

Segmentasi Pelanggan Menggunakan K-Means Clustering Berdasarkan Data Kepribadian dan Pola Konsumsi

Iqbal^{1*}, Nurul Hidayat², Daiva Paundra Gevano³, Andhika Putra Restu Iahi⁴

¹Informatika, Universitas Almuslim, Indonesia

^{2,3,4}Informatika, Universitas Jenderal Soedirman, Indonesia

Email: iqbal@umuslim.ac.id

Received : Jul 19, 2025; Revised : Oct 21, 2025; Accepted : Oct 22, 2025; Published : Oct 23, 2025

Abstract

In today's competitive business landscape, a deep understanding of customer behavior and preferences is crucial for strategic success. Customer segmentation emerges as a vital approach to identify distinct customer subgroups, enabling personalized and efficient marketing strategies. However, many companies still struggle to achieve this understanding due to suboptimal data utilization and inaccurate manual grouping methods. To address these challenges, this research proposes and implements a data mining approach using the K-Means Clustering algorithm for automated and measurable customer segmentation. Leveraging the "Customer Personality Analysis" dataset from Kaggle, this study aims to uncover hidden patterns in customer demographics (age, income, marital status, number of children) and purchasing behavior (number and frequency of transactions). A comprehensive data pre-processing pipeline, including handling missing values, feature engineering, irrelevant column removal, categorical transformation, and numerical scaling, ensures data quality and readiness. Using the Elbow Method, four optimal clusters were identified: "Balanced Spenders with Teenagers" (Cluster 0), "Budget-Conscious Families" (Cluster 1), "High-Value Engaged Buyers" (Cluster 2), and "Active Mature Buyers" (Cluster 3). Visualization using Principal Component Analysis (PCA) further confirms significant characteristic differences between these segments. Cluster 2, being the most valuable and responsive segment, requires premium marketing strategies, while Cluster 1, the largest segment, demands a value-oriented approach. The results of this segmentation provide deep strategic insights, enabling companies to allocate marketing resources more efficiently, craft more relevant messages, and ultimately enhance customer satisfaction and business profitability. These findings demonstrate the potential of unsupervised learning in enhancing data-driven customer profiling systems in marketing and business informatics.

Keywords : *Customer Behavior, Customer Segmentation, Data Mining, K-Means Clustering, Personalized Marketing.*

This work is an open access article and licensed under a Creative Commons Attribution-Non Commercial 4.0 International License



1. PENDAHULUAN

Dalam dunia bisnis yang semakin sengit saat ini, pengetahuan yang mendalam mengenai perilaku dan preferensi pelanggan menjadi faktor penting untuk mencapai keberhasilan strategis. Perusahaan senantiasa mencari metode untuk mengenali segmen pelanggan yang berbeda sehingga dapat menyusun strategi pemasaran yang lebih personal dan efisien[1]. Tanpa pemahaman yang tepat tentang basis pelanggan, usaha pemasaran bisa terlalu umum dan tidak efisien, menghasilkan *Return on Investment* (ROI) yang kurang optimal. Sebagai akibatnya, segmentasi pelanggan muncul sebagai pendekatan yang sangat krusial untuk mencapai tujuan ini[2][3]. Pendekatan ini memungkinkan perusahaan untuk mengelompokkan konsumen ke dalam kategori-kategori homogen berdasarkan karakteristik, perilaku pembelian, dan preferensi mereka yang unik[4], [5].

Segmentasi pelanggan merupakan proses memecah basis pelanggan yang besar dan beragam menjadi sub kelompok yang lebih kecil dan seragam berdasarkan karakteristik tertentu[6]. Sasaran utamanya adalah untuk mengenali segmen pelanggan dengan kebutuhan, keinginan, dan perilaku yang mirip, agar perusahaan dapat menjangkau mereka dengan pesan dan tawaran yang relevan. Metode ini memungkinkan distribusi sumber daya yang lebih efektif, peningkatan kepuasan konsumen, dan akhirnya, peningkatan laba[7], [8], [9].

Namun, banyak perusahaan masih mengalami kesulitan dalam memahami karakteristik dan kebiasaan pelanggan secara mendalam karena penggunaan data yang belum optimal[10], [11], [12]. Selama ini, pengelompokan pelanggan berdasarkan faktor-faktor seperti usia, pendapatan, status pernikahan, jumlah anak, dan kebiasaan belanja biasanya dilakukan secara manual atau dengan metode konvensional, sehingga hasilnya kurang akurat. Minimnya data terstruktur tentang profil pelanggan juga menyulitkan perusahaan dalam menyusun strategi pemasaran yang tepat. Penelitian sebelumnya menunjukkan bahwa segmentasi pelanggan yang akurat dapat meningkatkan efektivitas strategi pemasaran dan profitabilitas perusahaan[13].

