Journal of Information System, Applied, Management, Accounting and Research.
https://publikasi.binainternusa.org/index.php/jisamar ,
jisamar@stmik.jayakarta.ac.id , jisamar2017@gmail.com
e-ISSN: 2598-8719 (Online), p-ISSN: 2598-8700 (Printed) , Vol. 10 No.4 (November 2026)

PERBANDINGAN ALGORITMA K-MEANS DAN HIERARCHICAL CLUSTERING UNTUK SEGMENTASI PELANGGAN PADA MALL CUSTOMER DATASET

Putri Nazwa Salsabila Sufi[1*], Hafiyyan Putra Pratama[2]

Program Studi Sistem Telekomunikasi[12]
Universitas Pendidikan Indonesia
Kampus Daerah Purwakarta

Correspondent Author: putrinazwa.29@upi.edu,

Email Author: haffiyan@upi.edu

Received: July 15, 2026**. Revised:** October 28, 2026**. Accepted:** November 9, 2026**. Issue Period:** Vol.10 No.4 (2026), Pp. 1062-1071

Abstrak: Pusat perbelanjaan memerlukan strategi segmentasi pelanggan yang tepat untuk meningkatkan efektivitas pemasaran dan loyalitas pelanggan di tengah persaingan industri ritel yang semakin ketat. Penelitian ini membandingkan dua algoritma clustering, yaitu K-Means dan Hierarchical Clustering dengan Ward linkage, diterapkan pada Mall Customer Dataset dari Kaggle yang terdiri atas 200 data pelanggan dengan fokus pada fitur Annual Income dan Spending Score. Praproses data mencakup pemeriksaan missing values, label encoding variabel Gender, dan normalisasi fitur menggunakan StandardScaler untuk menyetarakan skala variabel dalam perhitungan jarak Euclidean. Jumlah kluster optimal ditentukan melalui Elbow Method untuk K-Means dan analisis dendrogram untuk Hierarchical Clustering, keduanya menghasilkan K=5. Evaluasi menggunakan Silhouette Score menunjukkan K-Means memperoleh nilai 0,5547 dan Hierarchical Clustering 0,5538 dengan selisih sebesar 0,0009. Kedua algoritma secara konsisten mengidentifikasi lima segmen pelanggan, yaitu pelanggan menengah moderat, berpendapatan tinggi yang aktif berbelanja, muda antusias berpendapatan rendah, kaya konservatif, dan hemat berpendapatan rendah. Kemiripan hasil disebabkan oleh struktur klaster yang jelas secara visual, kualitas data yang bersih tanpa missing values dan duplikat, serta kesamaan objektif matematis kedua algoritma dalam meminimalkan within-cluster variance. K-Means lebih direkomendasikan untuk dataset berskala besar karena efisiensi komputasinya, sedangkan Hierarchical Clustering unggul dalam visualisasi struktural melalui dendrogram dan sifatnya yang deterministik. Implikasi praktis berupa panduan strategi pemasaran berbasis segmen yang dapat langsung diterapkan oleh manajemen pusat perbelanjaan.

Kata kunci: K-Means Clustering; Hierarchical Clustering; Segmentasi Pelanggan; Silhouette Score; Mall Customer Dataset; Data Mining

Abstract: *Shopping centers require appropriate customer segmentation strategies to improve marketing effectiveness and customer loyalty amid increasingly competitive retail industry conditions. This study compares two clustering algorithms, namely K-*

DOI: 10.52362/jisamar.v10i4.2448
Ciptaan disebarluaskan di bawah Lisensi Creative Commons Atribusi 4.0 Internasional.

Journal of Information System, Applied, Management, Accounting and Research.
https://publikasi.binainternusa.org/index.php/jisamar ,
jisamar@stmik.jayakarta.ac.id , jisamar2017@gmail.com
e-ISSN: 2598-8719 (Online), p-ISSN: 2598-8700 (Printed) , Vol. 10 No.4 (November 2026)

Means and Hierarchical Clustering with Ward linkage, applied to the Mall Customer Dataset from Kaggle consisting of 200 customer records focusing on Annual Income and Spending Score features. Data preprocessing includes missing value inspection, Gender variable label encoding, and feature normalization using StandardScaler to equalize variable scales in Euclidean distance calculations. The optimal number of clusters is determined using the Elbow Method for K-Means and dendrogram analysis for Hierarchical Clustering, both yielding K=5. Evaluation using the Silhouette Score shows K-Means achieves 0.5547 and Hierarchical Clustering 0.5538, with a negligible difference of 0.0009. Both algorithms consistently identify five customer segments: moderate middle-income, high-income active shoppers, young enthusiastic low-income, affluent conservative, and frugal low-income customers. The similarity of results is attributed to clearly visible cluster structures, clean data quality without missing values and duplicates, and the shared mathematical objective of both algorithms in minimizing within-cluster variance. K-Means is more recommended for large-scale datasets due to its computational efficiency, while Hierarchical Clustering excels in structural visualization through dendrograms and its deterministic nature. Practical implications include segment-based marketing strategies directly applicable by shopping center management.

Keywords: *K-Means Clustering; Hierarchical Clustering; Customer Segmentation; Silhouette Score; Mall Customer Dataset; Data Mining*

I. PENDAHULUAN

Lanskap bisnis ritel modern yang semakin kompetitif menuntut perusahaan untuk memahami karakteristik dan perilaku pelanggan secara lebih mendalam sebagai dasar pengambilan keputusan strategis [1], [2]. Seiring berkembangnya teknologi informasi, perusahaan ritel menghasilkan data pelanggan dalam jumlah besar yang berpotensi digunakan untuk mendukung strategi pemasaran berbasis data [3], [4]. Pemahaman mendalam terhadap pola konsumsi pelanggan menjadi faktor penting dalam mempertahankan loyalitas dan daya saing di pasar yang terus berubah [5], [6].

