

ANALISIS SEGMENTASI PELANGGAN MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING BERDASARKAN FAKTOR UMUR, PENDAPATAN, DAN SKOR PENGELUARAN

Dandi Yohananda Saputra Utama¹, Muawan Bisri²

^{1,2}Universitas Indonesia Mandiri

Email: dandivohanandasaputra@uimandiri.ac.id, muawanbisri@uimandiri.ac.id

Abstrak

Penelitian ini bertujuan untuk menganalisis segmentasi pelanggan menggunakan algoritma K-Means Clustering berdasarkan tiga faktor utama, yaitu umur, pendapatan tahunan, dan skor pengeluaran. Data yang digunakan berasal dari *Mall Customers Dataset* yang terdiri dari 200 data pelanggan. Tahapan penelitian meliputi *data preprocessing*, *normalisasi data*, penerapan algoritma K-Means, serta evaluasi menggunakan *Elbow Method* untuk menentukan jumlah cluster optimal.

Hasil penelitian menunjukkan bahwa jumlah cluster optimal adalah empat ($k = 4$). Setiap cluster memiliki karakteristik yang berbeda, yaitu: Cluster 0 pelanggan berusia dewasa dengan pendapatan sedang dan pola pengeluaran moderat, Cluster 1 pelanggan muda produktif dengan pendapatan tinggi dan pengeluaran tinggi, Cluster 2 pelanggan muda berpendapatan menengah dengan pengeluaran cukup tinggi, dan Cluster 3 pelanggan berpendapatan tinggi tetapi pengeluaran rendah. Variabel *Annual Income* dan *Spending Score* terbukti paling berpengaruh dalam pembentukan cluster.

Penelitian ini menunjukkan bahwa algoritma K-Means efektif untuk mengidentifikasi pola perilaku pelanggan dan membantu perusahaan dalam merancang strategi pemasaran yang lebih tepat sasaran. Hasilnya dapat dimanfaatkan untuk meningkatkan loyalitas pelanggan dan mendukung pengambilan keputusan berbasis data.

Kata kunci: Data mining, K-Means clustering, segmentasi pelanggan, pendapatan, skor pengeluaran.

Abstract

This study aims to analyze customer segmentation using the K-Means Clustering algorithm based on three main factors: age, annual income, and spending score. The dataset used is the *Mall Customers Dataset* containing 200 customer records. The research stages include data preprocessing, normalization, implementation of the K-Means algorithm, and evaluation using the *Elbow Method* to determine the optimal number of clusters.

The results indicate that the optimal number of clusters is four ($k = 4$). Each cluster has distinct characteristics: Cluster 0 represents middle-aged customers with moderate income and spending, Cluster 1 includes young, high-income, and high-spending customers, Cluster 2 consists of younger customers with moderate income but active spending, and Cluster 3 represents high-income customers with low spending levels. The *Annual Income* and *Spending Score* variables are the most influential in cluster formation.

This study demonstrates that K-Means Clustering is effective for identifying customer behavior patterns and supporting businesses in designing more targeted marketing strategies. The findings can help enhance customer loyalty and facilitate data-driven decision-making.

Keywords: Data mining, K-Means clustering, customer segmentation, income, spending score.

1. Pendahuluan

Perkembangan teknologi informasi telah mendorong perusahaan untuk memanfaatkan data pelanggan sebagai dasar pengambilan keputusan yang lebih tepat dan efisien. Data pelanggan yang tersimpan dalam sistem informasi perusahaan mengandung berbagai informasi penting, seperti usia, jenis kelamin, tingkat pendapatan, hingga pola pengeluaran. Informasi ini dapat diolah untuk memahami perilaku dan karakteristik pelanggan, sehingga perusahaan dapat merancang strategi pemasaran yang lebih efektif dan terarah. Namun, banyaknya jumlah data yang dihasilkan setiap hari sering kali membuat proses analisis menjadi kompleks dan sulit dilakukan secara manual.