Untuk mengatasi masalah ini, diperlukan pendekatan berbasis data mining, seperti K-Means Clustering, yang dapat melakukan segmentasi pelanggan secara otomatis, dapat diukur, dan didukung oleh analisis data yang kuat[14], [15]. Penelitian ini bertujuan untuk melaksanakan segmentasi pelanggan pada dataset kampanye pemasaran yang tersedia, dengan memanfaatkan teknik penambangan data. Penelitian ini akan menggunakan algoritma K-Means clustering untuk mengenali berbagai segmen pelanggan. K-Means merupakan algoritma pemisahan non-hierarkis yang sering digunakan karena kesederhanaan dan efisiensinya dalam mengelompokkan data ke dalam sejumlah kluster yang sudah ditentukan sebelumnya[6].

Metode berbasis kecerdasan buatan seperti K-Means clustering memiliki potensi dalam mengungkap pola tersembunyi dalam perilaku pelanggan, serta menawarkan alternatif yang kuat dibandingkan pendekatan segmentasi tradisional[16], [17], [18], [19]. Melalui penerapan K-Means, kami berharap bisa mengidentifikasi pola-pola tersembunyi dalam perilaku belanja dan demografi pelanggan, yang selanjutnya dapat dianalisis untuk merumuskan profil segmen yang relevan[14].

Penelitian ini bertujuan untuk melakukan segmentasi pelanggan berdasarkan data demografis, perilaku pembelian, dan respons terhadap kampanye pemasaran dengan menggunakan algoritma K-Means Clustering. Secara khusus, penelitian ini bertujuan untuk mengelompokkan pelanggan ke dalam beberapa segmen berdasarkan atribut-atribut seperti usia, pendapatan, status pernikahan, jumlah anak, serta kebiasaan belanja guna memahami karakteristik utama dari setiap segmen. Selain itu, penerapan teknik clustering berbasis data mining diharapkan mampu menghasilkan segmentasi yang lebih akurat dan terukur dibandingkan dengan pendekatan konvensional, serta menyediakan informasi yang berguna bagi perusahaan untuk merancang strategi pemasaran yang lebih tepat sasaran, efektif, dan efisien[20].

Ruang lingkup dari penelitian ini dibatasi hanya pada proses segmentasi pelanggan menggunakan metode K-Means Clustering terhadap dataset *Customer Personality Analysis* yang diambil dari situs Kaggle[17]. Penelitian ini berfokus pada karakteristik demografis seperti usia, status pernikahan, dan tingkat pendapatan serta perilaku pembelian pelanggan seperti jumlah produk yang dibeli dan frekuensi transaksi. Penelitian ini tidak mencakup prediksi tindakan pelanggan di masa depan, evaluasi efektivitas strategi pemasaran, maupun penambahan data eksternal di luar dataset yang digunakan.

2. METODOLOGI PENELITIAN

2.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari sebuah kumpulan data yang lengkap berjudul "*Customer Personality Analysis*" yang disediakan dalam format CSV. Dataset ini diperoleh dari sebuah kampanye pemasaran, yang mencatat berbagai data demografis, perilaku pembelian, dan

reaksi pelanggan terhadap kampanye pemasaran. Dataset ini digunakan sebagai landasan untuk analisis segmentasi pelanggan kami. Dataset diambil dari website bernama Kaggle. Dataset ini dipilih karena kecukupan dan relevansinya dalam konteks analisis perilaku pelanggan serta kampanye pemasaran.

Dataset asli terdiri dari 2240 baris dan 29 kolom. Setiap baris menggambarkan satu pelanggan yang berbeda, sedangkan setiap kolom memuat informasi khusus terkait pelanggan itu. Atribut-atribut dalam kumpulan data dapat dikelompokkan Tabel 1.

Tabel 1. Atribut-atribut yang digunakan

Demografi	- ID	
Pelanggan	- Year_Birth (tahun kelahiran)	
	- Education (gelar pendidikan)	
	- Marital_Status (status pernikahan)	
	- Income (penghasilan tahunan)	
Properti Keluarga	- Termasuk Kidhome (jumlah anak di rumah)	
	- Teenhome (jumlah remaja di rumah)	
Perilaku Pembelian	- Recency (jumlah hari sejak pembelian terakhir)	
	- MntWines (jumlah uang yang dikeluarkan untuk anggur)	
	- MntFruits (jumlah uang yang dikeluarkan untuk buah)	
	- MntMeatProducts (jumlah uang yang dikeluarkan untuk produk daging)	
	- MntFishProducts (jumlah uang yang dikeluarkan untuk produk ikan)	
	- MntSweetProducts (jumlah uang yang dikeluarkan untuk produk manis)	
	- MntGoldProds (jumlah uang yang dikeluarkan untuk produk emas)	
	- NumDealsPurchases (jumlah pembelian dengan potongan harga)	
	- NumWebPurchases (jumlah pembelian melalui web)	
	- NumCatalogPurchases (jumlah pembelian melalui katalog)	
	- NumStorePurchases (jumlah pembelian di toko fisik).	
	Aktivitas Web dan Kampanye	- NumWebVisitsMonth (jumlah kunjungan situs bulanan)
		- AcceptedCmp1 sampai AcceptedCmp5 (indikator penerimaan kampanye 1 hingga 5)
- Complain (indikator apakah pelanggan mengajukan keluhan)		
Informasi Tambahkan	- Response (indikator penerimaan kampanye paling akhir)	
	- Dt_Customer (tanggal registrasi pelanggan)	
	- Z_CostContact	
	- Z_Revenue	

