

Implementasi Segmentasi Pelanggan Menggunakan Metode K-Means pada Dataset Marketing Campaign Berbasis Website

Varel Deva Dewangga^{1*}, Anggi Aziyah Kusumaningrum², Andy Aldyansyah³, Ahmad Dwi Rifa'i⁴

¹ Teknik Informatika/ Fakultas Ilmu
Komputer

Universitas Duta Bangsa

^{1*} 230103120@mhs.udb.ac.id

² Teknik Informatika/ Fakultas Ilmu
Komputer

Universitas Duta Bangsa

² 230103091@mhs.udb.ac.id

³ Teknik Informatika / Fakultas Ilmu
Komputer

Universitas Duta Bangsa

³ 230103090@mhs.udb.ac.id

⁴ Teknik Informatika/ Fakultas Ilmu
Komputer

Universitas Duta Bangsa

⁴ 230103085@mhs.udb.ac.id

Abstrak— Perkembangan teknologi informasi menghasilkan data pelanggan dalam jumlah besar yang dapat dimanfaatkan untuk mendukung pengambilan keputusan bisnis. Namun, data tersebut sering kali belum diolah secara optimal sehingga diperlukan penerapan teknik data mining untuk memperoleh informasi yang bermanfaat. Penelitian ini bertujuan mengimplementasikan algoritma K-Means Clustering pada Marketing Campaign Dataset untuk melakukan segmentasi pelanggan berdasarkan karakteristik dan perilaku pembelian. Tahapan penelitian meliputi data cleaning, feature engineering, normalisasi data menggunakan Min-Max Scaling, penentuan jumlah cluster dengan Elbow Method, serta proses clustering menggunakan algoritma K-Means. Dataset yang digunakan berjumlah 2.240 data pelanggan, kemudian setelah proses data cleaning diperoleh 2.215 data yang digunakan dalam proses analisis. Hasil segmentasi selanjutnya diimplementasikan ke dalam aplikasi berbasis web yang dikembangkan menggunakan React, Supabase, dan Vercel, sehingga pengguna dapat mengimpor data, menjalankan proses clustering, serta melihat hasil segmentasi pelanggan secara interaktif. Implementasi sistem mampu mengelompokkan pelanggan ke dalam beberapa cluster berdasarkan karakteristik yang dimiliki sehingga dapat memberikan informasi yang berguna dalam penyusunan strategi pemasaran, penentuan target promosi, serta pengambilan keputusan bisnis yang lebih efektif dan berbasis data. Hasil evaluasi menunjukkan bahwa jumlah cluster optimal adalah 4 dengan nilai Silhouette Score sebesar 0.3723 sehingga menghasilkan segmentasi pelanggan yang baik.

Kata kunci— K-Means Clustering, Segmentasi Pelanggan, Data Mining, Marketing Campaign Dataset, Aplikasi Berbasis Web.

Abstract— The rapid development of information technology has enabled organizations to collect large amounts of customer data that can be utilized to support business decision-making. However, these data are often not optimally analyzed, making it necessary to apply data mining techniques to extract valuable insights. This study aims to implement the K-Means Clustering algorithm on the Marketing Campaign Dataset to perform customer segmentation based on demographic characteristics and purchasing behavior. The research methodology consists of data cleaning, feature engineering, data normalization using the Min-Max Scaling method, determination of the optimal number of clusters using the Elbow Method, and customer segmentation using the K-Means algorithm. The dataset initially contained 2,240 customer records, and after the data cleaning process, 2,215 records were retained for analysis. The clustering results were implemented in a web-based application developed using React, Supabase, and Vercel, enabling users to import customer data, execute the clustering process, and visualize customer segmentation interactively. The implemented system successfully groups customers into several clusters based on similar characteristics, providing useful insights for marketing strategy development, customer targeting, and data-driven business decision-making.

Keywords— K-Means Clustering; Customer Segmentation; Data Mining; Marketing Campaign Dataset; Web-Based Application.

I. PENDAHULUAN

Perkembangan teknologi informasi dan digitalisasi telah mendorong perusahaan untuk mengumpulkan data pelanggan dalam jumlah

yang semakin besar. Data tersebut mencakup informasi mengenai karakteristik pelanggan, kondisi demografis, serta riwayat transaksi yang dapat dimanfaatkan untuk mendukung

pengambilan keputusan bisnis. Namun, banyak perusahaan masih belum mampu mengolah data tersebut secara optimal sehingga potensi dalam meningkatkan efektivitas strategi pemasaran belum dapat dimanfaatkan secara maksimal. Oleh karena itu, diperlukan penerapan teknik data mining yang mampu mengolah data pelanggan menjadi informasi yang bermanfaat bagi perusahaan [1], [2].

Segmentasi pelanggan merupakan salah satu teknik dalam data mining yang digunakan untuk mengelompokkan pelanggan berdasarkan karakteristik dan perilaku yang serupa. Melalui proses segmentasi, perusahaan dapat memahami kebutuhan setiap kelompok pelanggan sehingga strategi pemasaran, promosi, maupun program loyalitas dapat disusun secara lebih tepat sasaran. Selain itu, hasil segmentasi juga dapat membantu perusahaan dalam mengoptimalkan sumber daya pemasaran dan meningkatkan kepuasan pelanggan [3], [4], [5].

Salah satu metode yang banyak digunakan dalam proses segmentasi pelanggan adalah algoritma K-Means Clustering. Algoritma ini merupakan metode unsupervised learning yang mengelompokkan data berdasarkan tingkat kemiripan karakteristik antarobjek sehingga setiap kelompok memiliki tingkat homogenitas yang tinggi. K-Means dipilih karena memiliki proses komputasi yang relatif sederhana, mudah diimplementasikan, serta mampu menghasilkan proses clustering yang efisien pada data dengan jumlah besar [6], [7].