Salah satu pendekatan yang dapat digunakan untuk memahami pola perilaku pelanggan adalah melalui segmentasi pelanggan (*customer segmentation*). Segmentasi ini berfungsi untuk mengelompokkan pelanggan ke dalam beberapa kategori berdasarkan kesamaan karakteristik tertentu. Dengan melakukan segmentasi, perusahaan dapat mengetahui kelompok pelanggan potensial, pelanggan loyal, maupun pelanggan dengan daya beli rendah. Hasil segmentasi ini menjadi dasar penting dalam pengambilan keputusan, seperti menentukan strategi promosi, menetapkan harga, hingga melakukan pengembangan produk.

Dalam bidang analisis data, data mining menjadi teknik yang banyak digunakan untuk menggali pengetahuan tersembunyi dari kumpulan data besar (*large datasets*). Salah satu metode dalam data mining yang populer digunakan untuk segmentasi pelanggan adalah algoritma K-Means Clustering. K-Means merupakan algoritma *unsupervised learning* yang berfungsi mengelompokkan data ke dalam sejumlah *cluster* berdasarkan kemiripan nilai antar data. Algoritma ini bekerja dengan mencari pusat cluster (*centroid*) dan mengelompokkan data yang memiliki jarak paling dekat dengan centroid tersebut. Kelebihan K-Means terletak pada kesederhanaan perhitungannya, efisiensi dalam menangani data berukuran besar, serta kemampuannya dalam memberikan hasil pengelompokan yang jelas dan mudah diinterpretasikan.

Penelitian ini menggunakan lima variabel utama yaitu CustomerID, Gender, Age, Annual Income (k\$), dan Spending Score (1–100) untuk menganalisis perilaku pelanggan. Variabel *Age* dan *Annual Income* digunakan untuk menggambarkan kondisi demografis dan ekonomi pelanggan, sedangkan *Spending Score* menunjukkan tingkat pengeluaran yang mencerminkan perilaku konsumsi. Melalui penerapan algoritma K-Means Clustering, penelitian ini bertujuan untuk mengidentifikasi kelompok pelanggan yang memiliki karakteristik serupa berdasarkan kombinasi variabel tersebut.

Dengan demikian, hasil dari penelitian ini diharapkan dapat membantu perusahaan dalam memahami pola perilaku pelanggan serta mendukung pengambilan keputusan strategis di bidang pemasaran. Selain itu, penelitian ini juga diharapkan dapat memberikan kontribusi dalam penerapan teknik *data mining* untuk analisis perilaku konsumen menggunakan pendekatan *unsupervised learning* yang efisien dan mudah diterapkan.

2. Tinjauan Pustaka

2.1 Data Mining

Data mining merupakan proses penggalian informasi berharga dari kumpulan data yang besar dengan menggunakan metode statistik, matematika, dan kecerdasan buatan. Menurut Han, Kamber, dan Pei (2012), data mining didefinisikan sebagai proses penemuan pola yang menarik dan pengetahuan tersembunyi dari sejumlah besar data. Tujuan utama dari data mining

adalah memperoleh wawasan baru yang dapat digunakan dalam pengambilan keputusan yang lebih baik.

Dalam konteks bisnis, data mining menjadi alat penting dalam memahami perilaku pelanggan, mendeteksi pola pembelian, dan memprediksi tren pasar. Proses data mining umumnya terdiri atas beberapa tahap, yaitu data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation, dan knowledge presentation (Han et al., 2012). Teknik-teknik dalam data mining dapat dibedakan menjadi dua kategori besar, yaitu supervised learning (misalnya classification dan regression) dan unsupervised learning (misalnya clustering dan association rules).

2.2 Segmentasi Pelanggan (Customer Segmentation)

Segmentasi pelanggan merupakan proses pengelompokan konsumen ke dalam beberapa segmen berdasarkan karakteristik tertentu seperti demografi, perilaku, psikografis, atau geografis (Kotler & Keller, 2016). Melalui segmentasi pelanggan, perusahaan dapat memahami kebutuhan dan preferensi setiap kelompok sehingga strategi pemasaran dapat disesuaikan secara lebih efektif.

Dalam dunia bisnis modern, segmentasi pelanggan menjadi langkah awal dalam strategi customer relationship management (CRM). Analisis segmentasi yang tepat memungkinkan perusahaan untuk memfokuskan sumber daya pada kelompok pelanggan yang paling menguntungkan. Menurut Wedel dan Kamakura (2012), segmentasi berbasis data (data-driven segmentation) memberikan hasil yang lebih objektif dibandingkan pendekatan tradisional karena menggunakan teknik statistik dan algoritmik untuk menentukan kelompok pelanggan yang memiliki kesamaan karakteristik.

