalui tiga indikator. *PromoHitRate* menyatakan rata-rata proporsi kombinasi departemen, promosi, dan periode yang benar-benar menghasilkan penjualan. Nilai mendekati 1 berarti hampir seluruh kesempatan promosi berhasil. *QuantitySold_norm* menyatakan rata-rata intensitas penjualan terstandardisasi per departemen, sedangkan *ShareQty* menyatakan kontribusi persentase klaster terhadap total penjualan terstandardisasi. Hasilnya disajikan pada Tabel 5.

Tabel 5. Statistik Inti Empat Klaster Hasil Ward's Linkage

Cluster	<i>PromoHitRate</i> (avg)	<i>QuantitySold_norm</i> (avg)	<i>ShareQty</i> (%)	Jumlah Departemen
1	0,9856	0,004735	99,11	317
2	0,2248	0,000033	0,16	86
3	0,6226	0,000171	0,70	71
4	0,0000	0,000000	0,00	316

Sebaran ini memperlihatkan ketimpangan yang ekstrem. Klaster 1 yang beranggotakan 317 departemen menyerap 99,11% dari seluruh kuantitas terjual yang berkaitan dengan promosi, dengan *PromoHitRate* 0,986 yang berarti hampir setiap kesempatan promosi langsung berbuah penjualan. Pada ujung yang berlawanan, klaster 4 dengan 316 departemen sama sekali tidak mencatat penjualan dari promosi. Dua klaster dengan jumlah anggota yang nyaris identik ini menghasilkan kinerja yang sepenuhnya bertolak belakang, sebuah temuan yang menuntut penjelasan lebih lanjut.

Penjelasan pertama yang perlu diuji adalah kemungkinan bias paparan promosi, yaitu dugaan bahwa klaster berkinerja rendah sesungguhnya hanya menerima lebih sedikit atau ragam promosi yang lebih sempit. Analisis distribusi menunjukkan sebaliknya. Proporsi *Complex Promotion* terhadap *Simple Promotion* nyaris seragam di seluruh klaster, berkisar 70% berbanding 30%, dengan klaster 4 justru mencatat proporsi *Complex Promotion* tertinggi sebesar 72,64%. Distribusi periode promosi juga merata: seluruh klaster berpartisipasi pada seluruh periode utama, dengan *Independence Day* berkisar 19,28% hingga 21,61% dan *Peak Session Nataru* berkisar 17,41% hingga 21,65% di semua klaster.

Dengan demikian, selisih kinerja antarklaster tidak berasal dari keterbatasan jenis maupun jadwal promosi yang diterima, melainkan dari efektivitas internal masing-masing departemen dalam memanfaatkan promosi yang sama. Temuan ini merupakan kontribusi utama penelitian,

karena membalikkan asumsi umum bahwa kategori berkinerja rendah perlu diberi lebih banyak promosi. Data menunjukkan bahwa penambahan volume promosi pada klaster 4 kemungkinan besar hanya menambah biaya tanpa menghasilkan penjualan.

Indikator keragaman memberikan nuansa tambahan. Rata-rata jumlah periode promosi berbeda yang dialami tiap departemen adalah 6,9 pada klaster 1, 6,2 pada klaster 2, 6,1 pada klaster 3, dan hanya 4,1 pada klaster 4. Klaster 3 justru mencatat keragaman jenis dan nama promosi tertinggi, yaitu rata-rata 2,0 jenis per departemen, namun hanya menghasilkan *PromoHitRate* 0,62 dan kontribusi sebesar 0,70%. Fakta bahwa keragaman promosi tidak berbanding lurus dengan kinerja penjualan menegaskan kembali bahwa determinan utama responsivitas terletak pada karakteristik produk itu sendiri, bukan pada intensitas perlakuan promosi. Adapun *Percentage Off* dan *Message of Receipt* konsisten menjadi dua nama promosi dengan kontribusi penjualan terbesar di seluruh klaster.

3.4. Interpretasi Klaster dan Implikasi Strategi Promosi

Untuk menerjemahkan hasil segmentasi ke dalam keputusan operasional, komposisi 790 departemen ditelusuri satu level di atasnya, yaitu pada level *group* dalam hierarki produk. Penelusuran ini memungkinkan manajemen mengenali kategori produk yang terdapat di setiap klaster. Ringkasan karakteristik beserta implikasi strateginya disajikan pada Tabel 6.