Pemeriksaan awal terhadap dataset menunjukkan bahwa terdapat nilai yang hilang pada kolom Pendapatan. Atribut lainnya umumnya sudah lengkap, tetapi ada kolom dengan tipe data non-numerik (Education, Marital_Status, Dt_Customer) yang perlu ditransformasi sebelum analisis dilakukan. Informasi tambahan mengenai pengelolaan kualitas data akan dijelaskan pada bagian Pra-pemrosesan data.

2.2. Pra-pemrosesan Data

Tahap pra-pemrosesan data adalah langkah penting dalam setiap penelitian data mining, yang bertujuan untuk mentransformasi data mentah menjadi format yang rapi, konsisten, dan siap untuk dianalisis. Algoritma K-Means sangat peka terhadap kualitas dan skala data, sehingga pra-pemrosesan yang sesuai sangat krusial untuk memperoleh hasil klustering yang tepat dan relevan. Tahapan pra-

pemrosesan yang dilaksanakan mencakup pengelolaan nilai yang hilang, rekayasa fitur, penghapusan kolom yang tidak penting, transformasi fitur kategorikal, serta penskalaan fitur numerik

2.2.1. Penanganan Data yang Hilang

Langkah pertama adalah mengidentifikasi apakah terdapat nilai-nilai yang hilang dalam data. Ditemukan bahwa kolom Income memiliki sejumlah data yang kosong. Untuk mengatasinya, digunakan pendekatan dengan mengganti nilai-nilai kosong tersebut menggunakan rata-rata (mean) dari seluruh nilai yang tersedia pada kolom tersebut. Pendekatan ini dipilih agar tidak menghilangkan baris data yang masih mengandung informasi penting lainnya.

2.2.2. Rekayasa Fitur (Feature Engineering)

Agar data menjadi lebih informatif, dibuat dua fitur baru berdasarkan kolom yang sudah ada:

- Usia Pelanggan dihitung dari selisih antara tahun sekarang (2025) dan tahun lahir pelanggan (Year_Birth).
- Lama menjadi pelanggan (Customer_Tenure) dihitung berdasarkan selisih waktu antara tanggal saat ini dengan tanggal pertama kali pelanggan mendaftar (Dt_Customer).

Setelah dua fitur baru tersebut dibuat, kolom Year_Birth dan Dt_Customer dihapus karena nilainya telah diwakili oleh fitur yang lebih relevan.

2.2.3. Menghapus Kolom yang Tidak Relevan

Beberapa kolom diketahui tidak memberikan informasi yang berguna untuk proses klustering, sehingga dihapus dari dataset. Kolom ID, yang hanya berfungsi sebagai pengenalan unik untuk setiap entri, dihapus karena tidak memiliki nilai analitis. Selain itu, kolom Z_CostContact dan Z_Revenue juga dihapus karena nilai di seluruh barisnya sama, sehingga tidak berkontribusi pada variasi data.

2.2.4. Mengubah Data Kategorikal ke Bentuk Numerik

Beberapa kolom dalam dataset berisi data kategorikal, seperti Education dan Marital_Status. Karena algoritma K-Means hanya dapat bekerja dengan data numerik, maka nilai-nilai kategorikal ini perlu diubah. Digunakan metode One-Hot Encoding, di mana setiap kategori diubah menjadi kolom tersendiri yang bernilai 0 atau 1. Dengan demikian, data kategorikal dapat dimasukkan ke dalam proses klustering tanpa mengasumsikan adanya urutan antar kategori.

2.2.5. Menyamakan Skala Nilai (Standard Scaling)

Langkah terakhir dalam pra-pemrosesan adalah menyamakan skala antar fitur numerik. Hal ini penting karena algoritma K-Means sangat bergantung pada penghitungan jarak, dan fitur dengan nilai besar dapat mendominasi hasil klustering. Oleh karena itu, dilakukan standarisasi, yaitu mengubah setiap nilai fitur agar memiliki rata-rata nol dan standar deviasi satu. Dengan begitu, setiap fitur akan berkontribusi secara seimbang dalam proses klustering.

