

Application of K-means Clustering Algorithm on Customer Segmentation for Effective Marketing Strategy

Akhiyar Waladi

Universitas Jambi, Fakultas Sains dan Teknologi, Jambi, Indonesia

E-mail: akhiyar.waladi@unja.ac.id

*Corresponding Author

Raffi Rizaldi

Universitas Jambi, Fakultas Sains dan Teknologi, Jambi, Indonesia

E-mail: raffir6@gmail.com

Mochammad Arief Hermawan Sutoyo

Universitas Jambi, Fakultas Sains dan Teknologi, Jambi, Indonesia

E-mail: ariefhermawansutoyo@unja.ac.id

Received: March, 2025; Accepted: June, 2025; Published: June, 2025

Abstract: This study discusses the application of K-Means Clustering algorithm in customer segmentation to develop more effective and personalized marketing strategies. In the competitive digital era, companies require deep understanding of customer characteristics to design targeted campaigns. This study uses a dataset of 2,627 customers with 10 demographic and behavioral variables, including age, marital status, education level, work experience, spending score, and family size. The research methodology includes comprehensive data preprocessing (handling missing values, encoding categorical variables, and feature normalization), determining optimal clusters using Elbow method and Silhouette Score, and implementing the K-Means algorithm. The analysis results identified four distinct customer segments: (1) Cluster 0 - premium customers with high purchasing power (12.9%, spending score 2.16), (2) Cluster 1 - mainstream segment with moderate financial stability (47.6%, spending score 1.91), (3) Cluster 2 - large families with moderate spending (22.9%, spending score 1.00), and (4) Cluster 3 - young consumers with growth potential (22.6%, spending score 1.02). Evaluation using Silhouette Score (0.174), Davies-Bouldin Index, and Calinski-Harabasz Index demonstrates good segmentation quality with clear cluster separation. Each segment has unique characteristics that enable the development of specific marketing strategies, such as digital campaigns for young segments, loyalty programs for mainstream market, and premium offerings for high-value customers. This research contributes to the development of data-driven marketing strategies and provides a practical framework for implementing customer segmentation using machine learning in the Indonesian business context.

Keywords: Customer Segmentation, K-Means Clustering, Marketing Strategy, Machine Learning, Behavioral Analytics

1. INTRODUCTION

Di era digital, perusahaan menghadapi persaingan yang semakin ketat dan perlu memahami pelanggan secara mendalam untuk menyusun strategi pemasaran yang efektif. Segmentasi pelanggan memungkinkan perusahaan mengelompokkan pelanggan berdasarkan karakteristik tertentu, sehingga kampanye pemasaran dapat disesuaikan dengan kebutuhan tiap kelompok, yang pada akhirnya meningkatkan efektivitas dan efisiensi pemasaran [1].

Tradisional, segmentasi dilakukan berdasarkan karakteristik demografis sederhana, seperti usia dan jenis kelamin. Namun, pendekatan ini kurang mampu menangkap kompleksitas perilaku pelanggan saat ini. Dengan berkembangnya teknik data mining, algoritma *K-Means Clustering* kini banyak digunakan untuk segmentasi pelanggan yang lebih mendalam. *K-Means* adalah algoritma yang mengelompokkan data ke dalam beberapa cluster berdasarkan kemiripan atribut, membuatnya cocok untuk analisis segmentasi.

Penelitian ini bertujuan untuk menerapkan algoritma *K-Means Clustering* pada data pelanggan dengan variabel seperti usia, status pernikahan, pengeluaran, dan ukuran keluarga. Dengan metode ini, diharapkan perusahaan dapat mengidentifikasi segmen pelanggan yang berbeda dan merancang strategi pemasaran yang lebih relevan bagi tiap kelompok. Metode Elbow digunakan untuk menentukan jumlah cluster optimal, sehingga setiap cluster memiliki

karakteristik unik yang dapat dioptimalkan dalam strategi pemasaran [2].

2. LITERATURE REVIEW

2.1. Segmentasi Pelanggan

Segmentasi pelanggan adalah proses pengelompokan pelanggan berdasarkan karakteristik atau perilaku yang sama, sehingga perusahaan dapat menargetkan kelompok tertentu dengan strategi pemasaran yang lebih relevan. Segmentasi bertujuan untuk meningkatkan efektivitas pemasaran dengan memahami kebutuhan spesifik dari setiap kelompok pelanggan [1]. Secara umum, segmentasi pelanggan dapat dibagi menjadi empat kategori utama: demografis, geografis, psikografis, dan perilaku.

- **Segmentasi Demografis:** Ini adalah jenis segmentasi yang paling umum, di mana pelanggan dikelompokkan berdasarkan karakteristik demografis seperti usia, jenis kelamin, pendapatan, status pernikahan, dan pendidikan. Meskipun segmentasi demografis mudah diimplementasikan dan sering kali efektif, segmentasi ini seringkali kurang fleksibel dalam menangkap perbedaan perilaku yang lebih halus antar pelanggan [3].
- **Segmentasi Geografis:** Segmentasi ini mengelompokkan pelanggan berdasarkan lokasi geografis seperti negara, kota, atau wilayah tertentu. Jenis segmentasi ini sering digunakan untuk perusahaan yang beroperasi di banyak lokasi atau pasar internasional, karena memungkinkan penyesuaian kampanye pemasaran sesuai dengan kebudayaan, iklim, atau preferensi lokal [3].
- **Segmentasi Psikografis:** Segmentasi psikografis memperhitungkan variabel seperti gaya hidup, minat, dan kepribadian pelanggan. Segmentasi ini memberi pemahaman yang lebih mendalam tentang pola pikir dan nilai-nilai pelanggan, yang dapat membantu dalam menciptakan kampanye pemasaran yang lebih personal.
- **Segmentasi Perilaku:** Berdasarkan perilaku konsumen seperti frekuensi pembelian, pola penggunaan, dan loyalitas, segmentasi perilaku bertujuan untuk mengidentifikasi pelanggan yang sering membeli, pelanggan yang loyal, atau pelanggan baru. Segmentasi ini dianggap sangat efektif dalam mendorong loyalitas karena memungkinkan perusahaan menargetkan pelanggan berdasarkan kebiasaan mereka.