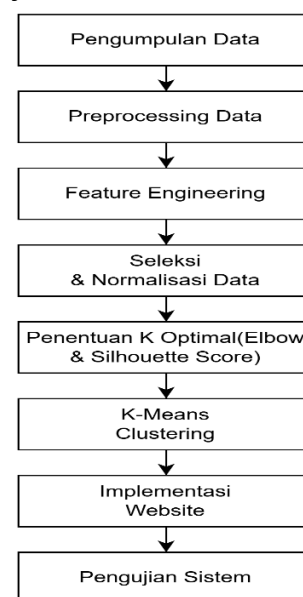
Penelitian ini menggunakan Marketing Campaign Dataset yang diperoleh dari Kaggle sebagai sumber data pelanggan. Sebelum proses clustering dilakukan, dataset terlebih dahulu melalui tahap preprocessing yang meliputi pembersihan data (data cleaning), penanganan missing value, serta feature engineering. Pada tahap ini dibentuk beberapa atribut baru, seperti Age, Spending, Seniority, Has_child, dan Children, sehingga data yang digunakan lebih mampu merepresentasikan karakteristik pelanggan dibandingkan hanya menggunakan atribut mentah dari dataset. Selanjutnya, atribut hasil feature engineering dinormalisasi sebelum digunakan pada proses K-Means Clustering [8].

Berbeda dengan beberapa penelitian sebelumnya yang hanya berfokus pada proses clustering, penelitian ini juga mengimplementasikan hasil segmentasi ke dalam sebuah aplikasi berbasis web. Aplikasi dikembangkan menggunakan React sebagai antarmuka pengguna, Supabase sebagai basis data, dan Vercel sebagai media deployment. Melalui aplikasi tersebut, pengguna dapat mengimpor data pelanggan, menjalankan proses segmentasi secara otomatis menggunakan algoritma K-Means, serta melihat hasil pengelompokan pelanggan secara interaktif [2], [9].

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk mengimplementasikan algoritma K-Means Clustering pada Marketing Campaign Dataset serta mengembangkan aplikasi berbasis web yang mampu melakukan segmentasi pelanggan secara otomatis. Hasil penelitian diharapkan dapat membantu pengguna dalam memahami karakteristik setiap kelompok pelanggan sehingga dapat mendukung penyusunan strategi pemasaran yang lebih efektif dan berbasis data [10], [11].

II. METODOLOGI PENELITIAN

Penelitian ini dilakukan melalui beberapa tahapan, dimulai dari pengumpulan data, preprocessing data, proses segmentasi menggunakan algoritma K-Means Clustering, implementasi hasil segmentasi ke dalam aplikasi berbasis web, hingga pengujian sistem [8], [10].



Gambar 1. Tahapan Penelitian

1. Pengumpulan Data

Penelitian ini menggunakan Marketing Campaign Dataset yang diperoleh dari platform Kaggle tahun 2021. Dataset tersebut berisi informasi mengenai karakteristik pelanggan, seperti usia, tingkat pendidikan, status pernikahan, pendapatan, jumlah anak, serta riwayat pembelian pada berbagai kategori produk. Dataset ini dipilih karena memiliki atribut yang mampu merepresentasikan perilaku pelanggan sehingga sesuai digunakan dalam proses segmentasi menggunakan algoritma K-Means Clustering. Data yang digunakan kemudian diolah sebagai dasar dalam pembentukan model segmentasi pelanggan.

2. Preprocessing Data

Tahap preprocessing dilakukan untuk meningkatkan kualitas data sebelum proses clustering dilakukan. Pada tahap awal dilakukan pemeriksaan terhadap missing value dan penghapusan data yang memiliki nilai kosong pada atribut Income. Selanjutnya dilakukan feature engineering untuk membentuk atribut baru yang lebih representatif terhadap karakteristik pelanggan. Atribut Age diperoleh dari hasil konversi Year_Birth, sedangkan atribut Spending dihitung dari total pengeluaran pelanggan pada kategori Wines, Fruits, Meat, Fish, Sweets, dan Gold. Selain itu dibentuk atribut Seniority, Has_child, dan Children berdasarkan informasi yang terdapat pada dataset.

Tahap berikutnya adalah pemilihan atribut yang digunakan dalam proses clustering, yaitu Age, Income, Spending, Seniority, serta atribut pendukung lainnya yang menggambarkan karakteristik pelanggan. Seluruh atribut numerik kemudian dinormalisasi menggunakan metode Min-Max Scaling agar setiap atribut memiliki rentang nilai yang sama sehingga tidak terjadi dominasi nilai tertentu pada proses pengelompokan data [4], [7]

3. Feature Engineering

Setelah proses data cleaning selesai, dilakukan tahap feature engineering untuk membentuk atribut baru yang lebih representatif terhadap karakteristik pelanggan. Proses ini bertujuan

meningkatkan kualitas data sehingga informasi yang digunakan dalam proses clustering lebih relevan. Beberapa atribut baru yang dibentuk meliputi Age, yang diperoleh dari konversi atribut Year_Birth, Spending, yang merupakan total pengeluaran pelanggan dari kategori Wines, Fruits, Meat, Fish, Sweets, dan Gold, serta atribut Seniority, Has_child, dan Children yang menggambarkan lama pelanggan bergabung serta kondisi keluarga pelanggan. Hasil dari feature engineering digunakan sebagai atribut utama dalam proses segmentasi menggunakan algoritma K-Mean [2]

4. Seleksi & Normalisasi Data

Tahap selanjutnya adalah melakukan seleksi atribut yang akan digunakan pada proses clustering. Atribut yang dipilih meliputi Age, Income, Spending, Seniority, serta atribut pendukung lainnya yang dianggap mampu merepresentasikan karakteristik pelanggan. Setelah proses seleksi selesai, dilakukan normalisasi data menggunakan metode Min-Max Scaling untuk mengubah seluruh atribut numerik ke dalam rentang nilai yang sama [12]. Normalisasi dilakukan agar tidak terdapat atribut yang mendominasi proses perhitungan jarak pada algoritma K-Means, sehingga hasil pengelompokan menjadi lebih optimal.