Segmentasi pelanggan merupakan pendekatan yang digunakan untuk mengelompokkan pelanggan yang heterogen menjadi kelompok-kelompok homogen berdasarkan karakteristik dan perilaku tertentu [1], [5]. Dengan segmentasi yang tepat, perusahaan dapat mengidentifikasi kelompok pelanggan potensial, pelanggan loyal, maupun pelanggan yang memerlukan pendekatan berbeda, sehingga strategi promosi dan personalisasi layanan dapat dirancang lebih terarah dan efisien [3], [4]. Pada konteks analisis pengunjung mall, variabel Annual Income dan Spending Score terbukti menjadi indikator utama yang merepresentasikan profil ekonomi dan perilaku belanja pelanggan [7], [6], [8].

Dalam analisis data pelanggan berskala besar, teknik data mining banyak digunakan untuk menemukan pola tersembunyi dari kumpulan data yang kompleks [9], [10]. Salah satu teknik yang paling umum diterapkan adalah clustering, yaitu metode unsupervised learning yang bertujuan mengelompokkan data berdasarkan tingkat kemiripan karakteristik tertentu [11]. Penerapan clustering pada segmentasi pelanggan memungkinkan perusahaan memperoleh wawasan mengenai pola perilaku pelanggan secara objektif tanpa memerlukan label awal pada data [10], [3].

Algoritma K-Means merupakan metode clustering berbasis partisi yang paling banyak digunakan karena kesederhanaan komputasi dan efisiensinya dalam mengelompokkan data berdasarkan jarak ke centroid [9], [5], [2]. Namun, K-Means memiliki keterbatasan seperti sensitivitas terhadap pemilihan centroid awal dan keharusan menentukan jumlah kluster sejak awal proses [12], [11]. Di sisi lain, Agglomerative Hierarchical Clustering (AHC) menggunakan pendekatan hierarkis yang membangun struktur klaster secara bertahap berdasarkan tingkat kedekatan antar data dan mampu memberikan visualisasi hubungan antar data melalui dendrogram [7], [13]. Perbedaan karakteristik kedua algoritma tersebut memungkinkan terbentuknya hasil segmentasi yang berbeda pada dataset yang sama.

DOI: 10.52362/jisamar.v10i4.2448
Ciptaan disebarluaskan di bawah Lisensi Creative Commons Atribusi 4.0 Internasional.

Journal of Information System, Applied, Management, Accounting and Research.
https://publikasi.binainternusa.org/index.php/jisamar ,
jisamar@stmik.jayakarta.ac.id , jisamar2017@gmail.com
e-ISSN: 2598-8719 (Online), p-ISSN: 2598-8700 (Printed) , Vol. 10 No.4 (November 2026)

Beberapa penelitian sebelumnya telah menerapkan algoritma clustering dalam segmentasi pelanggan ritel dan analisis perilaku konsumen. K-Means terbukti efektif dalam segmentasi pelanggan ritel berbasis perilaku dengan interpretasi bisnis yang langsung dapat diterapkan [4], serta mampu menghasilkan segmentasi berkualitas baik pada toko ritel dengan nilai Silhouette Coefficient mencapai 0,66 [1]. Kombinasi K-Means dan Hierarchical Clustering pada segmentasi berbasis RFMT juga menunjukkan adanya perbedaan karakteristik hasil antar metode pada dataset yang sama [10]. Selain itu, penerapan Hierarchical Clustering pada Mall Customer Dataset berhasil mengidentifikasi segmen pelanggan berbasis pendapatan dan pengeluaran, namun tanpa membandingkannya dengan K-Means [7]. Penelitian lain juga menunjukkan bahwa hasil perbandingan dua algoritma clustering dapat sangat bervariasi tergantung karakteristik dan struktur dataset yang digunakan [14], [13]. Namun demikian, Sebagian besar penelitian ini masih berfokus pada satu algoritma tanpa evaluasi komparatif terhadap kualitas, stabilitas, dan interpretasi bisnis antar metode pada data demografi pelanggan mall [9], [15], [11].

Berdasarkan kajian literatur tersebut, terdapat celah penelitian yang belum banyak dieksplorasi secara sistematis, yakni perbandingan langsung antara K-Means dan Hierarchical Clustering secara komparatif pada variabel demografi pelanggan mall, khususnya Annual Income dan Spending Score, disertai evaluasi kualitas klaster kuantitatif dan implikasi strategi bisnis yang konkret. Studi yang ada umumnya hanya menerapkan satu algoritma [6], [7], atau membandingkan pada dataset non-mall [9], [13], atau melakukan perbandingan tanpa menyertakan rekomendasi pemasaran berbasis segmen yang dapat ditinjaklanjuti [10], [14]. Oleh karena itu, penelitian ini memberikan kontribusi berupa: (1) evaluasi komparatif K-Means dan Hierarchical Clustering dengan Ward Linkage pada Mall Customer Dataset menggunakan protocol evaluasi yang identik; (2) analisis faktor-faktor yang menyebabkan konvergensi atau divergensi hasil algoritma; dan (3) panduan strategi pemasaran berbasis segmentasi yang dapat langsung diterapkan oleh manajemen pusat perbelanjaan. Penelitian ini berfokus pada Mall Customer Dataset dengan 200 data pelanggan menggunakan dua fitur utama, yaitu Annual Income dan Spending Score, sehingga generalisasi hasil perlu mempertimbangkan konteks dataset yang digunakan.