2.2. K-Means Clustering Algorithm

Clustering adalah salah satu metode dalam data mining yang bertujuan untuk mengelompokkan data ke dalam kelompok-kelompok berdasarkan kesamaan [4]. Di antara berbagai teknik *clustering*, algoritma *K-Means Clustering* merupakan salah satu metode yang paling populer dan banyak digunakan. *K-Means* adalah algoritma partisi yang bekerja dengan meminimalkan variansi di dalam cluster, di mana data yang mirip ditempatkan dalam satu cluster yang sama, sementara data yang berbeda ditempatkan di *cluster* yang berbeda.

- **Cara Kerja K-Means Clustering:** *K-Means Clustering* bekerja dengan cara memilih sejumlah k sebagai jumlah cluster yang diinginkan, kemudian secara iteratif memperbarui *centroid* dari setiap cluster hingga posisi *centroid* stabil. Proses ini melibatkan dua langkah utama dalam setiap iterasi, yaitu (1) pengelompokan data berdasarkan jarak terdekat dengan *centroid*, dan (2) perhitungan ulang *centroid* dari masing-masing cluster. Algoritma ini berhenti ketika *centroid* tidak berubah lagi atau ketika jumlah iterasi maksimum tercapai [5].
- **Kelebihan dan Kekurangan K-Means Clustering:** Kelebihan *K-Means* terletak pada kemudahannya dalam implementasi, efisiensi komputasi, dan hasil yang relatif baik pada data yang memiliki pola cluster yang jelas. Namun, algoritma ini juga memiliki beberapa kekurangan, antara lain, algoritma ini sensitif terhadap titik awal *centroid* yang dipilih secara acak, sehingga hasil yang diperoleh mungkin berbeda setiap kali algoritma dijalankan. Selain itu, *K-Means* memiliki kelemahan dalam menangani data dengan bentuk cluster yang tidak bulat atau variansi yang tidak seragam [3].
- **Metode Elbow untuk Menentukan Jumlah Cluster:** Salah satu tantangan dalam *K-Means Clustering* adalah menentukan jumlah cluster yang optimal. Metode Elbow adalah teknik yang digunakan untuk menentukan jumlah cluster dengan menganalisis perubahan *inertia* atau *within-cluster sum of squares* (WCSS). Ketika grafik WCSS diplot terhadap jumlah cluster, titik di mana penurunan *inertia* melambat (membentuk "elbow") menunjukkan jumlah cluster yang optimal [6].

2.3. Penerapan K-Means Clustering dalam Segmentasi Pelanggan

Algoritma *K-Means Clustering* telah banyak diterapkan dalam segmentasi pelanggan di berbagai industri, seperti ritel, perbankan, dan telekomunikasi, untuk memahami pola pembelian dan preferensi pelanggan. Segmentasi berbasis *K-Means* memungkinkan perusahaan untuk mengidentifikasi kelompok pelanggan yang berbeda, seperti pelanggan yang sering membeli, pelanggan dengan pengeluaran tinggi, atau pelanggan yang hanya sesekali membeli produk. Dengan mengetahui karakteristik dari setiap kelompok ini, perusahaan dapat merancang kampanye pemasaran yang lebih relevan dan personal [7].

- **Contoh Penerapan di Industri Ritel:** Dalam industri ritel, *K-Means Clustering* digunakan untuk mengelompokkan pelanggan berdasarkan frekuensi belanja, jenis produk yang dibeli, dan jumlah pengeluaran.

Sebagai contoh, penelitian yang dilakukan oleh [8] menunjukkan bahwa segmentasi berbasis K-Means mampu mengidentifikasi kelompok pelanggan yang berbelanja secara berkala dan kelompok yang hanya melakukan pembelian pada musim-musim tertentu. Informasi ini berguna bagi perusahaan ritel untuk merancang program loyalitas yang lebih efektif [8].

- **Contoh Penerapan di Industri Perbankan:** Di sektor perbankan, K-Means digunakan untuk mengidentifikasi segmen pelanggan yang memiliki potensi untuk menggunakan layanan tertentu, seperti kartu kredit atau pinjaman. Penelitian sebelumnya menunjukkan bahwa segmentasi pelanggan berbasis K-Means dapat membantu bank dalam menargetkan nasabah yang memiliki skor kredit tinggi dan potensi pengeluaran yang besar [1].

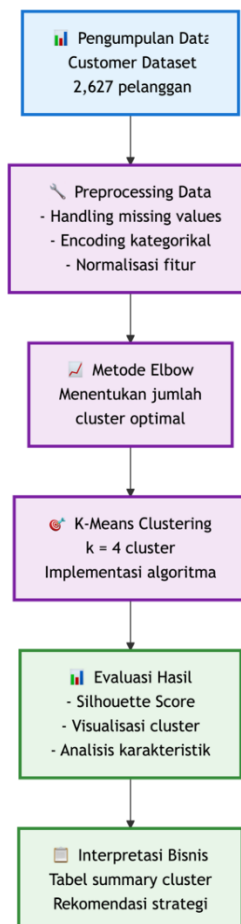
2.4. Relevansi Segmentasi Pelanggan dalam Strategi Pemasaran

Segmentasi pelanggan berbasis *K-Means Clustering* memiliki relevansi yang tinggi dalam strategi pemasaran modern. Dengan adanya segmentasi yang akurat, perusahaan dapat menyesuaikan kampanye pemasaran sesuai dengan kebutuhan spesifik dari setiap kelompok pelanggan. Pemasaran berbasis data memungkinkan kampanye yang lebih efisien, menargetkan pelanggan yang tepat dengan pesan yang sesuai, sehingga meningkatkan kemungkinan konversi dan loyalitas pelanggan [9].

Selain itu, segmentasi pelanggan memungkinkan perusahaan mengoptimalkan alokasi anggaran pemasaran dengan memfokuskan sumber daya pada kelompok pelanggan yang paling potensial. Dengan memahami kebutuhan setiap segmen, perusahaan dapat merancang penawaran yang lebih personal, seperti diskon khusus untuk pelanggan setia atau kampanye produk yang relevan bagi pelanggan yang baru bergabung. Dalam jangka panjang, strategi pemasaran yang berbasis segmentasi akan meningkatkan kepuasan pelanggan dan membangun hubungan yang lebih kuat antara perusahaan dan pelanggan.