5. Penentuan K optimal (Elbow & Silhouette Score)

Sebelum proses clustering dilakukan, terlebih dahulu ditentukan jumlah cluster (K) yang paling optimal. Penelitian ini menggunakan Elbow Method untuk mengidentifikasi nilai K terbaik berdasarkan perubahan nilai Within Cluster Sum of Squares (WCSS) [13]. Selanjutnya, hasil tersebut divalidasi menggunakan Silhouette Score untuk mengukur kualitas pemisahan antarcluster. Berdasarkan hasil pengujian, diperoleh jumlah cluster optimal sebanyak 4 cluster dengan nilai Silhouette Score sebesar 0,3723. Nilai tersebut menunjukkan bahwa hasil pengelompokan memiliki kualitas yang cukup baik sehingga digunakan sebagai parameter pada proses segmentasi pelanggan menggunakan algoritma K-Mea[2]

6. Implementasi Sistem

Implementasi sistem dilakukan dengan membangun aplikasi berbasis web menggunakan React sebagai antarmuka pengguna, Supabase sebagai basis data, dan Vercel sebagai media deployment[9]. Data hasil preprocessing disimpan pada basis data Supabase dan ditampilkan pada halaman utama aplikasi. Sistem menyediakan beberapa fitur utama, yaitu Seed dari CSV untuk mengimpor data pelanggan, Jalankan Analisis untuk menjalankan proses segmentasi menggunakan algoritma K-Means, Filter Segmentasi untuk menampilkan data berdasarkan kelompok pelanggan, serta Tambah Baru untuk menambahkan data pelanggan ke dalam basis data. Hasil segmentasi disimpan kembali ke database dan ditampilkan secara langsung pada dashboard aplikasi sehingga pengguna dapat melakukan analisis pelanggan secara lebih mudah dan interaktif.

7. Pengujian Sistem

Pengujian sistem dilakukan untuk memastikan seluruh fungsi aplikasi berjalan sesuai dengan kebutuhan pengguna. Pengujian meliputi proses impor data pelanggan dari file CSV, penyimpanan data ke basis data, pelaksanaan algoritma K-Means, penyimpanan hasil segmentasi, serta penampilan hasil segmentasi pada antarmuka aplikasi. Selain pengujian fungsional, kualitas hasil clustering dievaluasi menggunakan Silhouette Score untuk mengetahui tingkat kualitas pemisahan antarcluster yang dihasilkan [9].

III. HASIL DAN PEMBAHASAN

1. Pengumpulan Data

Pada penelitian ini digunakan Marketing Campaign Dataset yang diperoleh dari platform Kaggle. Dataset tersebut berisi informasi mengenai karakteristik pelanggan yang meliputi ID, Year_Birth, Marital_Status, Education, Income, Kidhome, Teenhome, Dt_Customer, Recency, MntWines, MntFruits, MntMeatProducts, MntFishProducts, MntSweetProducts, MntGoldProds, NumDealsPurchases, NumWebPurchases, NumCatalogPurchases,

NumStorePurchases, NumWebVisitsMonth, AcceptedCmp1, AcceptedCmp2, AcceptedCmp3, AcceptedCmp4, AcceptedCmp5, Complain, Z_CostContact, Z_Revenue, Response. Dataset ini dipilih karena memiliki atribut yang lengkap sehingga dapat digunakan untuk melakukan segmentasi pelanggan menggunakan algoritma K-Means Clustering.

Tabel 1. Atribut Dataset

Atribut Data	Type data	Keterangan
ID	Integer	Identitas unik pelanggan
Year_Birth	Integer	Tahun lahir pelanggan
Education	Object	Tingkat pendidikan pelanggan
Marital_Status	Object	Status Pernikahan pelanggan
Income	Float	Pendapatan tahunan keluarga pelanggan
Kidhome	Integer	Jumlah anak dalam keluarga
Teenhome	Integer	Jumlah remaja dalam keluarga
Dt_Customer	Object	Tanggal bergabung pelanggan pada perusahaan
Recency	Integer	Jumlah hari sejak pembelian terakhir
MntWines	Integer	Jumlah pembelian anggur dalam 2 tahun terakhir
MntFruits	Integer	Jumlah pembelian buah dalam 2 tahun terakhir
MntMeatProducts	Integer	Jumlah pembelian daging dalam 2 tahun terakhir
MntFishProducts	Integer	Jumlah pembelian ikan dalam 2 tahun terakhir
MntSweetProducts	Integer	Jumlah pembelian manisan dalam 2 tahun terakhir
MntGoldProds	Integer	Jumlah pembelian emas dalam 2 tahun terakhir
NumDealsPurchases	Integer	Jumlah pembelian menggunakan potongan harga
NumWebPurchases	Integer	Jumlah pembelian dalam website perusahaan
NumCatalogPurchases	Integer	Jumlah pembelian menggunakan katalog
NumStorePurchases	Integer	Jumlah pembelian dari toko
NumWebVisitsMonth	Integer	Jumlah kunjungan pada website perusahaan
AcceptedCmp1	Integer	Bernilai 1 jika pelanggan menerima promosi pertama, 0 jika tidak
AcceptedCmp2	Integer	Bernilai 1 jika pelanggan menerima promosi kedua, 0 jika tidak
AcceptedCmp3	Integer	Bernilai 1 jika pelanggan menerima promosi ketiga, 0 jika tidak
AcceptedCmp4	Integer	Bernilai 1 jika pelanggan menerima promosi keempat, 0 jika tidak
AcceptedCmp5	Integer	Bernilai 1 jika pelanggan menerima promosi kelima, 0 jika tidak
Complain	Integer	1 jika pelanggan melakukan komplain dalam 2 tahun terakhir, 0 jika tidak

Z_CostContact	Integer	biaya yang terkait dengan menghubungi pelanggan
Z_Revenue	Integer	pendapatan dari pelanggan
Response	Integer	1 jika pelanggan menerima promo terakhir, 0 jika tidak

Dataset yang berhasil diperoleh terdiri dari 2.240 data pelanggan dengan 29 atribut. Sebelum digunakan pada proses analisis, dataset terlebih dahulu melalui tahap data cleaning untuk memastikan kualitas data yang digunakan.