II. METODE DAN MATERI

Data yang digunakan dalam penelitian ini adalah Mall Customer Dataset yang diperoleh dari repositori publik Kaggle, terdiri dari 200 data pelanggan dengan atribut CustomerID, Gender, Age, Annual Income (k$), dan Spending Score (1-100) [7], [6], [8]. Atribut utama yang menjadi fokus analisis adalah Annual Income dan Spending Score. Pemilihan kedua filter ini didasarkan pada tiga pertimbangan. Pertama, keduanya secara langsung merepresentasikan kapasitas ekonomi dan perilaku konsumtif pelanggan, yang merupakan dimensi paling relevan untuk segmentasi berbasis pemasaran [5], [6]. Kedua, penggunaan dua fitur numerik kontinyu memungkinkan visualisasi hasil clustering secara langsung pada bidang dua dimensi sehingga interpretasi klaster lebih intuitif dan mudah dikomunikasikan kepada manajemen [7], [8]. Ketiga, fokus pada dua fitur utama menghindari masalah curse of dimensionality yang dapat menurunkan kualitas perhitungan jarak Euclidean [9]. Variabel Age tetap diikutsertakan sebagai atribut profiling tambahan dalam interpretasi segmen pasca-clustering, bukan sebagai fitur pembentuk klaster.

Karena K-Means bersifat non-deterministik akibat inisialisasi centroid secara acak, eksperimen dijalankan sebanyak 10 kali dengan nilai n_init=10 untuk memastikan stabilitas hasil. Parameter random_state=42 ditetapkan untuk keperluan reproduksibilitas. Pendekatan ini konsisten dengan praktik standar dalam penelitian clustering untuk meminimalkan risiko konvergensi ke solusi lokal yang tidak optimal [9], [12].

Tahap pra-pemrosesan data diawali dengan pemeriksaan dan konfirmasi tidak adanya missing values maupun data duplikat [1]. Selanjutnya dilakukan label encoding pada variabel Gender menggunakan LabelEncoder (Female=0, Male=1), serta normalisasi fitur menggunakan StandardScaler untuk menyetarakan skala variabel agar tidak ada fitur yang mendominasi perhitungan jarak Euclidean selama proses klastering [1], [13].

Metodologi penelitian ini menerapkan analisis komparatif antara algoritma K-Means dan Hierarchical Clustering [9], [14]. Pada K-Means, jumlah klaster optimal ditentukan melalui Elbow Method dengan menganalisis titik melandainya nilai inertia atau Within-Cluster Sum of Squares (WCSS) [9], [5]. Sementara itu, Hierarchical Clustering menggunakan Ward Linkage untuk meminimalkan varians internal klaster dari kedua algoritma kemudian dievaluasi menggunakan Silhouette Score untuk mengukur tingkat kepadatan dan keterpisahan antar Klaster yang dihasilkan [14], [12] [15].

DOI: 10.52362/jisamar.v10i4.2448
Ciptaan disebarluaskan di bawah Lisensi Creative Commons Atribusi 4.0 Internasional.

Journal of Information System, Applied, Management, Accounting and Research.
https://publikasi.binainternusa.org/index.php/jisamar ,
jisamar@stmik.jayakarta.ac.id , jisamar2017@gmail.com
e-ISSN: 2598-8719 (Online), p-ISSN: 2598-8700 (Printed) , Vol. 10 No.4 (November 2026)

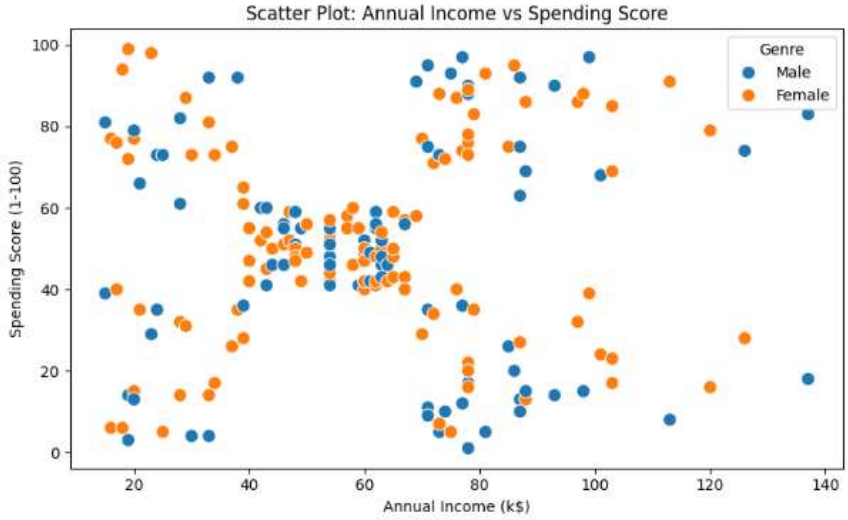
III. PEMBAHASAN DAN HASIL

3.1. Analisis Data Eksploratif

Pemeriksaan mengonfirmasi data terdiri atas 200 pelanggan dengan komposisi 112 perempuan (56%) dan 88 laki-laki (44%), tanpa missing values maupun duplikat pada seluruh atribut [1]. Kondisi ini memastikan proses clustering dapat dilakukan langsung tanpa imputasi tambahan, sehingga hasil segmentasi tidak terdistorsi oleh penanganan data hilang [1], [8].

Analisis histogram menunjukkan bahwa variabel Spending Score dan Age berdistribusi relative merata, sedangkan Annual Income sedikit condong ke kiri (left-skewed), mengindikasikan proporsi pelanggan berpendapatan menengah ke bawah yang lebih besar dalam dataset [7], [6]. Analisis boxplot mengkonfirmasi tidak adanya outlier ekstrem pada ketiga variabel numerik, menandakan kualitas data yang bersih representatif [1].