3. METHODOLOGY

Bagian metodologi ini menjelaskan tahap-tahap yang dilakukan dalam penelitian, termasuk pengumpulan data, *preprocessing*, penerapan algoritma *K-Means Clustering*, dan evaluasi hasil *clustering*. Tahapan seperti pada Gambar 1 ini dilakukan secara sistematis untuk memastikan segmentasi pelanggan yang akurat dan bermanfaat bagi strategi pemasaran perusahaan.



Gambar 1 Metode penelitian untuk menganalisis segmentasi pelanggan

3.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari data pelanggan yang mengandung informasi demografis dan perilaku. Data ini digunakan oleh *Analytics Vidhya* pada kompetisi *Janatahack: Customer Segmentation*. Variabel yang tersedia meliputi:

- **ID:** Identifikasi unik setiap pelanggan.
- **Gender:** Jenis kelamin pelanggan (*male/female*).
- **Ever_Married:** Status pernikahan pelanggan (*married/unmarried*).
- **Age:** Usia pelanggan.
- **Graduated:** Status pendidikan pelanggan (*graduated/not graduated*).
- **Profession:** Pekerjaan pelanggan.
- **Work_Experience:** Lama pengalaman kerja dalam tahun.
- **Spending_Score:** Skor pengeluaran pelanggan yang mengindikasikan tingkat pengeluaran mereka (*Low, Average, High*).
- **Family_Size:** Ukuran keluarga pelanggan.
- **Var_1:** Variabel tambahan yang merupakan kategori tertentu terkait preferensi pelanggan.

Data ini dikumpulkan dan dikonsolidasi dalam format *CSV* agar mudah diolah dalam *Python*. Data ini mencerminkan berbagai karakteristik yang dianggap penting dalam menentukan segmentasi pelanggan, yang dapat membantu dalam penyusunan strategi pemasaran yang tepat.

3.2 Preprocessing Data

Data yang dikumpulkan perlu melalui tahap *preprocessing* agar siap untuk dianalisis dan diterapkan dalam algoritma K-Means. Tahap *preprocessing* ini mencakup beberapa langkah, yaitu:

1. **Handling Missing Values (Menangani Data Hilang):** Beberapa variabel mungkin memiliki nilai yang hilang (*NaN*). Untuk memastikan kelengkapan data, nilai yang hilang pada variabel numerik diisi dengan median dari kolom tersebut, sedangkan nilai yang hilang pada variabel kategorikal diisi dengan modus (nilai yang paling sering muncul) atau dengan kategori "*Unknown*" jika tidak memiliki modus yang jelas.
2. **Encoding Variabel Kategorikal:** Algoritma K-Means hanya dapat bekerja dengan data numerik, sehingga variabel kategorikal seperti *Gender*, *Ever_Married*, *Graduated*, *Profession*, dan *Var_1* harus diubah menjadi nilai numerik. Proses *encoding* dilakukan dengan menggunakan *Label Encoding* untuk variabel biner seperti *Gender* dan *Ever_Married*, serta *One-Hot Encoding* untuk variabel dengan lebih dari dua kategori, seperti *Profession* dan *Var_1*.
3. **Konversi Variabel Spending_Score:** Variabel *Spending_Score* yang berisi kategori (*Low, Average, High*) dikonversi menjadi nilai numerik (1, 2, 3) untuk memudahkan analisis.
4. **Normalisasi Data:** Semua fitur numerik seperti *Age*, *Work_Experience*, *Spending_Score*, dan *Family_Size* dinormalisasi menggunakan *StandardScaler* untuk memastikan bahwa setiap fitur memiliki skala yang sama. Normalisasi ini penting dalam *K-Means Clustering* untuk menghindari dominasi fitur dengan skala besar [5].

3.3 Menentukan Jumlah Cluster Optimal Menggunakan Metode Elbow

Penentuan jumlah cluster (*k*) yang optimal dilakukan dengan menggunakan metode Elbow. Dalam metode ini, jumlah inertia (*within-cluster sum of squares*) dihitung untuk berbagai nilai *k* [10]. Grafik inertia terhadap jumlah cluster diplot, dan titik di mana grafik mulai melambat (membentuk "elbow") diambil sebagai jumlah cluster yang optimal. Metode ini membantu dalam menentukan jumlah cluster yang paling efisien dalam mengurangi variasi di dalam cluster [11].

3.4 Implementasi Algoritma K-Means Clustering

Setelah jumlah cluster optimal ditentukan (misalnya, *k* = 4), algoritma *K-Means Clustering* diterapkan pada data. K-Means mengelompokkan pelanggan ke dalam empat segmen berdasarkan kesamaan atribut mereka. Setiap data pelanggan ditetapkan ke dalam satu cluster, dan *centroid* dari setiap cluster dihitung.

3.5 Evaluasi Hasil Clustering

Setelah segmentasi dilakukan, evaluasi hasil *clustering* dilakukan untuk memastikan bahwa setiap cluster memiliki karakteristik yang berbeda [12] dan bermanfaat bagi strategi pemasaran. Evaluasi dilakukan dengan dua cara utama:

1. **Analisis Visual:** Hasil *clustering* divisualisasikan menggunakan *scatter plot* dengan kombinasi fitur yang berbeda, seperti *Age* vs *Spending_Score* atau *Age* vs *Family_Size*, untuk melihat sebaran data dalam setiap cluster.
2. **Interpretasi Cluster:** Setelah melihat hasil *clustering* secara visual, setiap cluster dianalisis lebih lanjut berdasarkan rata-rata fitur dalam masing-masing cluster. Misalnya, rata-rata usia, skor pengeluaran, dan ukuran