2.3. K-Means Algorithm

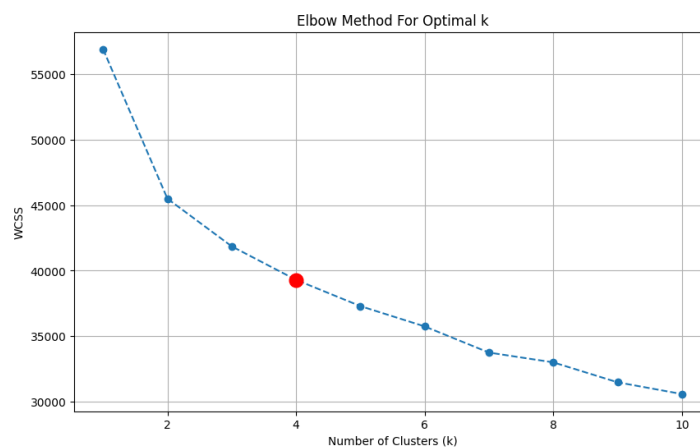
Algoritma K-Means adalah salah satu teknik pengelompokan (*clustering*) data non-hierarkis yang paling mendasar, banyak dipakai, dan diakui efektivitasnya dalam komunitas penelitian data mining. Intinya adalah kemampuan untuk mengatur sekumpulan n objek data atau pengamatan ke dalam k grup atau kluster berbeda. Tujuan utama algoritma ini adalah untuk mengecilkan varians atau dispersi dalam setiap kluster, sekaligus meningkatkan perbedaan antara kluster-kluster yang ada.

Secara konseptual, K-Means beroperasi dengan menemukan k titik pusat, yang disebut sebagai centroid, untuk setiap kluster. Setiap centroid ini berfungsi sebagai prototipe atau perwujudan dari kluster yang relevan. Proses pengelompokan mencakup penugasan setiap pengamatan data ke kluster

yang pusatnya memiliki jarak terdekat (biasanya menggunakan jarak *Euclidean*) dengan pengamatan tersebut[21], [22], [23]. Penugasan ini secara berulang disesuaikan, dan posisi centroid diperbarui dengan menggunakan rata-rata observasi yang termasuk dalam kluster tersebut, sampai kondisi konvergensi tercapai di mana penugasan kluster tidak lagi berubah atau perubahan yang terjadi sangat minimal[24], [25], [26].

2.4. Analisis dan Penarikan Kesimpulan

Bagian ini memaparkan hasil analisis klustering yang telah dilaksanakan dengan menggunakan algoritma K-Means pada dataset kampanye pemasaran. Setelah menyelesaikan tahap pra-pemrosesan data dan menentukan jumlah kluster optimal dengan *Elbow Method*, dataset berhasil dibagi menjadi empat kluster yang berbeda. Analisis mendetail terhadap sifat setiap kluster memberikan pemahaman penting tentang profil dan perilaku segmen pelanggan.



Gambar 1. Grafik Elbow Method

Berikut adalah profil karakteristik rata-rata untuk setiap kluster yang teridentifikasi, yang diinterpretasikan untuk menggambarkan perilaku dan preferensi masing-masing segmen:

1. Kluster 0: "Pembelanja Seimbang dengan Remaja"

- Pelanggan dengan pendapatan sedikit di atas rata-rata dan cenderung memiliki lebih banyak remaja di rumah.
- Pengeluaran sedang untuk berbagai kategori produk, dengan sedikit preferensi untuk anggur dan produk emas.
- Terlibat dalam berbagai saluran pembelian (web, toko, katalog) dan cukup responsif terhadap kampanye.

2. Kluster 1: "Keluarga Sadar Anggaran"

- Segmen terbesar dengan pendapatan lebih rendah dan lebih banyak anak kecil di rumah.
- Pengeluaran yang jauh lebih rendah di semua kategori produk, menunjukkan bahwa mereka lebih sadar anggaran.
- Memiliki jumlah pembelian keseluruhan yang lebih sedikit.

3. Kluster 2: "Pembeli Bernilai Tinggi yang Terlibat"

- Segmen ini terdiri dari pelanggan kaya tanpa atau dengan sedikit anak.
- Mereka adalah pembelanja besar di semua kategori produk dan sangat responsif terhadap kampanye pemasaran.
- Lebih memilih pembelian melalui katalog dan di toko daripada pembelian web, dan memiliki jumlah kunjungan web yang lebih sedikit. Ini kemungkinan adalah pelanggan yang paling berharga.

4. Kluster 3: "Pembeli Dewasa Aktif"

- Pelanggan dengan pendapatan di atas rata-rata dan biasanya tidak memiliki atau memiliki sedikit anak.
- Pengeluaran tinggi di semua kategori produk, terutama untuk buah-buahan, daging, ikan, dan produk manis.
- Juga lebih memilih pembelian melalui katalog dan di toko, mirip dengan Kluster 2, tetapi kurang terlibat dalam kampanye pemasaran dibandingkan Kluster 2.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Clustering

Proses klustering menggunakan algoritma K-Means telah berhasil membagi dataset pelanggan menjadi empat kluster yang berbeda. Penentuan jumlah kluster dilakukan sebelumnya menggunakan Elbow Method, yang menunjukkan bahwa nilai $k=4$ merupakan titik siku (*elbow*) yang paling jelas, yaitu perubahan dari penurunan inerti yang tidak lagi signifikan pada grafik.

