

SEGMENTASI PELANGGAN DENGAN ALGORITMA K-MEANS UNTUK STRATEGI PEMASARAN YANG EFEKTIF

Raihan insan pratama siagian, Ega pratama, Fauzan azmi Lubis, Selfi audy priscilia, Fanny Ramadhani

Ilmu Komputer, Universitas Negeri Medan

Jalan Willem Iskandar, Pasar V Medan Estate, Percut Sei Tuan, Deli Serdang, Indonesia

raihansiagian218@mhs.unimed.ac.id

ABSTRAK

Segmentasi pelanggan merupakan strategi penting dalam pemasaran untuk memahami karakteristik dan perilaku pelanggan yang beragam. Salah satu metode yang sering digunakan adalah K-Means Clustering, yang mampu mengelompokkan pelanggan berdasarkan kesamaan karakteristik tertentu. Namun, pemilihan jumlah cluster yang optimal dan analisis hasil segmentasi sering kali menjadi tantangan utama dalam implementasi metode ini. Oleh karena itu, penelitian ini bertujuan untuk mengimplementasikan algoritma K-Means dalam segmentasi pelanggan serta menganalisis efektivitas hasil klusterisasi. Pengujian dilakukan pada dataset Mall Customer yang didapat dari kagle, lalu dianalisis menggunakan algoritma K-Means. Hasil pengujian menunjukkan bahwa jumlah cluster optimal yang diperoleh adalah 5, dengan distribusi anggota sebagai berikut: Cluster 0 terdiri dari 81 pelanggan, Cluster 1 sebanyak 39 pelanggan, Cluster 2 sebanyak 22 pelanggan, Cluster 3 sebanyak 35 pelanggan Dan Cluster 4 sebanyak 23 pelanggan. Dengan demikian, penelitian ini menunjukkan bahwa metode K-Means dapat mengelompokkan pelanggan secara efektif untuk mendukung strategi pemasaran yang lebih terarah..

Kata kunci : *k-means, clustering, big data, data mining*

1. PENDAHULUAN

Dalam era digital yang semakin berkembang, data menjadi aset yang sangat berharga bagi perusahaan, khususnya dalam sektor bisnis dan pemasaran. Perusahaan kini dapat mengumpulkan berbagai informasi terkait pelanggan, mulai dari riwayat transaksi, pola pembelian, hingga preferensi produk. Data ini dapat dimanfaatkan untuk meningkatkan efektivitas strategi pemasaran, namun dalam jumlah yang besar dan beragam, analisis data secara manual menjadi tidak efisien. Oleh karena itu, diperlukan pendekatan berbasis data mining untuk menggali pola dan segmentasi pelanggan secara lebih sistematis.

Salah satu metode yang dapat digunakan dalam segmentasi pelanggan adalah K-Means Clustering. Menurut Yudi Agusta Phd, dalam penelitiannya yang berjudul, “K-Means Penerapan, Permasalahan dan Metode Terkait”. K-means clustering merupakan salah satu metode data clustering non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih cluster/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu cluster/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan cluster/kelompok yang lain sehingga data yang berada dalam satu cluster/kelompok memiliki tingkat variasi yang kecil [6][4].

Era yang serba digital ini, penjualan online telah mengubah cara berbelanja, hingga perusahaan e-commerce perlu memahami pelanggan mereka secara mendalam untuk menyediakan pengalaman yang relevan dan memuaskan. Selain itu, dalam persaingan bisnis yang semakin ketat, perusahaan harus menjaga hubungan yang baik dengan pelanggan. Salah satu

strategi yang efektif dalam menjaga hubungan tersebut adalah melalui segmentasi pelanggan, yang melibatkan pengelompokan pelanggan dengan kesamaan tertentu [1]. Dengan menerapkan K-Means Clustering, perusahaan dapat mengidentifikasi kelompok pelanggan dengan perilaku serupa dan merancang strategi pemasaran yang lebih tepat sasaran sesuai dengan karakteristik masing-masing segmen.

Dari ketiga penelitian tersebut, dapat disimpulkan bahwa algoritma K-Means Clustering telah terbukti efektif dalam membantu perusahaan memahami perilaku pelanggan dan menyusun strategi pemasaran yang lebih optimal. Oleh karena itu, penelitian ini akan mengadopsi pendekatan serupa dengan fokus pada segmentasi pelanggan berdasarkan data transaksi guna memberikan wawasan yang lebih mendalam dalam pengelolaan strategi pemasaran

2. TINJAUAN PUSTAKA

2.1. Big data & data mining

Data mining adalah suatu proses ekstraksi pengetahuan atau informasi yang berharga dari suatu set data yang besar dan kompleks. Tujuan utama dari data mining adalah mengidentifikasi pola, hubungan, atau informasi yang mungkin tidak terlihat secara langsung dalam data, sehingga dapat memberikan wawasan yang lebih dalam dan bernilai [10].

Peran Big Data dan Data Mining dalam Segmentasi Pelanggan Dalam proyek ini, Big Data dan Data Mining berperan dalam mengelompokkan pelanggan berdasarkan pola pembelian mereka. Dengan memanfaatkan algoritma K-Means Clustering, pelanggan dapat dikelompokkan berdasarkan atribut tertentu seperti frekuensi pembelian, jumlah transaksi, dan preferensi produk.