Tabel 6. Karakteristik Klaster dan Implikasi Strategi Promosi

<i>Cluster</i>	<i>Label</i>	<i>Komposisi Group Dominan</i>	<i>Implikasi Strategi Promosi</i>
1	Respons tinggi dan kontribusi besar	<i>Young Fashion, Men Young Casual, Ladies Contemp, Men Formal, Men Contemp Casual</i>	Menjadi <i>baseline</i> evaluasi promosi. Diskon berkala dapat dioptimalkan karena ragam dan frekuensi promosi sudah memadai serta terbukti meningkatkan penjualan.
2	Volume kecil dan respons terbatas	<i>Toys, Men Shoes, Accessories, Children Boys, Health & Wellness</i>	Respons tetap ada meski lemah. Lebih efektif bila promosi dikombinasikan dengan strategi <i>cross-selling</i> untuk klaster 1, bukan dilakukan secara berdiri sendiri.
3	Ragam tinggi dan kinerja menengah	<i>Premium Brand, Active Shoes, Handbags, Intimate, Ladies Shoes</i>	Keragaman promosi tertinggi, namun hasil yang menengah mengindikasikan kejenuhan konsumen atau promosi yang kurang tepat sasaran. Cocok sebagai ruang uji untuk mekanisme promosi baru.
4	Tidak merespons promosi	<i>Men Formal, Infants & Essentials, Home, Fragrance</i>	Penambahan promosi tidak menghasilkan penjualan. Anggaran promosi sebaiknya dialihkan, dengan pendekatan selektif berbasis momen tertentu saja.

Implikasi paling langsung dari pemetaan ini adalah pada alokasi anggaran. Selama ini, promosi cenderung disebarkan merata ke seluruh departemen, padahal 99,11% hasilnya terkonsentrasi pada satu klaster. Realokasi anggaran dari klaster 4 menuju penguatan mekanisme pada klaster 1 dan eksperimen terarah pada klaster 3 berpotensi meningkatkan efisiensi promosi secara signifikan tanpa menambah total belanja promosi.

Hasil ini melengkapi penelitian terdahulu dari sisi yang berbeda. Iswandaru dan Indriyanti [14] serta Sugianti *et al.* [15] menghasilkan model untuk menentukan produk yang layak dipromosikan, sedangkan penelitian ini menghasilkan klaster yang menunjukkan produk mana yang tidak layak dipromosikan lagi, sebuah informasi yang justru bernilai tinggi dalam konteks efisiensi biaya. Dibandingkan dengan pendekatan berbasis *K-Means* yang digunakan Pii *et al.* [16], struktur hierarki yang dihasilkan AHC memungkinkan manajemen menelusuri tingkat granularitas segmentasi sesuai kebutuhan pengambilan keputusan, dari segmentasi makro dengan dua klaster hingga segmentasi mikro dengan enam klaster, tanpa perlu menjalankan ulang proses pemodelan.

Beberapa keterbatasan perlu dicatat saat membaca hasil ini. Data terbatas pada satu tahun transaksi dari satu perusahaan, sehingga generalisasi ke seluruh industri *fashion retail* di

Indonesia memerlukan pengujian lebih lanjut. Atribut yang digunakan juga belum memasukkan dimensi harga, margin, maupun kanal distribusi, padahal ketiganya berpotensi menjelaskan sebagian variasi responsivitas yang teramati. Selain itu, keberadaan klaster 4 dengan *PromoHitRate* nol menghadirkan persoalan *sparsity* yang sensitif terhadap pilihan metode penanganan data, sehingga interpretasi atas klaster tersebut sebaiknya diperlakukan sebagai indikasi awal, bukan sebagai simpulan akhir.

4. KESIMPULAN DAN SARAN

Penelitian ini berhasil membentuk segmentasi kategori produk berdasarkan responsivitas terhadap promosi pada bisnis *fashion retail* menggunakan *Agglomerative Hierarchical Clustering* dengan metode *Ward's Linkage*. Pengujian terhadap 6.089 *record* yang mencakup 790 departemen menunjukkan bahwa *Ward's Linkage* konsisten lebih unggul dibandingkan Average Linkage pada seluruh nilai k yang diuji. Empat klaster merupakan konfigurasi optimal berdasarkan konvergensi Silhouette Coefficient tertinggi sebesar 0,474, Davies-Bouldin Index terendah sebesar 0,049, penurunan SSE terbesar sebesar 22%, serta lonjakan jarak penggabungan pada dendrogram pada ketinggian 72,90. Mekanisme promosi yang paling efektif berbeda antarklaster, dengan *Percentage Off* dan *Message of Receipt* secara konsisten memberikan kontribusi penjualan terbesar, sementara cakupan periode promosi per departemen bervariasi dari rata-rata 6,9 periode pada klaster berkinerja tinggi hingga hanya 4,1 periode pada klaster yang tidak merespons.