Tabel 2. Informasi Dataset

Keterangan	Nilai
Nama Dataset	Marketing Campaign
Sumber	Kaggle
Jumlah Data Awal	2.240
Jumlah Atribut	29

2. Hasil Preprocessing Data

Tahap preprocessing dilakukan untuk meningkatkan kualitas data sebelum proses clustering. Dataset awal berjumlah 2.240 data pelanggan. Pada tahap ini dilakukan penghapusan data yang memiliki missing value pada atribut Income serta menghilangkan outlier dengan nilai pendapatan lebih dari 600.000. Setelah proses data cleaning, jumlah data yang digunakan menjadi 2.215 data.

Tabel 3. Hasil Data Cleaning

Tahapan	Jumlah Data
Dataset awal	2.240
Missing value dan outlier dihapus	25
Dataset akhir	2.215

Setelah proses data cleaning selesai, dilakukan pembentukan atribut baru (feature engineering) meliputi Age, Spending, Seniority, Has_child, dan Children. Selanjutnya atribut numerik dinormalisasi menggunakan metode Min-Max Scaling sehingga seluruh atribut memiliki rentang nilai yang sama sebelum diproses menggunakan algoritma K-Means.

3. Hasil Feature Engineering

Setelah proses data cleaning selesai, dilakukan tahap feature engineering untuk membentuk atribut baru yang lebih representatif terhadap karakteristik pelanggan [2]. Tujuan tahap ini adalah meningkatkan kualitas data sehingga proses

segmentasi menggunakan algoritma K-Means dapat menghasilkan pengelompokan yang lebih baik.

Pada penelitian ini dibentuk beberapa atribut baru, yaitu Age yang diperoleh dari konversi atribut Year_Birth, Spending yang merupakan total pengeluaran pelanggan dari kategori Wines, Fruits, Meat, Fish, Sweets, dan Gold, serta Seniority yang menunjukkan lama pelanggan bergabung dengan perusahaan. Selain itu, atribut kategorikal juga disederhanakan menjadi Education, Marital_Status, Has_child, dan Children agar lebih mudah digunakan pada proses analisis.

Hasil feature engineering menghasilkan 14 atribut utama yang digunakan pada proses segmentasi pelanggan, yaitu Age, Education, Marital_Status, Income, Spending, Seniority, Has_child, Children, Wines, Fruits, Meat, Fish, Sweets, dan Gold.

	Age	Education	Marital_Status	Income	Spending	seniority
0	57	Postgraduate	Alone	58138.0	1617	25.333333
1	60	Postgraduate	Alone	46344.0	27	7.000000
2	49	Postgraduate	In couple	71613.0	776	13.633333
3	30	Postgraduate	In couple	26646.0	53	7.866667
4	33	Postgraduate	In couple	58293.0	422	8.600000

Gambar 1. Hasil Feature Engineering

4. Hasil Seleksi dan Normalisasi Data

Setelah proses feature engineering, dilakukan seleksi atribut yang akan digunakan dalam proses segmentasi pelanggan. Atribut yang dipilih meliputi Age, Education, Marital_Status, Income, Spending, Seniority, Has_child, Children, Wines, Fruits, Meat, Fish, Sweets, dan Gold. Atribut-atribut tersebut dipilih karena mampu merepresentasikan karakteristik demografi serta perilaku belanja pelanggan [12].

Selanjutnya dilakukan proses normalisasi terhadap atribut numerik menggunakan metode Min-Max Scaling sebelum data diproses menggunakan algoritma K-Means. Normalisasi bertujuan untuk menyamakan rentang nilai setiap atribut sehingga tidak terdapat atribut tertentu yang mendominasi proses pengelompokan. Persamaan normalisasi Min-Max Scaling ditunjukkan pada Persamaan (1).

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

dengan:

x = nilai awal atribut;
 x_{min} = nilai minimum atribut;
 x_{max} = nilai maksimum atribut;
 x' = hasil normalisasi.

Persamaan (1) digunakan untuk mengubah seluruh atribut numerik ke dalam rentang nilai 0–1 sehingga setiap atribut memiliki kontribusi yang seimbang pada proses perhitungan jarak dalam algoritma K-Means Clustering.

Proses normalisasi bertujuan untuk menyamakan rentang nilai setiap atribut sehingga tidak terjadi dominasi atribut tertentu dalam perhitungan jarak antar data. Data hasil normalisasi kemudian digunakan sebagai masukan pada tahap penentuan jumlah cluster optimal menggunakan Elbow Method dan Silhouette Score.

5. Hasil Penentuan Jumlah Cluster Optimal

Penentuan jumlah cluster merupakan salah satu tahapan penting dalam algoritma K-Means karena nilai K yang dipilih akan memengaruhi kualitas hasil pengelompokan pelanggan. Pada penelitian ini, penentuan jumlah cluster dilakukan menggunakan Elbow Method, yaitu dengan menghitung nilai Within Cluster Sum of Squares (WCSS) pada beberapa variasi jumlah cluster. Metode ini bertujuan untuk memperoleh jumlah cluster yang memberikan keseimbangan antara tingkat homogenitas dalam cluster dan kompleksitas model. Nilai WCSS menunjukkan total jarak kuadrat antara setiap data dengan centroid pada cluster yang ditempatinya. Semakin kecil nilai WCSS menunjukkan bahwa anggota dalam suatu cluster semakin homogen. Persamaan WCSS ditunjukkan pada Persamaan (2).

$$WCSS = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (2)$$

dengan:

k = jumlah cluster;
 C_i = cluster ke- i ;
 x = data pada cluster;
 μ_i = centroid pada cluster ke- i .

Persamaan (2) digunakan untuk menghitung total variasi data terhadap centroid pada setiap cluster. Nilai WCSS kemudian diplot ke dalam

grafik Elbow Method sehingga diperoleh titik siku (elbow point) yang menunjukkan jumlah cluster optimal.

Setelah jumlah cluster optimal diperoleh menggunakan Elbow Method, kualitas hasil pengelompokan dievaluasi menggunakan Silhouette Score. Nilai Silhouette Score digunakan untuk mengukur tingkat kemiripan suatu data terhadap cluster tempat data tersebut berada dibandingkan dengan cluster lainnya. Persamaan Silhouette Score ditunjukkan pada Persamaan (3).