Variabel yang sering digunakan dalam segmentasi pelanggan antara lain:

- a. Demografi: usia, jenis kelamin, pendapatan.
- b. Psikografi: gaya hidup, nilai, dan kepribadian.
- c. Perilaku: tingkat pembelian, frekuensi pembelian, serta skor pengeluaran (spending score).

Dengan memanfaatkan variabel seperti Age, Annual Income, dan Spending Score, perusahaan dapat mengidentifikasi pelanggan potensial dan merancang strategi pemasaran yang lebih tepat sasaran.

2.3 Clustering

Clustering merupakan salah satu teknik unsupervised learning yang bertujuan mengelompokkan data ke dalam beberapa grup atau cluster sehingga data dalam satu cluster memiliki kemiripan yang tinggi, sedangkan data antar cluster memiliki perbedaan yang signifikan (Tan, Steinbach, & Kumar, 2018).

Teknik ini banyak digunakan dalam berbagai bidang seperti pemasaran, biologi, pengenalan pola, dan analisis sosial. Dalam konteks bisnis, clustering digunakan untuk mengelompokkan pelanggan berdasarkan kesamaan perilaku atau karakteristik tertentu. Proses clustering biasanya diawali dengan menentukan jumlah cluster, kemudian menghitung jarak antar data menggunakan ukuran seperti Euclidean distance atau Manhattan distance.

Hasil dari proses clustering berupa pembagian data ke dalam beberapa kelompok yang dapat digunakan sebagai dasar analisis lanjutan, seperti menentukan segmen pelanggan premium, menengah, dan ekonomi berdasarkan pendapatan dan pola pengeluaran.

2.4 Algoritma K-Means Clustering

K-Means Clustering merupakan salah satu algoritma unsupervised learning yang paling populer dalam analisis data. Algoritma ini bekerja dengan cara membagi dataset ke dalam K cluster yang berbeda berdasarkan kemiripan nilai antar data (MacQueen, 1967). Setiap cluster

memiliki titik pusat (centroid) yang mewakili rata-rata posisi seluruh data di dalamnya.

Proses kerja algoritma K-Means melibatkan dua tahap utama, yaitu:

- a. Assignment step: setiap data dimasukkan ke dalam cluster dengan jarak terdekat terhadap centroid.
- b. Update step: centroid setiap cluster diperbarui dengan menghitung rata-rata posisi semua anggota cluster.

Kedua tahap ini diulang hingga nilai centroid konvergen atau tidak mengalami perubahan signifikan.

Salah satu keunggulan algoritma K-Means adalah efisiensinya dalam menangani dataset berukuran besar dan hasil pengelompokan yang mudah diinterpretasikan. Namun, kelemahan K-Means terletak pada sensitivitas terhadap pemilihan nilai K dan posisi awal centroid (Jain, 2010). Oleh karena itu, metode seperti Elbow Method atau Silhouette Score sering digunakan untuk menentukan jumlah cluster yang optimal.

Dalam penelitian ini, algoritma K-Means digunakan untuk mengelompokkan pelanggan berdasarkan variabel Age, Annual Income, dan Spending Score, dengan tujuan untuk menemukan pola segmentasi yang dapat mendukung pengambilan keputusan bisnis, seperti menentukan strategi promosi dan penawaran produk yang lebih sesuai dengan karakteristik setiap segmen pelanggan.

2.5 Penelitian Terkait

Beberapa penelitian sebelumnya telah menggunakan algoritma K-Means untuk analisis segmentasi pelanggan.

Singh dan Chadha (2017) melakukan penelitian tentang segmentasi pelanggan menggunakan K-Means dan menemukan bahwa variabel Annual Income dan Spending Score efektif dalam membedakan perilaku konsumen.

Patel dan Patel (2019) menerapkan algoritma K-Means untuk menganalisis data pelanggan pada toko ritel dan menghasilkan empat kelompok pelanggan dengan karakteristik yang berbeda.