Scatter plot antara Annual Income dan Spending Score, sebagaimana ditampilkan pada Gambar 1, merupakan visualisasi paling informatif dalam tahap eksplorasi ini. Terdapat lima gumpalan titik yang terpisah secara visual dan jelas sebelum algoritma apapun diterapkan, memberikan justifikasi kuat bahwa K=5 merupakan jumlah klaster yang paling representatif untuk kedua algoritma [7], [5], [6], [8]. Kejelasan struktur visual ini juga menjadi faktor penting yang akan menjelaskan konvergensi hasil antara kedua algoritma pada tahap selanjutnya.



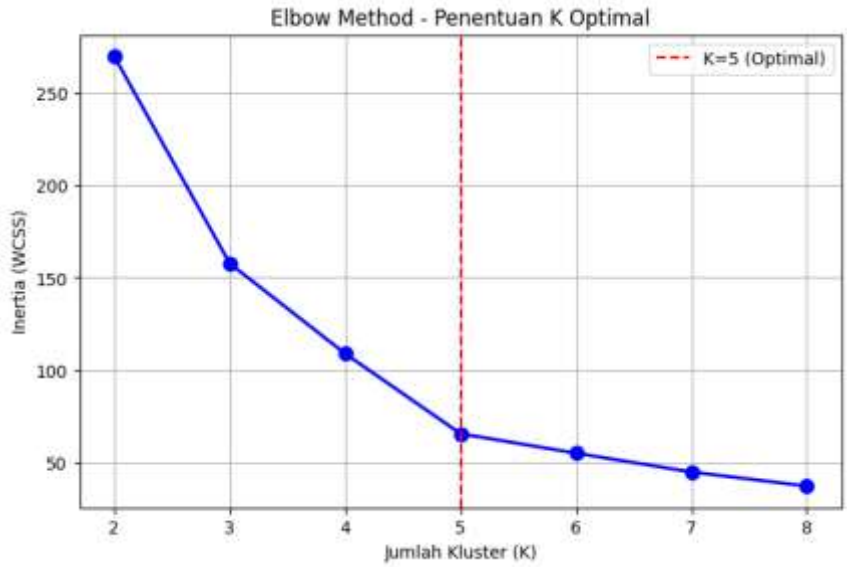
Gambar 1.Scatter Plot Income vs Spending Score

3.2. Hasil K-Means Clustering

Penentuan jumlah klaster optimal menggunakan Elbow Method menghasilkan grafik penurunan WCSS yang tajam dari K=2 hingga K=5, dengan kurva yang mulai mendatar setelahnya [9], [5], [8]. Titik siku (elbow) yang terbentuk secara jelas di K=5, sebagaimana ditampilkan pada Gambar 2, mengonfirmasi jumlah klaster optimal, konsisten dengan hasil observasi visual pada scatter plot eksplorasi. Model K-Means yang dilatih dengan n_clusters=5, random_state=42, dan n_init=10 menghasilkan total inertia sebesar 65,57.

DOI: 10.52362/jisamar.v10i4.2448
Ciptaan disebarluaskan di bawah Lisensi Creative Commons Atribusi 4.0 Internasional.

Journal of Information System, Applied, Management, Accounting and Research.
https://publikasi.binainternusa.org/index.php/jisamar ,
jisamar@stmik.jayakarta.ac.id , jisamar2017@gmail.com
e-ISSN: 2598-8719 (Online), p-ISSN: 2598-8700 (Printed) , Vol. 10 No.4 (November 2026)



Gambar 2. Grafik elbow method

Tabel 1 menyajikan karakteristik tiap klaster K-Means berdasarkan nilai rata-rata Annual Income dan Spending Score setelah denormalisasi:

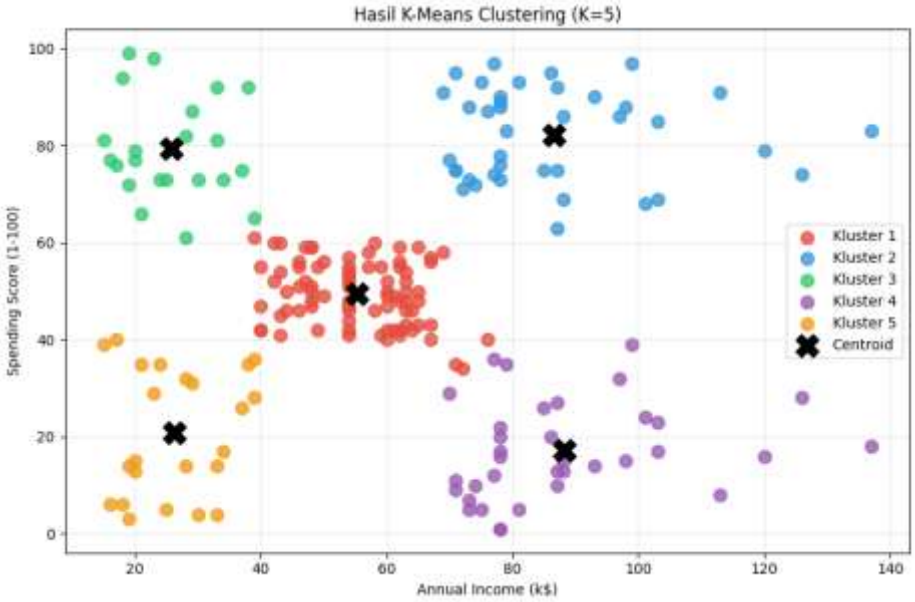
Tabel 1. Karakteristik Kluster K-Means

Klaster	Segmen	n	Income (~k$)	Score	Usia (~thn)
1	Menengah Moderat	81	55,3	49,5	42,7
2	Pendapatan Tinggi Aktif	39	86,5	82,1	32,7
3	Muda Antusias Income Rendah	22	25,7	79,4	25,3
4	Kaya Konservatif	35	88,2	17,1	40,7
5	Hemat Berpendapatan Rendah	23	26,3	21,2	45,2

Scatter plot hasil K-Means yang ditampilkan pada Gambar 3 memperlihatkan pemisahan antar klaster yang sangat jelas dengan tanda centroid pada tiap pusat klaster. Temuan lima segmen ini konsisten dengan penelitian yang mengidentifikasi pola serupa menggunakan Hierarchical Clustering pada dataset yang sama, serta sejalan dengan yang menemukan lima kelompok pelanggan serupa berdasarkan variabel Annual Income dan Spending Score [7], [6], [8].