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (3)$$

dengan :

$s(i)$ = nilai Silhouette untuk data ke- i ;
 $a(i)$ = rata-rata jarak data ke anggota dalam cluster yang sama;
 $b(i)$ = rata-rata jarak data ke cluster terdekat.

Nilai Silhouette Score berada pada rentang -1 hingga 1. Semakin mendekati 1, semakin baik kualitas pemisahan antarcluster, sedangkan nilai yang mendekati 0 menunjukkan bahwa batas antarcluster kurang jelas.

Berdasarkan hasil analisis pada dataset yang telah melalui tahap preprocessing, sistem menghasilkan nilai K optimal sebesar 4[13]. Nilai tersebut ditunjukkan oleh terbentuknya titik siku (elbow point) pada grafik Elbow Method. Selanjutnya dilakukan evaluasi menggunakan Silhouette Score dan diperoleh nilai sebesar 0,3723 [2].

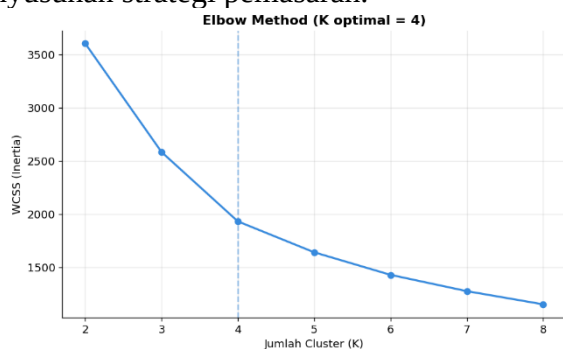
Tabel 4. Hasil Penentuan Jumlah Cluster

Parameter	Hasil
Jumlah Data	2.215
Metode	Elbow Method
K Optimal	4
Silhouette Score	0,3723

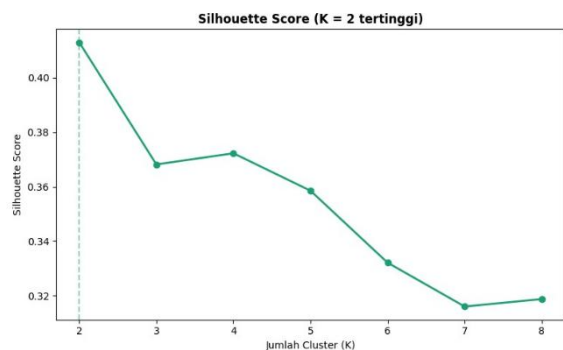
Hasil tersebut menunjukkan bahwa dataset pelanggan dapat dikelompokkan menjadi empat kelompok utama berdasarkan karakteristik demografi dan perilaku belanja pelanggan. Nilai Silhouette Score sebesar 0,3723 menunjukkan bahwa kualitas pemisahan antarcluster berada pada kategori cukup, sehingga setiap kelompok pelanggan telah memiliki karakteristik yang berbeda meskipun masih terdapat sebagian data yang memiliki kemiripan dengan cluster lain [8],

[14]. Kondisi tersebut umum dijumpai pada dataset pelanggan karena perilaku konsumen cenderung memiliki karakteristik yang saling beririsan.

Penggunaan kombinasi Elbow Method dan Silhouette Score memberikan hasil yang lebih objektif dibandingkan menentukan jumlah cluster secara manual[13]. Dengan demikian, proses segmentasi pelanggan yang dilakukan pada penelitian ini mampu menghasilkan jumlah cluster yang sesuai dengan pola distribusi data sehingga dapat digunakan sebagai dasar dalam penyusunan strategi pemasaran.



Gambar 2. Hasil Penentuan Jumlah Cluster Menggunakan Elbow Method



Gambar 3. Hasil Grafik Silhouette Score

Hasil tersebut menunjukkan bahwa penggunaan kombinasi Elbow Method dan Silhouette Score mampu menghasilkan jumlah cluster yang lebih objektif dibandingkan menentukan nilai K secara manual. Dengan demikian, nilai K = 4 digunakan sebagai parameter pada proses segmentasi pelanggan menggunakan algoritma K-Means [2], [13].

6. Hasil segmentasi pelanggan Menggunakan Algoritma K-Means

Setelah diperoleh jumlah cluster optimal sebanyak empat cluster. Algoritma ini bekerja dengan menghitung jarak antara setiap data pelanggan dengan centroid pada masing-masing cluster menggunakan metode Euclidean Distance. Setiap data akan ditempatkan pada cluster yang memiliki jarak paling dekat terhadap centroid. Persamaan Euclidean Distance ditunjukkan pada Persamaan (4).

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (4)$$

dengan :

$d(x, y)$ = jarak euclidean antara data dan centroid;

x_i = nilai atribut ke- i pada data pelanggan;

y_i = nilai atribut ke- i pada centroid;

n = jumlah atribut yang digunakan dalam proses clustering.

$$C_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_j \quad (5)$$

dengan:

C_i = centroid pada cluster ke- i ;

n_i = jumlah data pada cluster ke- i ;

x_j = data anggota pada cluster ke- i .

Persamaan (5) digunakan untuk menghitung centroid baru berdasarkan rata-rata seluruh anggota pada setiap cluster. Nilai centroid yang baru kemudian digunakan kembali pada proses perhitungan jarak hingga hasil clustering konvergen.

Berdasarkan hasil clustering, seluruh pelanggan berhasil dikelompokkan ke dalam empat cluster dengan distribusi sebagaimana ditunjukkan pada Tabel 5.

Tabel 5. Distribusi Hasil Segmentasi Pelanggan

Cluster	Jumlah Pelanggan	Persentase
Cluster 0	441	19,9%
Cluster 1	479	21,6%
Cluster 2	653	29,5%
Cluster 3	642	29,0%
Total	2.215	100%

Selanjutnya dilakukan analisis terhadap karakteristik rata-rata pada masing-masing cluster berdasarkan atribut Income, Spending, Seniority,

dan Age. Hasil analisis tersebut disajikan pada Tabel 6.