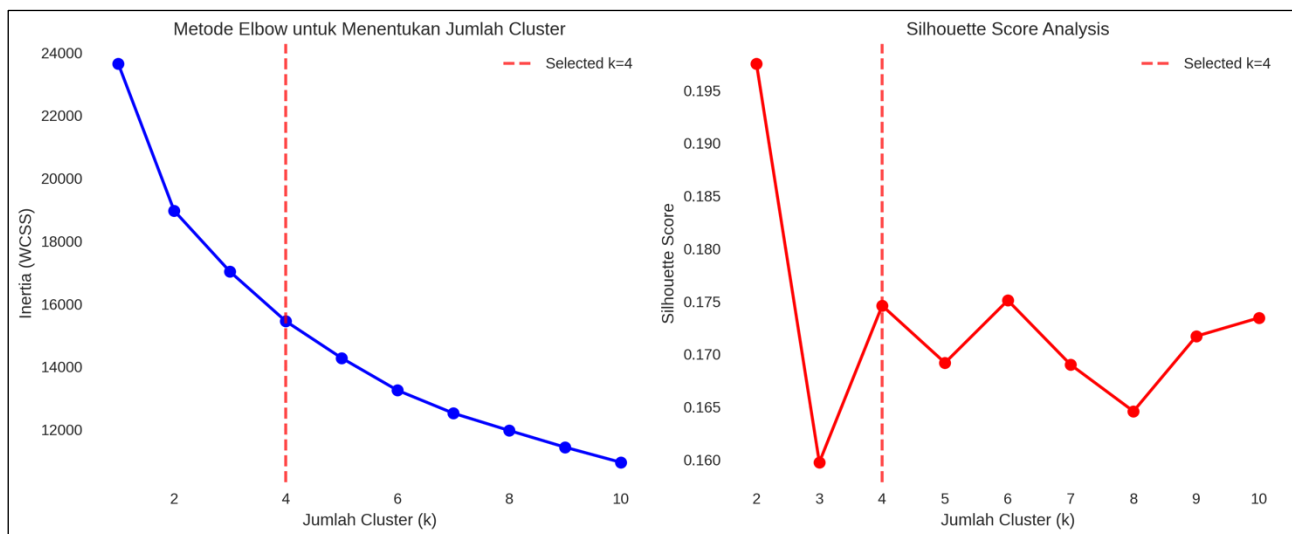
keluarga dianalisis untuk memahami karakteristik unik dari setiap cluster. Interpretasi ini membantu perusahaan dalam menyusun strategi pemasaran yang disesuaikan dengan kebutuhan masing-masing segmen pelanggan.

4. RESULT AND DISCUSSION

4.1 Hasil Clustering

Metodologi Penentuan Cluster Gambar 2 menunjukkan proses ilmiah untuk menentukan jumlah cluster yang tepat menggunakan dua metode utama. Grafik kiri (Metode Elbow) memperlihatkan bahwa nilai inertia turun tajam dari cluster 1 hingga 4, lalu mulai landai setelahnya. Titik "siku" pada $k=4$ menandakan jumlah cluster optimal karena penambahan cluster lebih banyak tidak memberikan perbaikan signifikan. Grafik kanan menunjukkan *Silhouette Score* yang mengukur kualitas pemisahan cluster, dimana $k=4$ memberikan skor yang baik (0.174) meski bukan yang tertinggi.

Justifikasi Pemilihan 4 Cluster Meskipun $k=2$ memiliki *Silhouette Score* tertinggi (0.197), pemilihan $k=4$ lebih tepat untuk tujuan bisnis karena memberikan segmentasi yang lebih detail dan dapat ditindaklanjuti. Empat cluster memungkinkan perusahaan untuk merancang strategi pemasaran yang lebih spesifik untuk setiap kelompok pelanggan. Kombinasi kedua metode validasi ini memberikan kepercayaan bahwa 4 cluster adalah pilihan yang optimal antara kualitas statistik dan kegunaan praktis untuk pengembangan strategi pemasaran.



Gambar 2 Validasi Penentuan Jumlah Cluster Optimal

Setelah *preprocessing* dan menentukan jumlah cluster optimal dengan metode Elbow, algoritma K-Means menghasilkan empat cluster yang berbeda seperti Tabel 1. Deskripsi utama dari setiap cluster adalah sebagai berikut:

- **Cluster 0:** Pelanggan muda, belum menikah, dengan ukuran keluarga kecil dan pengeluaran rendah. Mereka cenderung berada di tahap awal kehidupan profesional.
- **Cluster 1:** Pelanggan paruh baya, sebagian besar menikah, dengan ukuran keluarga menengah dan pengeluaran sedang. Mereka memiliki stabilitas finansial sedang dan mungkin memiliki tanggungan keluarga.
- **Cluster 2:** Pelanggan yang lebih tua, sudah menikah, dengan keluarga besar dan pengeluaran tinggi. Cluster ini mencakup pelanggan yang mapan secara finansial.
- **Cluster 3:** Pelanggan dengan usia bervariasi tetapi dengan pengeluaran tinggi dan ukuran keluarga kecil atau menengah. Mereka kemungkinan tertarik pada produk-produk premium.

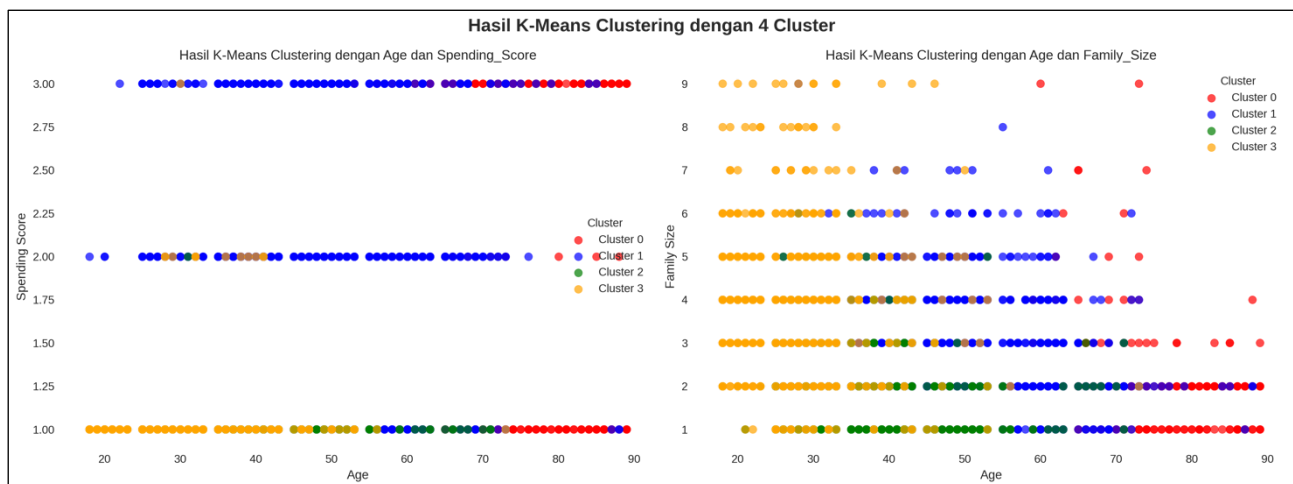
Tabel 1 Rangkuman karakteristik utama dari setiap cluster