Sementara itu, Prasad dan Reddy (2020) mengembangkan pendekatan K-Means yang dikombinasikan dengan metode Principal Component Analysis (PCA) untuk meningkatkan akurasi segmentasi pelanggan.

Hasil penelitian-penelitian tersebut menunjukkan bahwa algoritma K-Means dapat digunakan secara efektif dalam memahami perilaku pelanggan dan mendukung pengambilan keputusan strategis perusahaan.

3. Metode Penelitian

3.1 Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen komputasional di bidang *data mining*. Tujuan utama dari penelitian ini adalah untuk mengelompokkan pelanggan berdasarkan karakteristik tertentu menggunakan algoritma K-Means Clustering.

Metode kuantitatif dipilih karena penelitian ini berfokus pada pengolahan data numerik untuk menemukan pola dan hubungan antar variabel secara objektif. Sementara itu, pendekatan *unsupervised learning* digunakan karena data tidak memiliki label atau kategori yang telah ditentukan sebelumnya.

3.2 Sumber dan Jenis Data

Data yang digunakan dalam penelitian ini merupakan data sekunder, yaitu data pelanggan yang diperoleh dari dataset publik Mall Customers Dataset yang banyak digunakan untuk analisis segmentasi pelanggan. Dataset ini memuat informasi pelanggan yang meliputi:

Variabel	Deskripsi	Tipe Data
----------	-----------	-----------

CustomerID	Nomor identifikasi unik pelanggan	Numerik
Gender	Jenis kelamin pelanggan (Male/Female)	Kategorikal
Age	Umur pelanggan (tahun)	Numerik
Annual Income (k\$)	Pendapatan tahunan pelanggan (ribu dolar)	Numerik
Spending Score (1–100)	Skor pengeluaran pelanggan berdasarkan kebiasaan belanja	Numerik

Jumlah total data yang digunakan sebanyak 200 entri pelanggan, dengan masing-masing kolom menggambarkan karakteristik yang akan menjadi dasar proses pengelompokan.

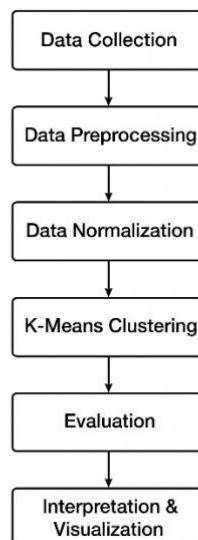
Data yang digunakan dalam penelitian ini merupakan data sekunder, yaitu data pelanggan yang diperoleh dari dataset publik Mall Customers Dataset yang banyak digunakan untuk analisis segmentasi pelanggan. Dataset ini memuat informasi pelanggan yang meliputi:

Variabel	Deskripsi	Tipe Data
CustomerID	Nomor identifikasi unik pelanggan	Numerik
Gender	Jenis kelamin pelanggan (Male/Female)	Kategorikal
Age	Umur pelanggan (tahun)	Numerik
Annual Income (k\$)	Pendapatan tahunan pelanggan (ribu dolar)	Numerik
Spending Score (1–100)	Skor pengeluaran pelanggan berdasarkan kebiasaan belanja	Numerik

Jumlah total data yang digunakan sebanyak 200 entri pelanggan, dengan masing-masing kolom menggambarkan karakteristik yang akan menjadi dasar proses pengelompokan.

3.5 Tahapan Penelitian

Tahapan penelitian secara umum terdiri atas enam langkah utama sebagaimana dijelaskan pada Gambar 3.1 berikut.



Gambar 3.1. Alur Tahapan Penelitian

3.6 Alat dan Perangkat Lunak

Penelitian ini dilakukan menggunakan perangkat keras dan lunak sebagai berikut:

- a. Perangkat keras: Laptop dengan prosesor Intel Core i5 dan RAM 8 GB.
- b. Perangkat lunak:
 1. Sistem operasi: Windows 10
 2. Bahasa pemrograman: Python 3.10