Tabel 6 . Profil Rata-rata Tiap Cluster

Cluster	Income	Spending	Seniority	Age
0	73.875,4	1.105,3	9,4	47,3
1	71.023,9	1.316,3	21,2	47,1
2	38.511,8	131,5	9,0	44,0
3	36.395,0	220,3	20,5	43,5

Perbedaan karakteristik pada masing-masing cluster menunjukkan bahwa pelanggan memiliki pola pendapatan, tingkat pengeluaran, usia, dan lama bergabung yang beragam. Informasi tersebut menjadi dasar bagi perusahaan untuk menentukan strategi pemasaran yang sesuai dengan karakteristik setiap kelompok pelanggan [5].

a. Cluster 0

Cluster 0 Cluster 0 terdiri atas 441 pelanggan (19,9%) dengan rata-rata pendapatan sebesar 73.875,4, rata-rata pengeluaran (Spending) sebesar 1.105,3, lama menjadi pelanggan (Seniority) sekitar 9,4 bulan, serta rata-rata usia 47,3 tahun. Karakteristik tersebut menunjukkan bahwa pelanggan pada cluster ini memiliki pendapatan dan tingkat pengeluaran yang tinggi, namun masih tergolong pelanggan yang relatif baru. Oleh karena itu, strategi yang sesuai adalah memberikan program loyalitas, penawaran produk premium, serta promosi yang bertujuan meningkatkan retensi pelanggan.

b. Cluster 1

Cluster 1 terdiri atas 479 pelanggan (21,6%) dengan rata-rata pendapatan sebesar 71.023,9 dan rata-rata pengeluaran sebesar 1.316,3, yang merupakan nilai pengeluaran tertinggi dibandingkan cluster lainnya. Selain itu, pelanggan pada cluster ini memiliki rata-rata lama bergabung selama 21,2 bulan, sehingga dapat dikategorikan sebagai pelanggan yang loyal dan bernilai tinggi. Strategi yang direkomendasikan adalah mempertahankan loyalitas pelanggan melalui pemberian reward, layanan eksklusif, serta penawaran produk premium.

c. Cluster 2

Cluster 2 merupakan kelompok terbesar dengan jumlah 653 pelanggan (29,5%). Kelompok ini memiliki rata-rata pendapatan sebesar 38.511,8 dan rata-rata pengeluaran sebesar 131,5, yang menunjukkan bahwa pelanggan memiliki

daya beli relatif rendah. Lama menjadi pelanggan sekitar 9 bulan mengindikasikan bahwa sebagian besar pelanggan masih berada pada tahap awal hubungan dengan perusahaan. Oleh karena itu, strategi yang dapat diterapkan adalah memberikan promosi entry-level, diskon, atau voucher untuk meningkatkan minat pembelian.

d. Cluster 3

Cluster 3 terdiri atas 642 pelanggan (29,0%) dengan rata-rata pendapatan sebesar 36.395,0 dan rata-rata pengeluaran sebesar 220,3. Meskipun memiliki tingkat pendapatan dan pengeluaran yang relatif rendah, pelanggan pada cluster ini memiliki rata-rata lama bergabung selama 20,5 bulan, sehingga menunjukkan bahwa mereka merupakan pelanggan lama dengan aktivitas pembelian yang masih rendah. Strategi yang dapat diterapkan adalah melakukan program reaktivasi pelanggan, memberikan penawaran khusus, serta menerapkan strategi cross-selling untuk meningkatkan nilai transaksi.

Secara keseluruhan, hasil segmentasi menunjukkan bahwa algoritma K-Means Clustering berhasil mengelompokkan pelanggan ke dalam empat segmen yang memiliki karakteristik berbeda. Cluster 0 dan Cluster 1 didominasi oleh pelanggan dengan pendapatan dan tingkat pengeluaran yang tinggi sehingga berpotensi menjadi pelanggan bernilai tinggi (high-value customer). Sebaliknya, Cluster 2 dan Cluster 3 memiliki tingkat pendapatan dan pengeluaran yang lebih rendah sehingga memerlukan strategi pemasaran yang berfokus pada peningkatan aktivitas pembelian. Dengan adanya segmentasi ini, perusahaan dapat menyusun strategi pemasaran yang lebih efektif, efisien, dan sesuai dengan karakteristik masing-masing kelompok pelanggan[1], [15].

7. Implementasi Sistem Berbasis Website

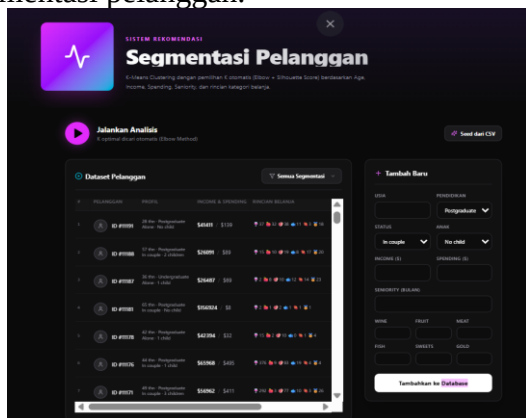
Hasil segmentasi pelanggan kemudian diimplementasikan ke dalam sebuah aplikasi berbasis web yang dikembangkan menggunakan React sebagai antarmuka pengguna (frontend), Supabase sebagai basis data (backend database), serta Vercel sebagai media deployment. Implementasi ini bertujuan untuk mempermudah pengguna dalam melakukan proses segmentasi pelanggan tanpa harus menjalankan proses

analisis secara manual menggunakan perangkat lunak data mining.

Sistem dirancang agar seluruh proses, mulai dari pengelolaan data pelanggan, pelaksanaan algoritma K-Means, hingga penyajian hasil segmentasi dapat dilakukan melalui satu aplikasi yang terintegrasi. Dengan demikian, pengguna hanya perlu mengunggah atau menambahkan data pelanggan, kemudian sistem akan secara otomatis melakukan proses analisis dan menampilkan hasil segmentasi pada dashboard.

a. Halaman Dashboard

Halaman dashboard merupakan tampilan utama yang digunakan pengguna untuk mengakses seluruh fungsi sistem. Pada halaman ini ditampilkan informasi mengenai jumlah data pelanggan, menu navigasi, serta beberapa fitur utama yang dapat digunakan dalam proses segmentasi pelanggan.