Tabel 1. Distribusi Data pada Setiap Kluster

Kluster	Jumlah Data
0	591
1	1017
2	170
3	462

Dari tabel di atas dapat dilihat bahwa kluster 1 merupakan kluster dengan jumlah anggota terbanyak, yaitu sebanyak 1.017 data, sedangkan kluster dengan anggota paling sedikit adalah kluster 2 dengan 170 data. Hal ini menunjukkan adanya ketidakseimbangan distribusi antar kluster, yang dapat mencerminkan perbedaan karakteristik signifikan di antara segmen pelanggan yang terbentuk.

3.2. Interpretasi Kluster

Untuk memahami profil masing-masing segmen pelanggan, analisis mendalam dilakukan dengan menghitung nilai rata-rata dari semua fitur untuk setiap kluster. Perbandingan nilai rata-rata ini memungkinkan identifikasi karakteristik dominan yang membedakan satu kluster dari kluster lainnya. Berikut adalah interpretasi profil untuk keempat kluster yang teridentifikasi:

- Kluster 0: "Pembelanja Seimbang dengan Remaja"

Kluster ini dicirikan oleh pelanggan dengan pendapatan yang sedikit di atas rata-rata dan jumlah remaja yang tinggal di rumah. Teenhome memiliki nilai yang lebih tinggi, sedangkan Kidhome memiliki nilai yang lebih rendah. Sebagian besar pelanggan berasal dari kelompok usia yang lebih tua. Mereka menunjukkan pola pengeluaran yang seimbang untuk berbagai jenis barang, dengan sedikit kecenderungan untuk membeli barang-barang emas dan anggur. Mereka beroperasi di toko fisik, melalui diskon (NumDealsPurchases), web (NumWebPurchases), dan di toko fisik (NumStorePurchases). Respon mereka terhadap kampanye pemasaran cenderung moderat. Penawaran yang berfokus pada nilai dan diskon keluarga, serta barang-barang yang terkait dengan keluarga remaja, menjanjikan di segmen ini.

- Kluster 1: "Keluarga Sadar Anggaran"

Dengan jumlah anak kecil yang lebih besar dan pendapatan relatif di bawah rata-rata, kluster ini merupakan segmen terbesar dari basis pelanggan. Sebagian besar pelanggan berasal dari kelompok usia yang lebih muda. Mereka adalah segmen yang paling hemat, dan mereka menurunkan biaya untuk semua barang, mulai dari anggur, buah, daging, ikan, manis, emas, dan lainnya. Selain itu, jumlah total pembelian mereka sangat rendah, yang menunjukkan sensitivitas harga yang tinggi. Selain itu, mereka tidak terlibat dan tidak menanggapi kampanye pemasaran karena profil ini. Strategi pemasaran untuk segmen ini harus berfokus pada promosi yang menekankan nilai uang dan penjualan produk penting dengan harga yang sangat kompetitif.

- Kluster 2: "Pembeli Bernilai Tinggi yang Terlibat"

Di antara semua kluster, segmen ini memiliki pendapatan tertinggi dan biasanya tidak memiliki atau hanya memiliki sedikit anak. Pelanggan dalam kluster ini cenderung dewasa-akhir. Mereka menunjukkan daya beli yang sangat tinggi karena mereka adalah pembeli terbesar di hampir semua kategori barang. Selain itu, kluster ini sangat menyukai pembelian di toko fisik dan melalui katalog, yang menunjukkan bahwa mereka mungkin menginginkan pengalaman belanja yang lebih premium atau terkurasi dengan jumlah kunjungan web yang cenderung lebih rendah. Yang paling menonjol, kluster ini secara konsisten memiliki tingkat respons dan penerimaan tertinggi untuk setiap kampanye pemasaran yang dijalankan; ini membuatnya segmen yang sangat berharga dan responsif. Tingkat keluhan yang sedikit lebih tinggi, bagaimanapun, mungkin menunjukkan bahwa pelanggan juga mengharapkan layanan yang baik.