DOI: 10.52362/jisamar.v10i4.2448
Ciptaan disebarluaskan di bawah Lisensi Creative Commons Atribusi 4.0 Internasional.

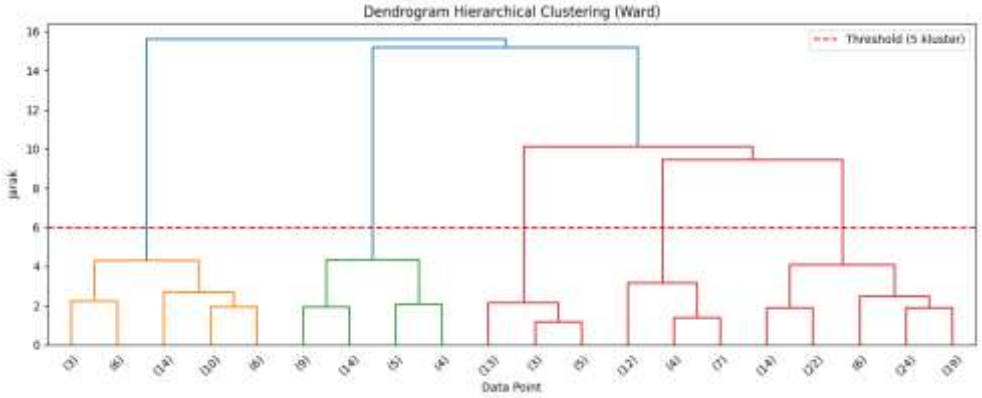
Journal of Information System, Applied, Management, Accounting and Research.
https://publikasi.binainternusa.org/index.php/jisamar ,
jisamar@stmik.jayakarta.ac.id , jisamar2017@gmail.com
e-ISSN: 2598-8719 (Online), p-ISSN: 2598-8700 (Printed) , Vol. 10 No.4 (November 2026)



Gambar 3. Scatter Plot K-Means

3.3. Hasil Hierarchical Clustering

Visualisasi dendrogram dengan Ward Linkage menampilkan struktur hierarkis penggabungan data yang jelas dan terstruktur [7], [13]. Pemilihan Ward Linkage didasarkan pada kemampuannya menghasilkan klaster yang paling seimbang dan kompak dibandingkan metode linkage lainnya, karena meminimalkan peningkatan total Within-Cluster variance pada setiap langkah penggabungan [10], [13]. Seperti yang terlihat pada Gambar 4, pemotongan dendrogram pada threshold jarak=6 memotong pohon menjadi lima cabang utama, menghasilkan K=5 klaster yang konsisten dengan hasil K-Means.



Gambar 4. Dendrogram

Tabel 2 menyajikan karakteristik tiap klaster Hierarchical Clustering, sedangkan Scatter Plot hasil klasterisasinya ditampilkan pada Gambar 5:

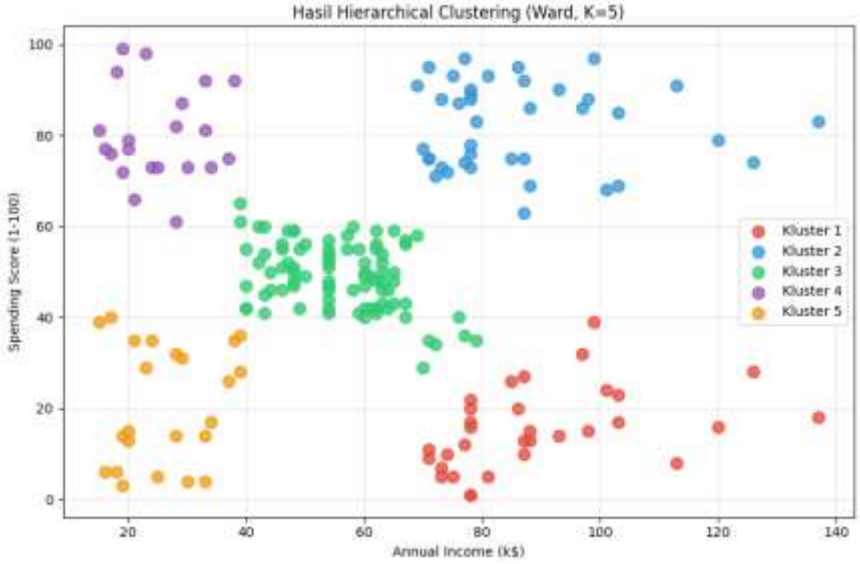
Tabel 2. Karakteristik Klaster Hierarchical Clustering

Klaster	Segmen	n	Income (~k$)	Score	Usia (~thn)
1	Menengah Moderat	85	54,9	49,1	43,1

DOI: 10.52362/jisamar.v10i4.2448
Ciptaan disebarluaskan di bawah Lisensi Creative Commons Atribusi 4.0 Internasional.