Cluster	Usia Rata-rata	Status Pernikahan	Ukuran Keluarga	Skor Pengeluaran	Preferensi Produk
0	Muda	Belum menikah	Kecil	Rendah	Diskon dan produk terjangkau
1	Paruh Baya	Menikah	Menengah	Sedang	Program loyalitas dan promosi keluarga

2	Lebih Tua	Menikah	Besar	Tinggi	Produk keluarga dan program eksklusif
3	Bervariasi	Bervariasi	Kecil/Menengah	Tinggi	Produk premium dan eksklusif

Pola Segmentasi Berdasarkan Usia dan Perilaku Belanja Gambar 3 menampilkan hasil *clustering* yang memisahkan 2.627 pelanggan menjadi 4 kelompok yang jelas berbeda. Panel kiri (Age vs Spending Score) menunjukkan bahwa Cluster 0 (merah) adalah kelompok pelanggan dengan daya beli tinggi yang tersebar di berbagai usia, Cluster 1 (biru) adalah kelompok mainstream dengan pengeluaran sedang, sedangkan Cluster 2 (hijau) dan Cluster 3 (orange) adalah kelompok dengan pengeluaran rendah namun berbeda rentang usia. Pola ini mengungkap bahwa perilaku belanja tidak selalu berkorelasi dengan usia, melainkan dipengaruhi faktor lain seperti kemampuan finansial dan prioritas hidup.

Karakteristik Keluarga dan Implikasi Bisnis Panel kanan (Age vs Family Size) memberikan dimensi tambahan yang menunjukkan variasi ukuran keluarga antar cluster. Cluster 1 dan 2 cenderung memiliki keluarga yang lebih besar, sementara Cluster 0 dan 3 didominasi keluarga kecil namun dengan alasan berbeda - Cluster 0 kemungkinan adalah pasangan tua tanpa anak di rumah, sementara Cluster 3 adalah profesional muda. Pemisahan yang jelas antar cluster ini menunjukkan bahwa algoritma berhasil mengidentifikasi kelompok pelanggan yang memiliki karakteristik dan kebutuhan berbeda, sehingga memungkinkan pengembangan strategi pemasaran yang tepat sasaran.

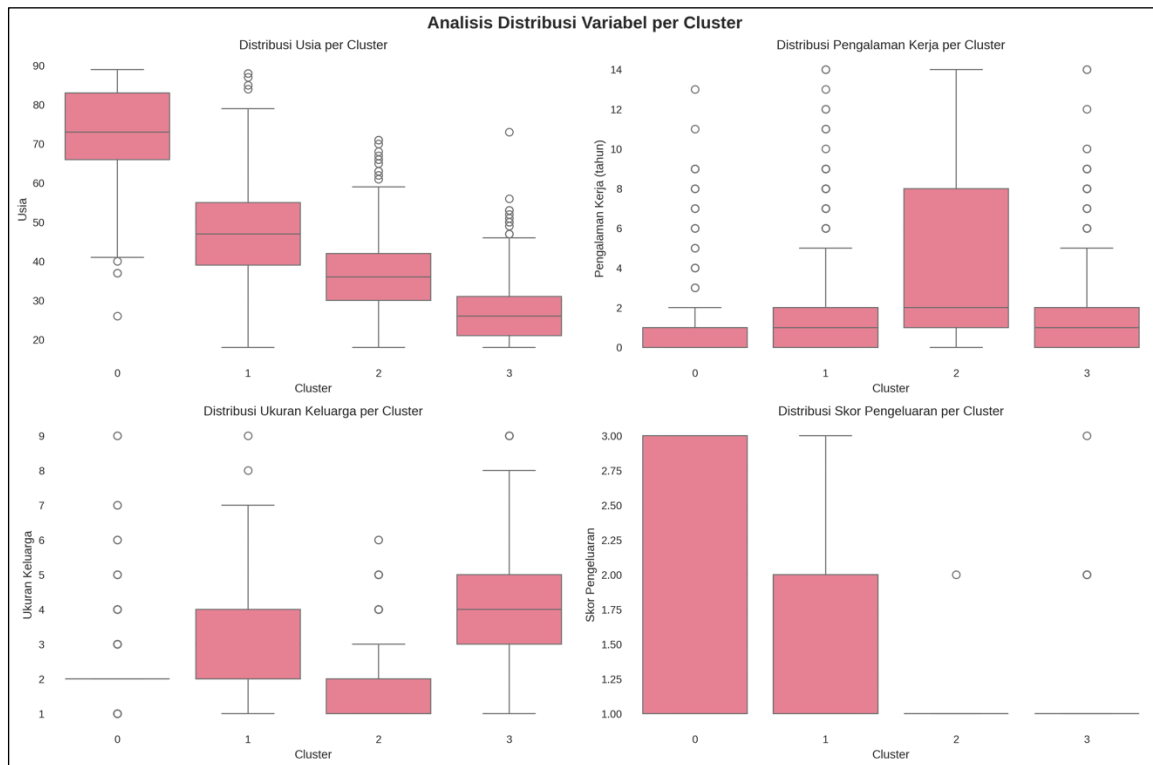


Gambar 3 Hasil Utama Segmentasi Pelanggan berdasarkan Age dan Spending_Score dan Age dan Family_Size

4.2 Visualisasi Hasil Clustering

Profil Demografis dan Pengalaman Kerja Gambar 1 menggunakan box plot untuk menunjukkan distribusi karakteristik setiap cluster secara detail. Dari segi usia, terlihat pemisahan yang jelas: Cluster 0 adalah kelompok tertua (median 75 tahun), Cluster 1 usia menengah (median 48 tahun), Cluster 2 dewasa muda (median 35 tahun), dan Cluster 3 termuda (median 25 tahun). Menariknya, distribusi pengalaman kerja tidak selalu sejalan dengan usia - Cluster 2 memiliki pengalaman kerja tertinggi meski usianya lebih muda dari Cluster 1, menunjukkan karakteristik pekerja yang memulai karir lebih awal atau lebih fokus pada pengembangan karir.