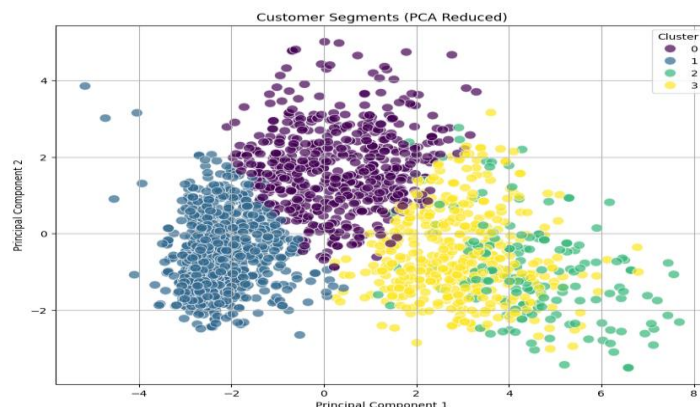
- Kluster 3: "Pembeli Dewasa Aktif"

Pelanggan dalam kluster ini memiliki pendapatan di atas rata-rata dan umumnya tanpa anak di rumah, mirip dengan Kluster 2, namun cenderung sedikit lebih muda. Pola pengeluaran mereka tinggi, terutama untuk produk tertentu seperti buah-buahan, daging, ikan, dan produk manis. Pilihan mereka untuk membeli produk melalui katalog dan toko fisik mirip dengan Kluster 2, tetapi tingkat respons mereka terhadap kampanye pemasaran cenderung lebih rendah daripada Kluster 2. Ini menunjukkan bahwa mereka mungkin lebih tertarik pada kualitas atau spesifikasi produk tertentu daripada promosi umum.

3.3. Visualisasi

Untuk memahami secara visual distribusi kluster-kluster ini dalam ruang fitur, teknik reduksi dimensi yaitu Principal Component Analysis (PCA) digunakan. PCA menurunkan dimensi data asli yang tinggi menjadi dua komponen utama (PC1 dan PC2) yang menangkap varians terbesar, sehingga memungkinkan visualisasi dalam dua dimensi.

Plot sebar (scatterplot) dari PC1 dan PC2 memperlihatkan setiap pelanggan diplot sesuai dengan dua komponen utama ini, dengan titik-titik yang diwarnai berdasarkan kluster keberadaan mereka.



Gambar 2. Visualisasi Segmen Pelanggan Menggunakan PCA

Visualisasi ini dengan jelas menggambarkan perbedaan antara kluster-kluster, meskipun terdapat beberapa tumpang tindih, terutama antara Kluster 0, 1, dan 3. Kluster 2, yang merupakan segmen pembelanja dengan nilai tinggi, tampak cukup terpisah dari kluster lainnya, menegaskan adanya perbedaan karakteristik yang signifikan. Kepadatan titik-titik dalam setiap kluster menggambarkan kekompakan internal kluster itu, sedangkan jarak antara kluster menunjukkan perbedaan antar segmen.

3.4. Pembahasan Hasil

Hasil klusterisasi yang diperoleh dari algoritma K-Means berhasil membagi data pelanggan ke dalam empat kluster yang berbeda. Berdasarkan evaluasi metode menggunakan Elbow Method, nilai optimal $k=4$ dipilih karena memberikan keseimbangan antara kompleksitas model dan kemampuan untuk membedakan segmen pelanggan secara jelas.

Distribusi jumlah anggota dalam masing-masing kluster menunjukkan adanya variasi yang cukup besar. Kluster 1 memiliki jumlah anggota terbanyak, yaitu 1.017 pelanggan, sedangkan kluster 2 menjadi segmen dengan jumlah anggota paling sedikit, yaitu 170 pelanggan. Hal ini menunjukkan bahwa sebagian besar pelanggan perusahaan termasuk dalam segmen "Keluarga Sadar Anggaran", sementara hanya sebagian kecil yang masuk ke dalam kategori "Pembeli Bernilai Tinggi yang Terlibat".

Analisis karakteristik rata-rata tiap fitur dalam kluster mengungkapkan perbedaan yang signifikan di antara keempat kluster, sehingga dapat disimpulkan bahwa segmentasi yang dilakukan efektif dalam membedakan perilaku dan preferensi pelanggan.

Secara detail, profil keempat kluster adalah sebagai berikut:

- Kluster 0: "Pembelanja Seimbang dengan Remaja"
Kluster ini berisi pelanggan dengan pendapatan sedikit di atas rata-rata dan jumlah remaja yang lebih banyak di rumah. Mereka memiliki pola pengeluaran yang seimbang untuk berbagai jenis barang dan cukup aktif dalam berbagai saluran pembelian. Respons terhadap kampanye pemasaran juga berada pada tingkat menengah. Strategi yang cocok untuk kluster ini melibatkan promosi produk keluarga serta diskon musiman.
- Kluster 1: "Keluarga Sadar Anggaran"
Segmen terbesar ini memiliki pendapatan lebih rendah dan jumlah anak kecil yang lebih banyak. Mereka melakukan pengeluaran minimal untuk semua kategori produk dan cenderung tidak responsif terhadap kampanye pemasaran. Strategi promosi berupa paket hemat dan diskon kebutuhan pokok akan lebih efektif untuk kluster ini.
- Kluster 2: "Pembeli Bernilai Tinggi yang Terlibat"
Kluster ini terdiri dari pelanggan dengan pendapatan tertinggi, sedikit atau tanpa anak, dan sangat responsif terhadap kampanye pemasaran. Mereka menunjukkan pengeluaran yang besar di hampir semua kategori produk. Pelanggan dalam kluster ini lebih suka berbelanja melalui katalog dan toko fisik. Karena nilai pelanggan yang tinggi, segmen ini memerlukan pendekatan pemasaran personalisasi dan layanan premium.
- Kluster 3: "Pembeli Dewasa Aktif"
Kluster ini memiliki pelanggan dengan pendapatan di atas rata-rata dan cenderung tidak memiliki anak. Pengeluaran mereka tinggi, khususnya untuk produk makanan. Mereka kurang responsif terhadap kampanye pemasaran dibandingkan kluster 2, sehingga penawaran produk berkualitas dan inovasi produk baru lebih relevan untuk kluster ini.