Journal of Information System, Applied, Management, Accounting and Research.
https://publikasi.binainternusa.org/index.php/jisamar ,
jisamar@stmik.jayakarta.ac.id , jisamar2017@gmail.com
e-ISSN: 2598-8719 (Online), p-ISSN: 2598-8700 (Printed) , Vol. 10 No.4 (November 2026)

2	Pendapatan Tinggi Aktif	39	86,5	82,1	32,7
3	Muda Antusias Income Rendah	21	25,3	79,8	25,1
4	Kaya Konservatif	32	89,1	15,2	41,2
5	Hemat Berpendapatan Rendah	23	26,3	21,2	45,2



Gambar 5. Scatter Plot Hierarchical

Perbandingan Tabel 1 dan Tabel 2 mengungkap dua temuan penting. Pertama, Klaster 2 (Pendapatan Tinggi Aktif) dan Klaster 5 (Hemat Berpendapatan Rendah) identik secara sempurna antara kedua metode (n=39 dan n=23), mengonfirmasi bahwa kedua segmen tersebut merupakan kelompok yang terdefinisi paling kuat dan konsisten dalam dataset ini [7]. Kedua, perbedaan utama terletak pada distribusi klaster Menengah Moderat (n=81 vs. n=85) dan Kaya Konservatif (n=35 vs. n=32), menunjukkan bahwa Hierarchical Clustering dengan Ward Linkage memberlakukan batas yang lebih ketat pada kelompok berpendapatan tinggi, sehingga memindahkan beberapa titik data perbatasan ke klaster Menengah Moderat [14], [13].

3.4. Analisis Kesamaan Hasil Antar Algoritma

Untuk mengukur tingkat kesepakatan antara hasil K-Means dan Hierarchical Clustering secara kuantitatif, digunakan Adjusted Rand Index (ARI). ARI menghasilkan nilai dalam rentang [-1, 1], di mana nilai 1 menunjukkan kesepakatan sempurna, dan nilai 0 menunjukkan kesepakatan setara acak [12]. Perhitungan ARI menggunakan sklearn.metrics.adjusted_rand_score menghasilkan nilai 0,9420, yang tergolong sangat tinggi. Nilai ini mengonfirmasi secara kuantitatif bahwa kedua algoritma tidak hanya menghasilkan jumlah klaster yang sama (K=5), tetapi juga mengalokasikan sebagian besar data point ke klaster yang secara substansial serupa. Perbedaan penempatan data point terjadi hanya pada zona perbatasan antara klaster Menengah Moderat dan Kaya Konservatif, konsisten dengan perbedaan distribusi anggota pada Tabel 1 dan Tabel 2.

3.5. Evaluasi Komparatif

Tabel 3 merangkum perbandingan performa kuantitatif dan karakteristik kualitatif kedua algoritma:

Tabel 3. Perbandingan K-Means vs Hierarchical Clustering

Aspek	K-Means	Hierarchical Clustering

DOI: 10.52362/jisamar.v10i4.2448
Ciptaan disebarluaskan di bawah Lisensi Creative Commons Atribusi 4.0 Internasional.

Journal of Information System, Applied, Management, Accounting and Research.
https://publikasi.binainternusa.org/index.php/jisamar ,
jisamar@stmik.jayakarta.ac.id , jisamar2017@gmail.com
e-ISSN: 2598-8719 (Online), p-ISSN: 2598-8700 (Printed) , Vol. 10 No.4 (November 2026)

Silhouette Score	0,5547	0,5538
Selisih Score	—	0,0009 (dapat diabaikan)
Inertia/WCSS	65,57	—
Jumlah Klaster Optimal	K=5 (Elbow Method)	K=5 (Dendrogram)
Kompleksitas Waktu	O(n·k·i) — linear	O(n^2) — kuadratik
Deterministik	Tidak (acak)	Ya (Ward Linkage)
Visualisasi Hierarki	Tidak	Ya (Dendrogram)
Skalabilitas Dataset Besar	Tinggi	Rendah

Perbedaan Silhouette Score antara K-Means (0,5547) dan Hierarchical Clustering (0,5538) hanya sebesar 0,0009, menunjukkan kinerja yang hampir setara secara statistik [14], [12]. Nilai 0,55 pada kedua algoritma tergolong dalam rentang moderate good structure (0,51–0,70) yang mengindikasikan pemisahan klaster yang cukup baik [12], [15]. Konvergensi hasil ini dapat dijelaskan melalui tiga faktor utama.

Pertama, struktur klaster dalam data sangat jelas secara visual (five clearly separated blobs) sehingga dapat dikenali oleh pendekatan algoritma yang berbeda sekalipun, baik partisi maupun hierarki [14], [11]. Kedua, kualitas data yang bersih tanpa missing values dan outlier ekstrem sehingga tidak ada distorsi yang dapat memengaruhi kedua algoritma secara asimetris [1]. Ketiga, kesamaan objektif matematis, baik K-Means maupun Hierarchical Ward sama-sama meminimalkan Within-Cluster variance, sehingga pada data yang terstruktur jelas keduanya konvergen ke solusi yang serupa [10], [13].

Temuan ini sejalan dengan temuan yang menunjukkan K-Means cenderung sedikit lebih unggul pada data berstruktur klaster jelas [9], serta temuan yang menemukan konvergensi hasil antara K-Means dan HC pada dataset pelanggan dengan distribusi yang terpisah baik [10]. Namun berbeda dengan lain yang menemukan selisih Silhouette Score yang jauh lebih besar antara kedua algoritma pada dataset dengan distribusi lebih kompleks [14], [13]. Perbedaan ini menegaskan bahwa karakteristik dan struktur dataset merupakan faktor penentu utama dalam konsistensi hasil perbandingan algoritma clustering [15], [11].

K-Means lebih direkomendasikan untuk penerapan pada dataset berskala besar karena kompleksitas waktu yang linear O(n·k·i) [9], [12], [2]. Hierarchical Clustering lebih direkomendasikan untuk tahap eksplorasi data awal karena sifatnya yang deterministik dan kemampuan visualisasi struktural melalui dendrogram yang membantu memahami hubungan hierarkis antar segmen pelanggan [7], [13], [11].