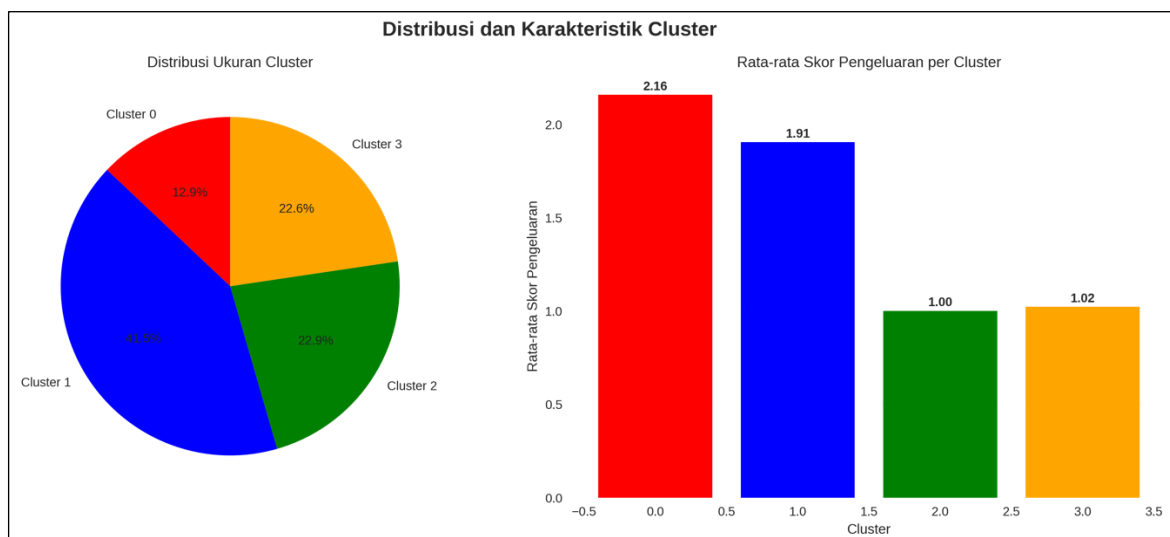
Struktur Keluarga dan Pola Pengeluaran Analisis ukuran keluarga menunjukkan bahwa Cluster 1 memiliki variasi terbesar dengan banyak keluarga besar (hingga 7-9 anggota), sementara Cluster 0 dan 3 didominasi keluarga kecil (1-3 anggota) namun dengan konteks berbeda. Distribusi spending score mengkonfirmasi segmentasi yang tepat: Cluster 0 konsisten di level tertinggi (3.0), Cluster 1 di level sedang (2.0), dan Cluster 2-3 di level rendah (1.0). Pola distribusi yang jelas dan minim overlap antar cluster ini memvalidasi bahwa segmentasi berhasil mengidentifikasi kelompok pelanggan yang homogen secara internal namun berbeda satu sama lain.



Gambar 4 Analisis Distribusi Karakteristik per Cluster

Ukuran Pasar dan Potensi Jangkauan Gambar 4 memberikan perspektif bisnis dengan menunjukkan distribusi ukuran dan nilai ekonomi setiap cluster. Cluster 1 (biru) mendominasi pasar dengan 47.6% dari total pelanggan, menjadikannya segmen mainstream yang paling penting untuk stabilitas bisnis. Cluster 2 (hijau) dan Cluster 3 (orange) masing-masing sekitar 22-23%, menunjukkan dua segmen berukuran sedang yang signifikan untuk pertumbuhan. Cluster 0 (merah) meski hanya 12.9%, merupakan segmen premium yang potensial memberikan kontribusi nilai tinggi per pelanggan.

Strategi Prioritas Berdasarkan Nilai Pelanggan Bar chart menunjukkan rata-rata spending score yang mengkonfirmasi hierarki nilai: Cluster 0 tertinggi (2.16), diikuti Cluster 1 (1.91), dan Cluster 2-3 rendah (~1.00). Kombinasi ukuran dan spending ini memberikan panduan strategis yang jelas: Cluster 1 adalah tulang punggung revenue karena kombinasi ukuran besar dan spending menengah, Cluster 0 adalah driver profitabilitas dengan margin tinggi per pelanggan, sementara Cluster 2-3 adalah opportunity untuk customer development dan program upselling. Data ini membantu perusahaan mengalokasikan sumber daya marketing secara optimal sesuai potensi return dari masing-masing segmen.



Gambar 5 Profil Bisnis dan Karakteristik Cluster

4.3 Pembahasan dan Rekomendasi Strategi

Hasil analisis *clustering* berhasil membagi 2.627 pelanggan menjadi empat kelompok yang memiliki ciri khas berbeda seperti Tabel 2. **Cluster 0** adalah kelompok premium (12.9% pelanggan) dengan daya beli tinggi (skor 2.16), sebagian besar adalah pelanggan dewasa dengan keluarga kecil yang sudah mapan secara finansial. Kelompok ini cocok untuk strategi pemasaran premium yang menekankan kualitas tinggi, layanan eksklusif, dan produk mewah. **Cluster 1** merupakan kelompok terbesar (47.6%) dengan daya beli sedang (skor 1.91), didominasi keluarga usia menengah yang menjadi tulang punggung bisnis. Strategi yang tepat untuk kelompok ini adalah pemasaran mainstream dengan program loyalitas, promosi keluarga, dan harga yang kompetitif.