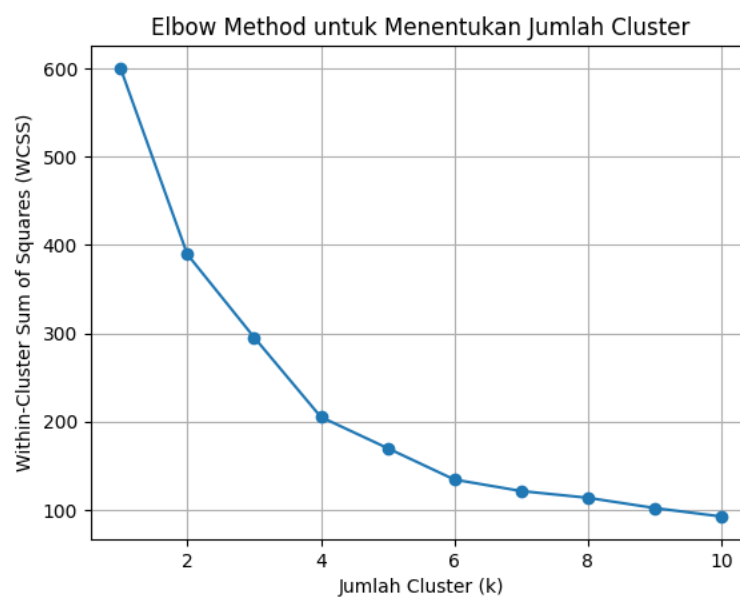
3. Lingkungan pengembangan: Google Colab / Jupyter Notebook
4. Pustaka: *pandas*, *numpy*, *matplotlib*, *seaborn*, dan *scikit-learn*.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Pengolahan Data

Setelah dilakukan proses data preprocessing dan normalisasi terhadap dataset yang berisi atribut CustomerID, Gender, Age, Annual Income (k\$), dan Spending Score (1–100), dilakukan pengelompokan menggunakan algoritma K-Means Clustering.

Untuk menentukan jumlah cluster yang optimal, digunakan metode Elbow Method. Metode ini dilakukan dengan menghitung nilai Within-Cluster Sum of Square (WCSS) untuk berbagai nilai k (jumlah cluster), kemudian hasilnya divisualisasikan dalam bentuk grafik seperti pada Gambar 4.1 berikut.



Gambar 4.1 Elbow Methode

Dari hasil grafik Elbow Method terlihat bahwa titik siku (elbow point) berada pada nilai $k = 4$, sehingga penelitian ini menggunakan empat cluster untuk menggambarkan segmentasi pelanggan yang optimal.

4.2 Hasil Clustering Menggunakan K-Means

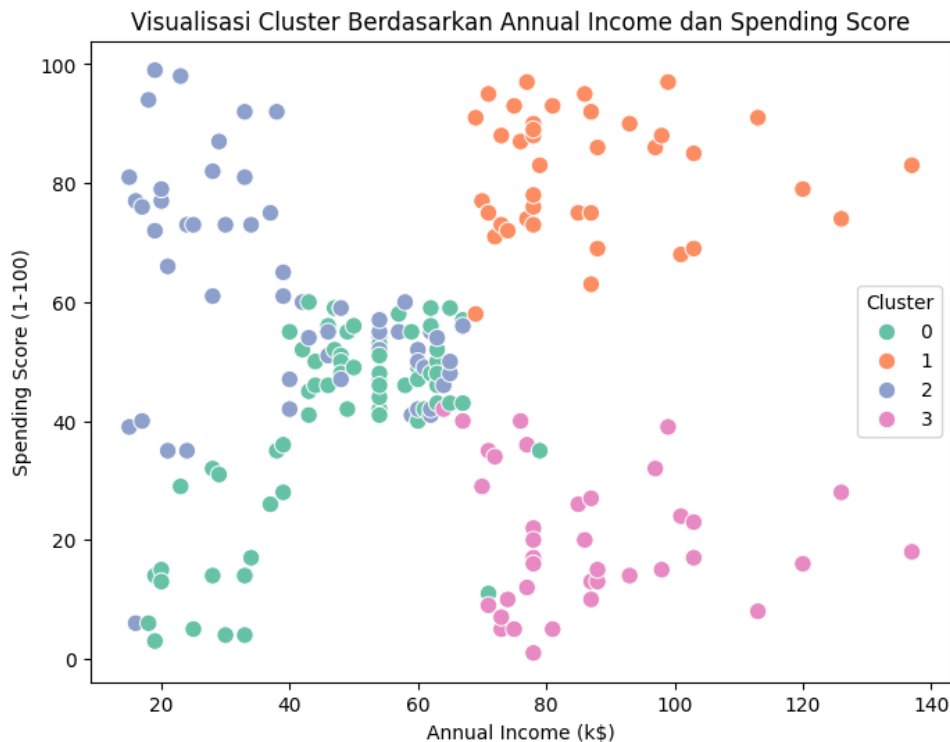
Setelah nilai k ditentukan, algoritma K-Means dijalankan untuk membentuk empat kelompok pelanggan (C0, C1, C2, dan C3). Berikut hasilnya pada tabel 4.1.