3.6. Interpretasi Segmen dan Rekomendasi Strategi Pemasaran

Lima segmen pelanggan yang teridentifikasi secara konsisten oleh kedua algoritma memberikan panduan strategi pemasaran yang dapat langsung ditindaklanjuti oleh manajemen pusat perbelanjaan [5], [3], [4]:

Segmen 1, Menengah Moderat (n=81–85, income ~$55 rb, score ~49): Merupakan basis pelanggan inti terbesar yang mewakili sekitar 40–43% total pelanggan. Segmen ini bersifat moderat dalam pengeluaran dengan pendapatan menengah yang stabil. Strategi yang direkomendasikan meliputi program loyalitas berbasis poin kumulatif, promosi lintas kategori produk, dan penawaran bundling bernilai tinggi dengan harga terjangkau untuk mendorong peningkatan frekuensi kunjungan [10], [2], [14].

Segmen 2, Pendapatan Tinggi Aktif Berbelanja (n=39, income ~$86,5 rb, score ~82): Segmen paling menguntungkan dengan spending score tertinggi dan usia rata-rata termuda (32,7 tahun). Segmen ini responsif terhadap pengalaman belanja premium dan produk eksklusif. Strategi yang direkomendasikan adalah layanan concierge premium, program membership eksklusif dengan benefit khusus, akses early bird produk luxury, dan komunikasi personal berbasis preferensi pembelian sebelumnya [5], [6], [4].

Segmen 3, Muda Antusias Berpendapatan Rendah (n=21–22, income ~$25 rb, score ~79): Segmen dengan motivasi belanja sangat tinggi meskipun pendapatan terbatas, dengan usia rata-rata termuda (25,3 tahun). Segmen ini merupakan pelanggan potensial jangka panjang seiring peningkatan karier. Strategi yang direkomendasikan mencakup opsi cicilan tanpa bunga, diskon khusus segmen pelajar/mahasiswa, program referral, dan kampanye pemasaran digital berbasis media sosial yang viral [3], [2], [4].

Segmen 4, Kaya Konservatif (n=32–35, income ~$88–89 rb, score ~15–17): Segmen berpendapatan tinggi namun dengan spending score terendah, menunjukkan perilaku belanja yang sangat selektif dan deliberatif. Segmen ini tidak responsif terhadap promosi massal namun menghargai eksklusivitas dan kualitas. Strategi yang direkomendasikan meliputi layanan personal shopper, undangan acara privat dan preview koleksi eksklusif,

DOI: 10.52362/jisamar.v10i4.2448
Ciptaan disebarluaskan di bawah Lisensi Creative Commons Atribusi 4.0 Internasional.

komunikasi pemasaran yang menekankan kualitas dan keunikan produk, serta program keanggotaan VIP [5], [15], [4].

Segmen 5, Hemat Berpendapatan Rendah (n=23, income ~$26 rb, score ~21): Segmen yang sangat sensitif terhadap harga dengan spending score dan pendapatan sama-sama rendah. Segmen ini tetap bernilai sebagai basis volume transaksi. Strategi yang direkomendasikan adalah flash sale terjadwal dengan notifikasi push, program diskon akhir musim, penawaran produk bundle hemat yang menonjolkan nilai ekonomis, dan program cicilan ringan untuk pembelian di atas ambang tertentu [2], [4].

IV. KESIMPULAN

Penelitian ini melakukan evaluasi komparatif antara K-Means dan Hierarchical Clustering pada Mall Customer Dataset menggunakan praproses, fitur, dan protokol evaluasi yang identik. Hasil evaluasi Silhouette Score menunjukkan kinerja yang hampir setara antara K-Means (0,5547) dan Hierarchical Clustering (0,5538) dengan selisih yang dapat diabaikan sebesar 0,0009 [14], [12]. Keduanya secara konsisten menghasilkan lima segmen pelanggan yang bermakna dengan karakteristik yang sebagian besar serupa, dikonfirmasi secara kuantitatif oleh nilai Adjusted Rand Index sebesar 0,9420 yang menunjukkan tingkat kesepakatan sangat tinggi antar kedua algoritma.

Konvergensi hasil ini disebabkan oleh tiga faktor utama: (1) struktur klaster yang sangat jelas secara visual sehingga dapat dikenali oleh pendekatan algoritma yang berbeda [14], [11]; (2) kualitas data yang bersih tanpa distorsi [1]; dan (3) kesamaan objektif matematis dalam meminimalkan Within-Cluster variance [10], [13]. Temuan ini berbeda dari hasil yang menemukan selisih jauh lebih besar, menegaskan bahwa karakteristik dataset sangat menentukan konsistensi perbandingan algoritma clustering [14], [13].

K-Means lebih direkomendasikan untuk segmentasi pada dataset berskala besar karena efisiensi komputasi yang lebih tinggi [9], [12], [2]. Hierarchical Clustering lebih direkomendasikan untuk eksplorasi data awal karena sifatnya yang deterministik dan kemampuan visualisasi hierarkis melalui dendrogram [4], [8]. Lima segmen yang teridentifikasi memberikan panduan strategi pemasaran konkret yang dapat langsung diterapkan oleh manajemen pusat perbelanjaan [5], [3], [2], [6], [4].

Penelitian ini memiliki keterbatasan pada cakupan dataset yang terbatas pada 200 data dari satu pusat perbelanjaan dengan dua fitur utama, sehingga generalisasi hasil perlu dilakukan dengan hati-hati pada konteks yang berbeda. Untuk penelitian selanjutnya, disarankan untuk memperluas fitur analisis dengan menambahkan variabel usia dan frekuensi kunjungan [5], [15], mengeksplorasi DBSCAN sebagai pembanding algoritma berbasis densitas [9], [12], serta menguji pendekatan komparatif ini pada dataset pusat perbelanjaan di Indonesia yang lebih representatif [3], [2].