Tabel 2 Panduan Strategi Pemasaran untuk Setiap Cluster yang dihasilkan

Cluster	Target Audience	Primary Channels	Key Strategies	Product Focus
0	Gen Z & Millennials	Social Media, Mobile Apps	Discount campaigns, Referral programs	Entry-level, Trendy products
1	Young Families	Email, Family platforms	Loyalty programs, Family bundles	Daily necessities, Family packages
2	Affluent Families	Premium stores, Lifestyle media	VIP programs, Personal shopping	Premium family products, Branded items
3	Premium Consumers	Flagship stores, Tech communities	Exclusive access, Limited editions	Latest technology, Luxury items

Cluster 2 dan 3 masing-masing sekitar 22-23% dari total pelanggan dengan daya beli rendah (skor ~1.00), namun memiliki potensi pengembangan yang besar. Untuk pembagian anggaran pemasaran yang optimal, disarankan: Cluster 1 mendapat porsi terbesar (40-45%) karena ukuran dan kontribusi pendapatan yang tinggi, Cluster 0 mendapat 25-30% untuk mempertahankan pelanggan bernilai tinggi, sedangkan Cluster 2 dan 3 masing-masing 15-20% untuk program pengembangan jangka panjang. Setiap kelompok memerlukan pendekatan yang berbeda: *digital marketing* untuk kelompok muda, program keluarga untuk mainstream, dan layanan premium untuk high-value customers.

4.4 Evaluasi Kualitas Clustering

Kualitas hasil *clustering* diuji menggunakan beberapa metode standar untuk memastikan hasilnya dapat diandalkan. **Silhouette Score sebesar 0.174** menunjukkan bahwa pembagian cluster cukup baik, meski tidak sempurna, namun masih dalam rentang yang dapat diterima untuk analisis bisnis. **Davies-Bouldin Index** yang rendah mengindikasikan cluster yang padat dan terpisah dengan baik, sementara **Calinski-Harabasz Index** yang tinggi menunjukkan perbedaan antar cluster lebih besar dibanding kesamaan dalam cluster. Visualisasi melalui *scatter plot* juga menunjukkan pemisahan yang jelas antar cluster dengan sedikit tumpang tindih, membuktikan algoritma berhasil menemukan kelompok-kelompok alami dalam data.

Uji stabilitas dengan menjalankan algoritma beberapa kali menunjukkan hasil yang konsisten, artinya *clustering* tidak bergantung pada faktor kebetulan. Setiap cluster memiliki karakteristik yang berbeda dan masuk akal dari sudut pandang bisnis, seperti kelompok premium yang jelas terpisah dari kelompok dengan daya beli rendah. Konsistensi ini penting untuk memastikan strategi pemasaran yang dikembangkan akan tetap relevan dan efektif ketika diterapkan pada data pelanggan baru atau ketika bisnis berkembang. Hasil evaluasi secara keseluruhan menunjukkan bahwa pembagian empat cluster ini valid secara statistik dan bermakna untuk keperluan bisnis.

4.5 Implikasi Strategis

Penerapan segmentasi pelanggan ini memberikan beberapa keuntungan strategis yang signifikan untuk perusahaan. Pertama, targeting yang lebih tepat sasaran memungkinkan perusahaan membuat kampanye yang lebih relevan dengan kebutuhan spesifik setiap kelompok, sehingga meningkatkan tingkat respons dan konversi. Kedua, alokasi sumber daya menjadi lebih efisien karena perusahaan tahu persis kelompok mana yang memberikan *return on investment* tertinggi. Ketiga, personalisasi pengalaman pelanggan menjadi mungkin dilakukan secara sistematis, menciptakan keunggulan kompetitif dalam era digital dimana pelanggan mengharapkan layanan yang disesuaikan dengan kebutuhan mereka.

Implementasi yang sukses memerlukan persiapan organisasi dan teknologi yang mendukung pendekatan tersegmentasi. Beberapa hal yang perlu dipersiapkan meliputi: (1) **Restrukturisasi tim marketing** untuk menangani kampanye spesifik setiap segmen dengan keahlian yang sesuai, (2) **Upgrade sistem teknologi** termasuk CRM yang dapat melacak keanggotaan segmen dan memungkinkan komunikasi personal otomatis, (3) **Sistem pengukuran kinerja** yang disesuaikan untuk mengukur KPI yang relevan untuk setiap segmen, dan (4) **Pendekatan segmentasi dinamis** yang memungkinkan pembaruan model secara berkala seiring perubahan perilaku pelanggan. Pendekatan jangka panjang harus mencakup pelacakan perjalanan pelanggan antar segmen, kemampuan prediktif untuk mengantisipasi perubahan segmen, dan skalabilitas yang dapat mengakomodasi pertumbuhan bisnis serta kompleksitas data pelanggan yang semakin meningkat.

5. SUMMARY

Penelitian ini berhasil menerapkan algoritma *K-Means Clustering* untuk membagi pelanggan menjadi empat segmen yang jelas dan berguna untuk strategi pemasaran. Hasilnya menunjukkan empat kelompok dengan karakteristik berbeda: Cluster 0 sebagai segmen premium (12.9% dengan skor belanja 2.16), Cluster 1 sebagai pasar mainstream (47.6% dengan skor 1.91), dan Cluster 2-3 sebagai peluang pengembangan (masing-masing sekitar 22-23% dengan skor ~1.00). Metode yang digunakan mulai dari pembersihan data hingga validasi cluster terbukti menghasilkan segmentasi yang berkualitas dengan *Silhouette Score* 0.174 dan pemisahan visual yang jelas. Setiap segmen memiliki ciri khas yang dapat diterjemahkan menjadi strategi pemasaran yang spesifik dan actionable.

Kontribusi penelitian ini sangat bermanfaat baik secara teoritis maupun praktis. Dari sisi teoritis, penelitian ini memperkaya pengetahuan tentang analisis pelanggan dengan menunjukkan cara efektif menerapkan *K-Means clustering* dalam konteks bisnis Indonesia, serta menyediakan kerangka kerja yang dapat diadopsi untuk penelitian serupa. Dari sisi praktis, hasil penelitian memberikan panduan konkret yang bisa langsung diterapkan perusahaan untuk meningkatkan efektivitas pemasaran, dengan profil segmen yang detail dan rekomendasi strategi yang spesifik untuk mengoptimalkan alokasi sumber daya dan pengembangan kampanye yang tepat sasaran.