Tabel 4.1 Rata-Rata Variabel Tiap Cluster

Cluster	Age	Annual Income (k\$)	Spending Score (1–100)
0	53.984615	47.707692	39.969231
1	32.875000	86.100000	81.525000
2	25.438596	40.000000	60.298246
3	39.368421	86.500000	19.578947

4.3 Visualisasi Hasil Clustering

Untuk memahami hasil pengelompokan, dilakukan visualisasi data dalam ruang dua dimensi menggunakan kombinasi variabel *Annual Income* dan *Spending Score*.



Gambar 4.2 Visualisasi Cluster Berdasarkan Annual Income dan Spending Score

Berdasarkan hasil tersebut, dilakukan analisis dan interpretasi terhadap masing-masing cluster sebagai berikut:

1. Cluster 0 – Pelanggan Dewasa dengan Pola Konsumsi Moderat
 - a. Karakteristik: Usia rata-rata sekitar 54 tahun, pendapatan sedang (± 47 k\$), dan skor pengeluaran menengah (sekitar 40).
 - b. Perilaku: Pelanggan pada kelompok ini tergolong stabil dan berhati-hati dalam melakukan pembelian. Mereka memiliki daya beli cukup baik namun cenderung berbelanja secara selektif.
 - c. Strategi Bisnis: Dapat difokuskan pada promosi yang menekankan kualitas dan keandalan produk. Program loyalitas seperti *membership point* atau *reward program* dapat menarik minat kelompok ini untuk tetap berbelanja secara rutin.
2. Cluster 1 – Pelanggan Produktif dengan Daya Beli dan Konsumsi Tinggi
 - a. Karakteristik: Usia muda produktif (sekitar 33 tahun), pendapatan tinggi (± 86 k\$), dan skor pengeluaran tinggi (± 81).
 - b. Perilaku: Kelompok ini merupakan segmen pelanggan premium yang aktif berbelanja dan memiliki potensi besar sebagai high-value customer. Mereka menyukai pengalaman berbelanja yang cepat, mudah, dan eksklusif.
 - c. Strategi Bisnis: Fokuskan pada layanan personal, exclusive offers, dan peningkatan pengalaman pelanggan (customer experience). Promosi berbasis gaya hidup dan kemudahan digital akan sangat efektif.
3. Cluster 2 – Pelanggan Muda dengan Pengeluaran Aktif tetapi Pendapatan Sedang
 - a. Karakteristik: Usia muda (sekitar 25 tahun), pendapatan menengah ke bawah (sekitar 40 k\$), dengan skor pengeluaran cukup tinggi (sekitar 60).
 - b. Perilaku: Pelanggan dalam segmen ini memiliki gaya hidup konsumtif dan cenderung responsif terhadap promosi atau diskon. Mereka membeli produk karena tren dan kesenangan, bukan semata kebutuhan.

- c. Strategi Bisnis: Gunakan pendekatan promosi yang agresif, seperti potongan harga, *flash sale*, dan program *buy one get one*. Media sosial dapat menjadi kanal utama untuk menjangkau segmen ini.
4. Cluster 3 – Pelanggan Pendapatan Tinggi namun Pengeluaran Rendah
 - a. Karakteristik: Usia menengah (sekitar 39 tahun), pendapatan tinggi (sekitar 86 k\$), namun skor pengeluaran rendah (sekitar 19).
 - b. Perilaku: Kelompok ini memiliki kemampuan finansial tinggi namun sangat berhati-hati dalam berbelanja. Mereka cenderung membeli produk berdasarkan kebutuhan, bukan keinginan.
 - c. Strategi Bisnis: Dapat difokuskan pada promosi produk premium dengan pendekatan nilai (*value-based marketing*), menonjolkan kualitas, daya tahan, dan layanan purna jual yang baik.