Temuan paling penting dari penelitian ini adalah bahwa perbedaan kinerja antarklaster tidak dapat dijelaskan oleh bias paparan promosi. Seluruh klaster menerima komposisi tipe promosi yang nyaris seragam, pada kisaran 70% berbanding 30%, serta berpartisipasi dalam seluruh periode utama. Namun, 99,11% kuantitas terjual terkonsentrasi pada satu klaster, sementara klaster lain dengan jumlah anggota yang hampir sama tidak mencatat penjualan sama sekali. Hal ini menunjukkan bahwa determinan responsivitas terletak pada efektivitas internal tiap departemen dalam memanfaatkan promosi yang sama, bukan pada jumlah promosi yang diterima. Konsekuensi praktisnya, peningkatan efisiensi promosi lebih mungkin dicapai melalui realokasi anggaran antarsegmen dibandingkan dengan penambahan volume promosi secara merata. Prototipe berbasis web yang dikembangkan pada tahap *deployment* memungkinkan manajemen menelusuri hasil segmentasi ini secara interaktif untuk merumuskan strategi promosi.

Penelitian lanjutan disarankan untuk mengintegrasikan variabel harga, margin, dan kanal distribusi guna memperkaya penjelasan atas variasi responsivitas yang belum terjelaskan oleh atribut saat ini. Perluasan rentang data melampaui satu tahun dan mencakup lebih dari satu perusahaan akan memperkuat generalisasi temuan, sekaligus memungkinkan pengujian stabilitas segmentasi lintas siklus musiman. Studi komparatif dengan metode klasterisasi lain, seperti *DBSCAN* [21] atau *K-Means++* [22], atau kombinasi keduanya [23], perlu dilakukan untuk menguji konsistensi struktur klaster yang dihasilkan, khususnya pada klaster dengan karakteristik *sparsity* yang tinggi. Arah pengembangan yang paling menjanjikan adalah pemanfaatan hasil klasterisasi dalam sistem rekomendasi promosi otomatis, sehingga keputusan promosi untuk suatu kategori produk dapat menyesuaikan diri secara adaptif berdasarkan profil klaster yang telah terbentuk.

DAFTAR PUSTAKA

- [1] PT Matahari Department Store Tbk., *Annual Report 2023*. Jakarta: PT Matahari Department Store Tbk., 2023.
- [2] Y. Sun, R. Wang, H. Bai, *et al.*, "Perceived values of luxury fashion brands—evidence from Indonesian Millennials," *SN Business & Economics*, vol. 4, art. 169, 2024, doi: 10.1007/s43546-024-00755-x.