Gambar 4. Halaman Dashboard Sistem

Berdasarkan Gambar 4, dashboard menyediakan beberapa menu utama, yaitu:

- Dashboard, untuk menampilkan ringkasan informasi sistem.
- Pelanggan, untuk menampilkan seluruh data pelanggan yang tersimpan pada basis data.
- Tambah Baru, digunakan untuk menambahkan data pelanggan secara manual.
- Seed dari CSV, digunakan untuk mengimpor dataset pelanggan dalam format CSV ke dalam basis data Supabase.
- Jalankan Analisis, digunakan untuk menjalankan proses preprocessing, penentuan jumlah cluster, dan algoritma K-Means secara otomatis.

Keberadaan fitur-fitur tersebut memungkinkan seluruh proses segmentasi dilakukan secara terintegrasi sehingga pengguna tidak perlu melakukan pengolahan data menggunakan aplikasi lain. Hasil implementasi menunjukkan bahwa aplikasi mampu mengelola sebanyak 2.215 data pelanggan dan menampilkan hasil segmentasi ke dalam empat cluster sesuai hasil analisis pada penelitian ini. Selain itu, sistem juga menyediakan fasilitas penyaringan (*filter*) berdasarkan cluster sehingga pengguna dapat melihat karakteristik pelanggan pada setiap kelompok secara lebih mudah [9].

Implementasi sistem berbasis web memberikan kemudahan bagi pengguna karena seluruh proses analisis dilakukan secara otomatis melalui antarmuka aplikasi tanpa memerlukan pengetahuan mengenai pemrograman maupun proses data mining. Dengan demikian, aplikasi dapat dimanfaatkan sebagai media pendukung dalam proses pengambilan keputusan, khususnya untuk menentukan strategi pemasaran berdasarkan karakteristik masing-masing segmen pelanggan[6].

8. Pengujian Sistem

Pengujian sistem dilakukan untuk memastikan bahwa seluruh fungsi aplikasi berjalan sesuai dengan kebutuhan pengguna. Pengujian dilakukan terhadap setiap fitur utama aplikasi, mulai dari proses impor dataset, pengelolaan data pelanggan, pelaksanaan algoritma K-Means Clustering, hingga penyajian hasil segmentasi.

Tabel 7. Hasil Pengujian Sistem

No	Skenario Pengujian	Hasil yang Di-harapkan	Hasil
1	Menampilkan dashboard	Dashboard berhasil ditampilkan	Berhasil
2	Menampilkan data pelanggan	Data pelanggan muncul pada tabel	Berhasil
3	Menambahkan data pelanggan	Data berhasil tersimpan	Berhasil
4	Mengimpor dataset	Dataset berhasil masuk ke database	Berhasil
5	Menjalankan preprocessing	Data berhasil diproses	Berhasil
6	Menentukan jumlah cluster	Nilai K optimal berhasil ditentukan	Berhasil
7	Menjalankan algoritma K-Means	Data berhasil dikelompokkan	Berhasil

8	Menghitung Silhouette Score	Nilai evaluasi berhasil ditampilkan	Berhasil
9	Menampilkan hasil segmentasi	Dashboard analisis berhasil ditampilkan	Berhasil
10	Melakukan filter berdasarkan cluster	Data sesuai cluster berhasil ditampilkan	Berhasil

Berdasarkan hasil pengujian pada Tabel 7, seluruh fungsi sistem memperoleh status berhasil. Hal ini menunjukkan bahwa integrasi antara React, Supabase, dan algoritma K-Means telah berjalan dengan baik. Sistem mampu mengelola data pelanggan, melakukan proses segmentasi secara otomatis, serta menyajikan hasil analisis dalam bentuk dashboard interaktif yang mudah dipahami oleh pengguna.

9. Pembahasan Hasil Penelitian

Penelitian ini berhasil mengimplementasikan algoritma K-Means Clustering untuk melakukan segmentasi pelanggan menggunakan Marketing Campaign Dataset yang terdiri atas 2.215 data pelanggan setelah melalui tahap preprocessing. Proses penentuan jumlah cluster menggunakan Elbow Method menghasilkan jumlah cluster optimal sebanyak 4 cluster, sedangkan evaluasi menggunakan Silhouette Score memperoleh nilai sebesar 0,3723, yang menunjukkan bahwa hasil pengelompokan memiliki kualitas yang cukup baik.

Hasil segmentasi menunjukkan bahwa setiap cluster memiliki karakteristik pelanggan yang berbeda berdasarkan atribut Age, Income, Spending, Seniority, Education, Marital Status, dan Children. Perbedaan karakteristik tersebut dapat dimanfaatkan sebagai dasar dalam penyusunan strategi pemasaran yang lebih tepat sasaran, seperti pemberian program loyalitas bagi pelanggan bernilai tinggi maupun promosi khusus bagi pelanggan dengan tingkat pengeluaran yang lebih rendah.

Selain itu, penelitian ini juga berhasil mengimplementasikan hasil segmentasi ke dalam aplikasi berbasis web menggunakan React, Supabase, dan Vercel. Aplikasi yang dikembangkan mampu mengelola data pelanggan, menjalankan proses segmentasi secara otomatis, serta menampilkan hasil analisis dalam bentuk dashboard interaktif sehingga mempermudah pengguna dalam

melakukan analisis pelanggan tanpa harus menjalankan proses data mining secara manual.

Meskipun demikian, penelitian ini masih memiliki keterbatasan karena menggunakan satu sumber dataset dan menghasilkan nilai Silhouette Score yang menunjukkan bahwa masih terdapat kemiripan karakteristik antarcluster. Oleh karena itu, penelitian selanjutnya dapat menggunakan dataset yang lebih beragam, menambahkan atribut perilaku pelanggan, atau membandingkan performa algoritma K-Means dengan metode clustering lain seperti DBSCAN, Agglomerative Clustering, maupun Gaussian Mixture Model (GMM) untuk memperoleh hasil segmentasi yang lebih optimal.