REFERENSI

[1] S. L. Achmad, A. Fauzi, R. Rahmat, and J. Indra, "Segmentasi Pelanggan Menggunakan K-Means Clustering Di Toko Retail," *J. Tek. Inf. dan Komput.*, vol. 7, no. 2, p. 736, 2024, doi: 10.37600/tekinkom.v7i2.1226.

[2] A. P. Sanevil and S. S. Hidayah, "Penerapan Teknik Data Mining Clustering Pengunjung Mall dengan Algoritma K-Means," *J. Teknol. Sist. Inf.*, vol. 6, no. 1, pp. 11–20, 2025, doi: 10.35957/jtsi.v6i1.10128.

[3] H. Safitri, S. Putri Lenggo Geni, F. Merry, and M. Wati, "Penerapan K-Means Clustering untuk Segmentasi Konsumen E-Commerce Berdasarkan Pola Pembelian," *JUKI J. Komput. dan Inform.*, vol. 7, no. 1, pp. 89–99, 2025.

[4] R. D. Nugraha, D. D. Adelia, and D. Rivaldi, "Segmentasi Pelanggan Retail Berbasis Perilaku menggunakan Algoritma K-Means Clustering," *Digit. Transform. Technol.*, vol. 5, no. 2, pp. 141–148, 2025, [Online]. Available: https://itscience.org/jurnal/index.php/digitech/article/view/6340

[5] B. T. Kristanti, A. Junaidi, and E. P. Mandyartha, "Implementasi K-Means Clustering Dalam Segmentasi Pelanggan Berdasarkan Usia, Pendapatan, Dan Model Rfm (Studi Kasus: Lantikya Store Jombang)," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4677.

[6] H. Muhammad, H. Haryadi, and Sularno, "Segmentasi Pelanggan Berdasarkan Pendapatan dan Skor Pengeluaran Menggunakan Metode K-Means," *J. Sains Dan Teknol.*, vol. 4, no. 3, pp. 325–337, 2024, doi: 10.47233/jsit.v4i3.2468.

[7] M. A. Pakpahan, S. Sahid, M. M. F. Simanullang, and R. P. Winanda, "Analisis Hierarchical Clustering untuk Segmentasi Pelanggan pada Dataset Mall Customers," *J. Sist. Inf. dan Inform.*, vol. 4, no. 1, pp. 58–63, 2026,

DOI: 10.52362/jisamar.v10i4.2448

Ciptaan disebarluaskan di bawah Lisensi Creative Commons Atribusi 4.0 Internasional.

doi:10.47233/jiska.v4i1.2615.

[8] R. Eva Yuliana, D. Mariatul Ulya, and M. Jamhuri, "Mall Customer Segmentation Using K-Means Clustering Optimized by the Elbow Method," *J. Ris. Mhs. Mat.*, vol. 4, no. 5, pp. 273–286, 2025, doi: 10.18860/jrmm.v4i5.33389.

[9] S. D. K. Wardani, A. S. Ariyanto, M. Umroh, and D. Rolliawati, "Comparison of K-Means, Db Scanner & Hierarchical Clustering Method Results for Market Segmentation Analysis," *JIKO (Jurnal Inform. dan Komputer)*, vol. 7, no. 2, p. 191, 2023.

[10] I. K. Y. Triantara, M. A. Raharja, and I. B. G. Sarasvananda, "Segmentasi Pelanggan Berbasis RFMT Menggunakan K-Means dan Hierarchical Clustering," *J. Nas. Teknol. Inf. dan Apl.*, vol. 3, no. 4, pp. 917–924, 2025.

[11] F. D. Wahyuningtyas, A. Arafat, A. Stiawan, and D. Rolliawati, "Komparasi Algoritma Hierarchical, K-Means, dan DBSCAN pada Analisis Data Penjualan Melalui Facebook," *Explor. J. Sist. Inf. dan Telemat.*, vol. 14, no. 1, p. 7, 2023, doi: 10.36448/jsit.v14i1.2931.

[12] M. S. Hasibuan, A. H. Lubis, and M. N. Sari, "Perbandingan algoritma clustering dbscan dan k-means dalam pengelompokan siswa terbaik Comparison of dbscan and k-means clustering algorithms in grouping the best students," *INFOTECH J. Inform. Teknol.*, vol. 5, pp. 301–309, 2024.

[13] T. A. Toto, "Perbandingan Algoritma K-Means Dan Agglomerative Hierarchical Clustering Untuk Pengelompokan Daerah Penghasil Padi Di Indonesia," *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 3, 2025, doi: 10.23960/jitet.v13i3.7251.

[14] N. L. A. N. Dewi, A. Z. Assyadida, S. M. Witanto, M. Nasrudin, and K. M. Hindrayani, "Komparasi Hasil Segmentasi Metode K-Means dan Agglomerative Hierarchical Terhadap Provinsi di Indonesia Berdasarkan Profil Perjalanan Wisata Tahun 2024," *STATMAT J. Stat. dan Mat.*, vol. 7, no. 3, pp. 482–502, 2025, doi: 10.32493/sm.v7i3.49999.

[15] A. Ramadhan, F. Achmad, I. Zulkarnain, and M. Aritsugi, "Evaluation of K-Means, DBSCAN, and Hierarchical Clustering for Strategic Segmentation of Tourism SMEs in Rembang, Indonesia," *J. Tek. Inform.*, vol. 6, no. 3, pp. 1605–1630, 2025, doi: 10.52436/1.jutif.2025.6.3.4602.

DOI: 10.52362/jisamar.v10i4.2448
Ciptaan disebarluaskan di bawah Lisensi Creative Commons Atribusi 4.0 Internasional.