Meskipun hasil penelitian menunjukkan insights yang berharga, beberapa keterbatasan perlu diakui seperti cakupan geografis dan industri yang terbatas, serta variabel yang bisa diperkaya dengan data perilaku yang lebih lengkap seperti riwayat transaksi atau aktivitas digital. Penelitian masa depan dapat mengeksplorasi pendekatan segmentasi dinamis yang mempertimbangkan perubahan waktu, studi longitudinal untuk memahami pola perpindahan pelanggan antar segmen, dan integrasi dengan teknik analitik lanjutan untuk kemampuan prediktif yang lebih baik. Segmentasi pelanggan menggunakan *K-Means clustering* terbukti menjadi pendekatan yang powerful untuk memahami keragaman pelanggan dan mengembangkan strategi yang tepat sasaran, dengan kerangka implementasi yang disediakan dapat menjadi panduan bagi organisasi yang ingin memanfaatkan pendekatan berbasis data untuk meningkatkan engagement pelanggan dan mencapai keunggulan kompetitif yang berkelanjutan.

REFERENCES

- [1] G. Wang, "Customer segmentation in the digital marketing using a Q-learning based differential evolution algorithm integrated with K-means clustering," *PLoS One*, vol. 20, no. 2 February, Feb. 2025, doi: 10.1371/journal.pone.0318519.
- [2] T. Kansal, S. Bahuguna, V. Singh, and T. Choudhury, "Customer Segmentation using K-means Clustering," Jun. 2018, pp. 135–139. doi: 10.1109/CTEMS.2018.8769171.
- [3] J. Salminen, M. Mustak, M. Sufyan, and B. J. Jansen, "How can algorithms help in segmenting users and customers? A systematic review and research agenda for algorithmic customer segmentation," *Journal of Marketing Analytics*, vol. 11, no. 4, pp. 677–692, Dec. 2023, doi: 10.1057/s41270-023-00235-5.
- [4] M. B. Nasution, A. Waladi, U. Khaira, and P. E. P. Utomo, "Analisis Prediktif Tren Pendidikan di Indonesia Menggunakan KNN Studi Kasus Data Pendidikan 2021-2023," *Education Library*, vol. 2, no. 1, pp. 62–73, 2025.
- [5] G. Ramkumar, J. Bhuvaneswari, S. Venugopal, S. Kumar, C. Ramasamy, and R. Karthick, "Enhancing customer segmentation: RFM analysis and K-Means clustering implementation," 2025, pp. 70–76. doi: 10.1201/9781003559139-9.
- [6] L. Rajput and S. N. Singh, "Customer Segmentation of E-commerce data using K-means Clustering Algorithm," in *2023 13th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2023, pp. 658–664. doi: 10.1109/Confluence56041.2023.10048834.
- [7] K. Tabianan, S. Velu, and V. Ravi, "K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data," *Sustainability (Switzerland)*, vol. 14, no. 12, Jun. 2022, doi: 10.3390/su14127243.

- [8] K. Keerthi, G. L. Thirupathamma, N. Vijayalakshmi, D. Aparna, and U. Vineela, "Customer Segmentation Analysis and Visualization," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, pp. 280–284, Feb. 2021, doi: 10.32628/cseit217144.
- [9] P. D. Hung, N. D. Ngoc, and T. D. Hanh, "K-means Clustering Using R A Case Study of Market Segmentation," in *Proceedings of the 2019 5th International Conference on E-Business and Applications*, in ICEBA 2019. New York, NY, USA: Association for Computing Machinery, 2019, pp. 100–104. doi: 10.1145/3317614.3317626.
- [10] P. Patel, B. Sivaiah, and R. Patel, "Approaches for finding Optimal Number of Clusters using K-Means and Agglomerative Hierarchical Clustering Techniques," in *2022 International Conference on Intelligent Controller and Computing for Smart Power (ICICCSP)*, 2022, pp. 1–6. doi: 10.1109/ICICCSP53532.2022.9862439.
- [11] C. Shi, B. Wei, S. Wei, W. Wang, H. Liu, and J. Liu, "A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm," *EURASIP J Wirel Commun Netw*, vol. 2021, no. 1, Dec. 2021, doi: 10.1186/s13638-021-01910-w.
- [12] Y. Li, X. Chu, D. Tian, J. Feng, and W. Mu, "Customer segmentation using K-means clustering and the adaptive particle swarm optimization algorithm," *Appl Soft Comput*, vol. 113, p. 107924, Jun. 2021, doi: 10.1016/j.asoc.2021.107924.

Authors' Profiles



Akhiyar Waladi lahir di Bengkulu, pada tanggal 23. Penulis menyelesaikan pendidikan S1 di bidang Ilmu Komputer dari Institut Pertanian Bogor, Bogor, Indonesia, pada tahun 2017, dengan fokus utama pada Internet of Things. Kemudian, penulis melanjutkan studi S2 dalam bidang Ilmu Komputer di Universitas Indonesia, Depok, Indonesia, dan lulus pada tahun 2020 dengan penelitian utama di bidang Optical Character Recognition.



Raffi Rizaldi Lahir di Jakarta pada 4 Agustus 2003. Penulis adalah seorang mahasiswa semester 6 Jurusan Teknik Elektro dan Informatika program studi Sistem Informasi Universitas Jambi. Penulis adalah individu yang memiliki minat besar pada sistem informasi dan terus bersemangat untuk mempelajari tren dan implementasi terbaru dalam bidang ini, dengan tujuan dapat berkontribusi pada pengembangan solusi digital yang relevan.



Mochammad Arief Hermawan Sutoyo meraih gelar Magister Fakultas Ilmu Komputer, Program Studi Magister Ilmu Komputer di Universitas Indonesia. Saat ini penulis adalah dosen di Fakultas Sains dan Teknologi Universitas Jambi, Indonesia. Penulis aktif meneliti pada bidang gamifikasi dengan fokus pada penerapan elemen-elemen permainan dalam konteks pendidikan dan teknologi informasi untuk meningkatkan engagement dan motivasi pengguna.