Untuk memperjelas pemisahan antar kluster, dilakukan visualisasi menggunakan teknik Principal Component Analysis (PCA). Hasil visualisasi scatterplot dua dimensi menunjukkan bahwa meskipun terdapat tumpang tindih di beberapa area terutama antara kluster 0, 1, dan 3 serta kluster 2 yang terlihat

cukup terpisah dari kluster lainnya. Ini mengindikasikan perbedaan yang nyata dalam karakteristik segmen pelanggan bernilai tinggi tersebut dibandingkan kluster lainnya.

Visualisasi ini mendukung interpretasi profil kluster yang telah dibahas sebelumnya, serta memperkuat validitas hasil segmentasi dalam memberikan pemahaman yang lebih baik mengenai struktur pelanggan.

4. KESIMPULAN

Penelitian ini berhasil menggunakan algoritma K-Means pada dataset kampanye pemasaran dan membagi pelanggan ke dalam empat segmen yang berbeda. Data disiapkan secara optimal melalui pra-pemrosesan data yang menyeluruh, yang mencakup penanganan nilai hilang, rekayasa fitur, penghapusan kolom yang tidak relevan, transformasi kategorikal, dan penskalaan numerik. Elbow Method menemukan empat kelompok pelanggan yang paling tepat, yang kemudian diubah menjadi profil pelanggan yang jelas: "Pembelanja Seimbang dengan Remaja", "Keluarga Sadar Anggaran", "Pembeli Bernilai Tinggi yang Terlibat", dan "Pembeli Dewasa Aktif". Pendekatan pemasaran yang disesuaikan sangat penting karena perbedaan besar dalam demografi, perilaku pembelian, dan respons kampanye antara segmen ini.

Sebuah temuan penting adalah penemuan "Pembeli Bernilai Tinggi yang Terlibat" (Kluster 2), segmen yang paling menguntungkan dan responsif, yang membutuhkan strategi pemasaran premium yang disesuaikan secara khusus. Sebaliknya, segmen terbesar, "Keluarga Sadar Anggaran" (Kluster 1), membutuhkan pendekatan yang berfokus pada nilai dan diskon. Kluster lain, "Pembeli Dewasa Aktif" (Kluster 3) dan "Pembelanja Seimbang dengan Remaja" (Kluster 0), juga menawarkan peluang penargetan spesifik berdasarkan karakteristik pengeluaran dan preferensi produk mereka.

Secara keseluruhan, wawasan strategis yang mendalam yang diberikan oleh segmentasi ini memungkinkan perusahaan untuk mengalokasikan sumber daya pemasaran dengan lebih efisien, menghasilkan pesan yang lebih relevan, dan pada akhirnya meningkatkan kepuasan pelanggan dan profitabilitas bisnis. Diharapkan bahwa hasil dari penerapan informasi ini di masa mendatang dapat membantu mengoptimalkan kinerja kampanye pemasaran.

REFERENCES

- [1] D. Müllensiefen, C. Hennig, and H. Howells, "Using clustering of rankings to explain brand preferences with personality and socio-demographic variables," Apr. 2017, [Online]. Available: <http://arxiv.org/abs/1704.00959>
- [2] 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.
- [3] 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, Dec. 2023, doi: 10.3390/analytics2040042.
- [4] L. A. S. Kristiyowati, F. M. Hana, and W. C. Wahyudin, "Penerapan Algoritma Naive Bayes pada Sistem Chatbot Persewaan Kos," *Sainteks*, vol. 22, no. 1, pp. 53–62, Apr. 2025, doi: 10.30595/sainteks.v22i1.25962.
- [5] D. Chen, S. L. Sain, and K. Guo, "Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining," *Journal of Database Marketing and Customer Strategy Management*, vol. 19, no. 3, pp. 197–208, Sep. 2012, doi: 10.1057/dbm.2012.17.
- [6] W. G. Mangold and D. J. Faulds, "Social media: The new hybrid element of the promotion mix," *Bus Horiz*, vol. 52, no. 4, pp. 357–365, Jul. 2009, doi: 10.1016/j.bushor.2009.03.002.