- [3] Y. W. Handranata and S. Kalila, "Attitude toward fashion influencers and its impact on purchase behavior: the roles of brand attitude and purchase intention," *Frontiers in Communication*, vol. 10, art. 1583602, 2025, doi: 10.3389/fcomm.2025.1583602.
- [4] W. N. M. Ariffin *et al.*, "Product Pairing Selection for Promotion using Partitioning Method," *Journal of Advanced Research in Applied Sciences and Engineering Technology*, vol. 30, no. 3, pp. 91–103, 2023, doi: 10.37934/araset.30.3.91103.
- [5] N. Phumchusri, W. Kosawanitchakarn, S. Chawanapranee, *et al.*, "Evaluating promotional pricing effectiveness using convenience store daily sales data," *Journal of Revenue and Pricing Management*, vol. 22, no. 5, pp. 362–373, 2023, doi: 10.1057/s41272-022-00415-5.
- [6] C. Binsriavanich and N. Phumchusri, "An analysis of retail promotional pricing effectiveness using agent-based modeling," *Journal of Revenue and Pricing Management*, vol. 24, no. 1, pp. 60–79, 2025, doi: 10.1057/s41272-024-00512-7.
- [7] A. Agussalim, K. P. R. Wardhani, and A. Anjani, "Sales Product Clustering Using RFM Calculation Model And K-Means Algorithm on Primskystore," *Technium: Romanian Journal of Applied Sciences and Technology*, vol. 16, pp. 176–182, 2023, doi: 10.47577/technium.v16i.9978.
- [8] C. P. Ufeli, M. U. Sattar, R. Hasan, and S. Mahmood, "Enhancing Customer Segmentation Through Factor Analysis of Mixed Data (FAMD)-Based Approach Using K-Means and Hierarchical Clustering Algorithms," *Information*, vol. 16, no. 6, art. 441, 2025, doi: 10.3390/info16060441.
- [9] V. Wulandari *et al.*, "Comparison of Density-Based Spatial Clustering of Applications with Noise (DBSCAN), K-Means and X-Means Algorithms on Shopping Trends Data," *IJATIS: Indonesian Journal of Applied Technology and Innovation Science*, vol. 1, no. 1, pp. 1–8, 2024, doi: 10.57152/ijatis.v1i1.1135.
- [10] R. M. F. Lubis *et al.*, "Agglomerative Hierarchical Clustering (AHC) Method for Data Mining Sales Product Clustering," *Building of Informatics, Technology and Science (BITS)*, vol. 5, no. 1, 2023, doi: 10.47065/bits.v5i1.3569.
- [11] J. M. John, O. Shobayo, and B. Ogunleye, "An Exploration of Clustering Algorithms for Customer Segmentation in the UK Retail Market," *Analytics*, vol. 2, no. 4, pp. 809–823, 2023, doi: 10.3390/analytics2040042.
- [12] E. Febrianty, L. Awalina, and W. I. Rahayu, "Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail," *Jurnal Teknologi dan Informasi (JATI)*, vol. 13, no. 2, 2023, doi: 10.34010/jati.v13i2.
- [13] Y. Arvio, J. E. Riwurohi, and E. K. Dewantara, "Pemodelan Segmentasi Transaksi Jual Beli Produk Menggunakan Pendekatan Model K-Means dan Subtractive Clustering Studi Kasus Survey Pada Beberapa Cabang Optik Retail," *PETIR*, vol. 16, no. 1, pp. 39–47, 2023, doi: 10.33322/petir.v16i1.1897.
- [14] R. S. Iswandar and A. D. Indriyanti, "Sales Performance Classification of Promotional Products Using Data Mining," *Journal of Emerging Information Systems and Business Intelligence (JEISBI)*, vol. 7, no. 2, pp. 213–220, 2026, doi: 10.26740/jeisbi.v7i2.75897.
- [15] D. Sugianti *et al.*, "Analisis Cluster Penentuan Promosi Produk Pasca Pandemi Covid-19 dengan Metode K-Means," *Jurnal Manajemen Informatika & Komputerisasi Akuntansi*, vol. 7, no. 1, pp. 20–24, 2023, doi: 10.46880/jmika.Vol7No1.pp20-24.
- [16] I. Pii, N. Suarna, and N. Rahaningsih, "Penerapan Data Mining pada Penjualan Produk Pakaian Dameyra Fashion Menggunakan Metode K-Means Clustering," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 1, 2023, doi: 10.36040/jati.v7i1.6336.
- [17] D. Bakkelund, "Order Preserving Hierarchical Agglomerative Clustering," *Machine Learning*, vol. 111, pp. 1851–1901, 2022, doi: 10.1007/s10994-021-06125-0.

- [18] F. Martinez-Plumed *et al.*, "CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 8, pp. 3048–3061, 2021, doi: 10.1109/TKDE.2019.2962680.
- [19] D. Chicco, A. Campagner, A. Spagnolo, D. Ciucci, and G. Jurman, "The Silhouette coefficient and the Davies-Bouldin index are more informative than Dunn index, Calinski-Harabasz index, Shannon entropy, and Gap statistic for unsupervised clustering internal evaluation of two convex clusters," *PeerJ Computer Science*, vol. 11, art. e3309, 2025, doi: 10.7717/peerj-cs.3309.
- [20] A. Anfossi and D. Chicco, "An easy guide to the Davies-Bouldin index for unsupervised internal clustering evaluation," *Discover Computing*, vol. 29, no. 1, art. 212, 2026, doi: 10.1007/s10791-026-10094-0.
- [21] S. R. Anggraeni, "Implementasi Algoritma Clustering DBSCAN Terhadap Pola Navigasi Pengguna di Perpustakaan Digital untuk Mengungkap Zona Buta Akses Informasi dan Optimalisasi Antarmuka Sistem," *SKANIKA: Sistem Komputer dan Teknik Informatika*, vol. 8, no. 2, pp. 232–243, 2025, doi: 10.36080/skanika.v8i2.3524.
- [22] M. Y. Wulang, S. A. Wibowo, and E. H. Susanto, "SIG dengan K-Means++ untuk Klasterisasi Pengembangan UMKM Kain Tenun (Studi Kasus: Kabupaten Nagekeo)," *SKANIKA: Sistem Komputer dan Teknik Informatika*, vol. 9, no. 1, pp. 47–60, 2026, doi: 10.36080/skanika.v9i1.3630.
- [23] D. Krisnawati *et al.*, "Penerapan Text Mining pada Analisis Sentimen Pengguna Twitter Layanan Transportasi Online Menggunakan Metode Density Based Spatial Clustering of Applications with Noise (DBSCAN) dan K-Means," *J Statistika*, vol. 15, no. 1, pp. 184–194, 2022, doi: 10.36456/jstat.vol15.no1.a5983.