5. Kesimpulan dan Saran

5.1. Kesimpulan

Berdasarkan hasil penelitian dan analisis yang telah dilakukan, dapat disimpulkan beberapa hal sebagai berikut:

1. Algoritma K-Means Clustering berhasil digunakan untuk mengelompokkan pelanggan menjadi empat segmen yang berbeda berdasarkan atribut Age, Annual Income (k\$), dan Spending Score (1–100). Pengelompokan ini mempermudah identifikasi karakteristik dan perilaku pelanggan dalam konteks strategi pemasaran.
2. Hasil dari proses Elbow Method menunjukkan bahwa jumlah cluster optimal adalah $k = 4$, di mana setiap cluster menunjukkan pola perilaku konsumsi yang berbeda:
 - a. Cluster 0: Pelanggan berusia dewasa dengan pendapatan sedang dan pola pengeluaran moderat (pelanggan stabil).
 - b. Cluster 1: Pelanggan muda produktif dengan pendapatan tinggi dan skor pengeluaran tinggi (pelanggan premium dan aktif).
 - c. Cluster 2: Pelanggan muda berpendapatan sedang dengan tingkat pengeluaran cukup tinggi (pelanggan potensial dengan respons tinggi terhadap promosi).
 - d. Cluster 3: Pelanggan berpendapatan tinggi namun dengan skor pengeluaran rendah (pelanggan konservatif dan selektif).
3. Faktor pendapatan tahunan (Annual Income) dan skor pengeluaran (Spending Score) memiliki pengaruh terbesar dalam pembentukan cluster, sebagaimana divisualisasikan dalam Gambar 4.2. Kedua variabel ini menentukan posisi pelanggan dalam kategori konsumtif, stabil, atau hemat.
4. Penerapan algoritma K-Means Clustering terbukti dapat memberikan gambaran yang jelas mengenai segmentasi pelanggan dan dapat dijadikan dasar bagi perusahaan untuk menyusun strategi pemasaran yang lebih efektif dan efisien. Dengan memahami karakteristik tiap kelompok pelanggan, perusahaan dapat mengarahkan sumber daya promosi, desain produk, dan layanan sesuai dengan kebutuhan spesifik tiap segmen.

5.2. Saran

1. Menambahkan variabel lain seperti frekuensi transaksi, jenis produk yang dibeli, atau durasi hubungan pelanggan untuk meningkatkan ketepatan segmentasi.
2. Menerapkan algoritma clustering lain seperti Hierarchical Clustering, DBSCAN, atau Gaussian Mixture Model untuk dibandingkan hasilnya dengan K-Means.
3. Menggunakan pendekatan visualisasi 3D atau PCA (Principal Component Analysis) untuk menggambarkan hubungan antar variabel dengan lebih komprehensif.

4. Dari sisi implementasi sistem, model clustering ini dapat dikembangkan menjadi dashboard interaktif berbasis web yang menampilkan segmentasi pelanggan secara dinamis. Hal ini akan memudahkan tim pemasaran dalam menganalisis dan mengambil keputusan berbasis data (*data-driven decision making*).

DAFTAR PUSTAKA

- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666.
- Kotler, P., & Keller, K. L. (2016). *Marketing Management* (15th ed.). Pearson Education.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281–297). University of California Press.
- Patel, D., & Patel, K. (2019). Customer segmentation using K-Means clustering algorithm. *International Journal of Computer Applications*, 178(45), 10–15.
- Prasad, S., & Reddy, M. (2020). Customer segmentation using K-Means clustering and PCA. *International Journal of Scientific & Technology Research*, 9(2), 517–521.
- Singh, S., & Chadha, S. (2017). A data mining approach for customer segmentation using K-Means clustering. *International Journal of Computer Applications*, 170(8), 22–27.
- Tan, P. N., Steinbach, M., & Kumar, V. (2018). *Introduction to Data Mining* (2nd ed.). Pearson.
- Wedel, M., & Kamakura, W. A. (2012). *Market Segmentation: Conceptual and Methodological Foundations* (2nd ed.). Springer.