-
- [7] M. S. E. Kasem, M. Hamada, and I. Taj-Eddin, "Customer profiling, segmentation, and sales prediction using AI in direct marketing," *Neural Comput Appl*, vol. 36, no. 9, pp. 4995–5005, Mar. 2024, doi: 10.1007/s00521-023-09339-6.
- [8] M. Alves Gomes and T. Meisen, "A review on customer segmentation methods for personalized customer targeting in e-commerce use cases," *Information Systems and e-Business Management*, vol. 21, no. 3, pp. 527–570, Sep. 2023, doi: 10.1007/s10257-023-00640-4.
- [9] D. Paundra Gevano, "Segmentasi Pelanggan Menggunakan K-Means Clustering Berdasarkan Data Kepribadian dan Pola Konsumsi: Studi Kasus pada Dataset Customer Personality Analysis H1D023075 H1D023089," 2025.
- [10] S. Prisilia, N. R. Fitriya, A. Febriana, and N. N. A. Putri, "Perbandingan Kinerja Keuangan Perusahaan Menggunakan Analisis Rasio," *GEMILANG: Jurnal Manajemen dan Akuntansi*, vol. 3, Jan. 2023.
- [11] R. S. Fams and A. W. Lubis, "Konsep Penguat Usaha Ekonomi Rakyat Dengan Menggunakan Cloud Computing Literature Study," *Al-Kharaj : Jurnal Ekonomi, Keuangan & Bisnis Syariah*, vol. 6, no. 2, pp. 685–697, Feb. 2023, doi: 10.47467/alkharaj.v6i2.3227.
- [12] F. Sutomo *et al.*, "OPTIMIZATION OF THE K-NEAREST NEIGHBORS ALGORITHM USING THE ELBOW METHOD ON STROKE PREDICTION," *Jurnal Teknik Informatika (JUTIF)*, vol. 4, no. 1, 2023, doi: 10.20884/1.jutif.2023.4.1.839.
- [13] 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.
- [14] 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.
- [15] B. Benaissa, M. Kobayashi, and H. Takenouchi, "Enhancing Consumer Agent Modeling Through Openness-Based Consumer Traits and Inverse Clustering," *Mach Learn Knowl Extr*, vol. 7, no. 1, Mar. 2025, doi: 10.3390/make7010009.
- [16] S. N. Lathifah and Z. F. Azzahra, "AI-Driven Customers Segmentation Using K-Means Clustering," *G-Tech: Jurnal Teknologi Terapan*, vol. 9, no. 1, pp. 320–329, Jan. 2025, doi: 10.70609/gtech.v9i1.6202.
- [17] N. Jain and V. Ahuja, "Segmenting online consumers using K-means cluster analysis," 2014.
- [18] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," Aug. 01, 2020, *MDPI AG*. doi: 10.3390/electronics9081295.
- [19] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit Lett*, vol. 31, no. 8, pp. 651–666, Jun. 2010, doi: 10.1016/j.patrec.2009.09.011.
- [20] L. Afuan, N. Hidayat, H. Hamdani, H. Ismanto, B. C. Purnama, and D. I. Ramdhani, "Optimizing BERT Models with Fine-Tuning for Indonesian Twitter Sentiment Analysis," *J Wirel Mob Netw Ubiquitous Comput Dependable Appl*, vol. 16, no. 2, pp. 248–267, Jun. 2025, doi: 10.58346/JOWUA.2025.12.016.
- [21] N. Miftahul Janna and D. Pembimbing, "KONSEP UJI VALIDITAS DAN RELIABILITAS DENGAN MENGGUNAKAN SPSS."
- [22] S. N. Agni, M. I. Djomiy, R. Fernando, and C. Apriono, "Evaluasi Penerapan Smart Mobility di Jakarta (Evaluation of Smart Mobility Implementation in Jakarta)," 2021.
- [23] M. Shutaywi and N. N. Kachouie, "Silhouette analysis for performance evaluation in machine learning with applications to clustering," *Entropy*, vol. 23, no. 6, Jun. 2021, doi: 10.3390/e23060759.
- [24] F. P. Rachman, H. Santoso, and A. Djajadi, "Machine Learning Mini Batch K-means and Business Intelligence Utilization for Credit Card Customer Segmentation." [Online]. Available: www.ijacsa.thesai.org
-

- [25] H. M. Tenkam, J. C. Mba, and S. M. Mwambi, "Optimization and Diversification of Cryptocurrency Portfolios: A Composite Copula-Based Approach," *Applied Sciences (Switzerland)*, vol. 12, no. 13, Jul. 2022, doi: 10.3390/app12136408.
- [26] E. U. Oti, M. O. Olusola, F. C. Eze, and S. U. Enogwe, "Comprehensive Review of K-Means Clustering Algorithms," *International Journal of Advances in Scientific Research and Engineering*, vol. 07, no. 08, pp. 64–69, 2021, doi: 10.31695/ijasre.2021.34050.