IV. KESIMPULAN

Penelitian ini berhasil mengimplementasikan algoritma K-Means Clustering untuk melakukan segmentasi pelanggan menggunakan Marketing Campaign Dataset yang terdiri atas 2.215 data pelanggan setelah melalui proses preprocessing. Tahapan preprocessing yang meliputi data cleaning, feature engineering, seleksi atribut, dan normalisasi data mampu meningkatkan kualitas data sehingga siap digunakan dalam proses clustering. Penentuan jumlah cluster menggunakan Elbow Method menghasilkan jumlah cluster optimal sebanyak 4 cluster ($K = 4$), sedangkan evaluasi menggunakan Silhouette Score memperoleh nilai sebesar 0,3723, yang menunjukkan bahwa hasil pengelompokan memiliki kualitas yang cukup baik dalam membedakan karakteristik pelanggan.

Hasil segmentasi menunjukkan bahwa setiap cluster memiliki karakteristik yang berbeda berdasarkan atribut Age, Income, Spending, Seniority, Education, Marital Status, dan Children. Perbedaan karakteristik tersebut memungkinkan perusahaan menyusun strategi pemasaran yang lebih tepat sasaran, seperti pemberian program loyalitas kepada pelanggan bernilai tinggi, promosi bagi pelanggan baru, maupun penawaran diskon kepada pelanggan dengan tingkat pengeluaran yang rendah. Dengan demikian, hasil segmentasi dapat mendukung perusahaan dalam meningkatkan efektivitas pemasaran serta pengambilan keputusan berbasis data.

Selain menghasilkan model segmentasi, penelitian ini juga berhasil mengimplementasikan algoritma K-Means Clustering ke dalam aplikasi berbasis web menggunakan React, Supabase, dan Vercel. Sistem yang dikembangkan mampu mengelola data pelanggan, menjalankan proses segmentasi secara otomatis, serta menampilkan hasil analisis melalui dashboard interaktif. Berdasarkan hasil implementasi dan pengujian sistem, seluruh fitur utama aplikasi dapat berjalan sesuai dengan fungsinya sehingga mampu membantu pengguna dalam melakukan analisis pelanggan secara lebih mudah, cepat, dan efisien.

REFERENSI

- [1] N. Zai, A. A. Kenap, and A. Hasibuan, "Perbandingan Clustering Berbasis RFM dan Implementasi K-Means untuk Segmentasi Pelanggan Bisnis Laundry," *semanTIK*, vol. 11, no. 1, pp. 22–32, 2026.
- [2] 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.
- [3] 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.
- [4] C. Ismaili and M. El Hamzaoui, "Data-Driven Customer Segmentation Using K-Means and PCA: Leveraging Meteorological and Behavioral," *International Journal of Computing*, vol. 25, no. 1, pp. 56–64, Mar. 2026, doi: 10.47839/ijc.25.14488.
- [5] O. Akande, E. O. Asani, and B. Dautare, "Customer Segmentation Through RFM Analysis and K-Means Clustering: Leveraging Data-Driven Insights for Effective Marketing Strategy," *Ceddi Journal of Information System and Technology (JST)*, vol. 3, no. 1, pp. 14–25, Apr. 2024, doi: 10.56134/jst.v3i1.81.
- [6] Fathoni Dwi Atmoko, "Penerapan K-Means Clustering untuk Segmentasi Pelanggan," *Uranus : Jurnal Ilmiah Teknik Elektro, Sains dan Informatika*, vol. 3, no. 2, pp. 174–191, Jun. 2025, doi: 10.61132/uranus.v3i2.1214.
- [7] 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)," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, Aug. 2024, doi: 10.23960/jitet.v12i3.4677.
- [8] A. M. Artiarno, P. Setiaji, and F. Nugraha, "K-Means Clustering untuk Segmentasi Pelanggan: Mengungkap Pola Pembelian Strategi Pemasaran pada Sektor Ritel," *Edumatic: Jurnal Pendidikan Informatika*, vol. 9, no. 2, pp. 442–451, Aug. 2025, doi: 10.29408/edumatic.v9i2.30336.
- [9] A. H. Fazri, A. Muhammad, P. Ayu, and W. Purnama, "Computer Based Information System Journal ALGORITMA K-MEANS CLUSTER UNTUK SEGMENTASI PELANGGAN KATA KUNCI CRM, K-means cluster, Customer Segmentation. KORESPONDENSI," *CBIS JOURNAL*, vol. 11, no. 02, 2023, [Online]. Available: <http://ejournal.upbatam.ac.id/index.php/cbishttp://ejournal.upbatam.ac.id/index.php/cbis>
- [10] C. Hafidz Ardana *et al.*, "Segmentasi Pelanggan Penjualan Online Menggunakan Metode K-means Clustering," 2024.
- [11] F. Nazihah, A. Danniswara, and A. Wibowo, "Analisis Segmentasi Pelanggan dengan Algoritma K-Means pada Data Penjualan," *Jurnal Algoritma*, vol. 22, no. 2, Nov. 2025, doi: 10.33364/algoritma/v.22-2.2489.
- [12] Baiq Nikum Yulisasih, H. Herman, and S. Sunardi, "K-Means Clustering Method For Customer Segmentation Based On Potential Purchases," *Jurnal ELTIKOM*, vol. 8, no. 1, pp. 83–90, Jun. 2024, doi: 10.31961/eltikom.v8i1.1137.
- [13] 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.
- [14] 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.
- [15] I. YUNITA, P. R. Ali, M. A. Kartawidjaja, and R. Sukwadi, "Segmentasi Pelanggan Menggunakan K-Means Clustering: Menganalisis Metrik RFM untuk Strategi Pemasaran," *Jurnal Media Teknik dan Sistem Industri*, vol. 9, no. 1, p. 58, Mar. 2025, doi: 10.35194/jmtsi.v9i1.4452.