Hal ini memungkinkan bisnis untuk menerapkan strategi pemasaran yang lebih efektif, seperti rekomendasi produk yang dipersonalisasi dan penentuan target promosi yang lebih tepat sasaran.

2.2. Clustering Dalam Data Mining

Clustering merupakan teknik data mining yang berfungsi untuk mengelompokkan data berdasarkan kemiripan data pada suatu kelompok dan meminimalkan kemiripan pada kelompok lain [4].

Metode ini tergolong dalam unsupervised learning, di mana algoritma secara otomatis mencari pola dalam data untuk membentuk kelompok atau klaster yang memiliki kesamaan tertentu. Proses clustering memiliki peranan penting dalam berbagai bidang, termasuk analisis pemasaran, pengelompokan pelanggan, pengenalan pola, hingga deteksi anomali. Dalam konteks segmentasi pelanggan, clustering dapat membantu bisnis dalam memahami perilaku konsumen dan menyusun strategi pemasaran yang lebih efektif berdasarkan pola belanja atau preferensi pelanggan. Cluster adalah kumpulan record yang memiliki kemiripan suatu dengan yang lainnya dan memiliki ketidakmiripan dengan record dalam cluster lain [5].

Ada dua metode clustering yang kita kenal, yaitu hierarchical clustering dan partitioning [7].

2.3. ALGORITMA K-MEANS CLUSTERING

Algoritma K-Means Clustering merupakan salah satu metode pengelompokan data dalam data mining yang berbasis partisi (partitioning clustering). Algoritma ini bekerja dengan membagi sekumpulan data ke dalam sejumlah klaster yang telah ditentukan sebelumnya (K klaster) berdasarkan kemiripan karakteristik setiap data[2].

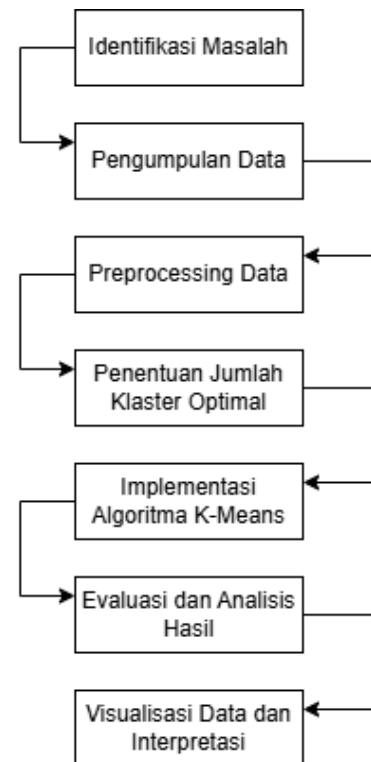
K-Means banyak digunakan dalam berbagai bidang, seperti analisis pasar, pengelompokan pelanggan, dan sistem rekomendasi, karena memiliki efisiensi komputasi yang tinggi serta kemudahan dalam implementasi.

2.4. KONSEP SEGMENTASI PELANGGAN

Penerapan segmentasi pelanggan dalam strategi bisnis dapat memberikan berbagai manfaat signifikan. Perusahaan dapat meningkatkan efektivitas komunikasi pemasaran dengan menyampaikan pesan yang lebih relevan kepada masing-masing segmen pelanggan. Selain itu, segmentasi pelanggan juga membantu dalam pengembangan produk atau layanan yang lebih sesuai dengan kebutuhan pasar sasaran. Dengan memahami perilaku pelanggan secara lebih mendalam, perusahaan dapat meningkatkan loyalitas pelanggan, mengoptimalkan pengalaman pengguna, serta meningkatkan profitabilitas bisnis secara keseluruhan[3][8].

3. METODE PENELITIAN

3.1. Tahapan Penelitian



Gambar 1. Alur Tahapan Penelitian

Penelitian ini dilakukan melalui beberapa tahapan utama. Tahap pertama adalah identifikasi masalah, di mana dilakukan analisis mengenai pentingnya segmentasi pelanggan dalam strategi pemasaran serta tantangan dalam memahami perilaku pelanggan secara efektif. Selain itu, ditinjau pula metode yang paling sesuai untuk mengelompokkan pelanggan berdasarkan karakteristik tertentu guna meningkatkan efektivitas pemasaran.

Tahap berikutnya adalah pengumpulan data, di mana dataset dikumpulkan dari berbagai sumber, seperti kagle. Dataset yang telah dikumpulkan kemudian melalui tahap pra-pemrosesan, yang mencakup pembersihan data untuk menghilangkan nilai yang tidak lengkap atau duplikat, normalisasi data agar setiap fitur memiliki skala yang seragam, serta pemilihan fitur yang paling relevan untuk proses klasterisasi.

Setelah pra-pemrosesan, dilakukan penentuan jumlah klaster optimal menggunakan metode Elbow Method dan Silhouette Score untuk menentukan jumlah segmen pelanggan yang paling representatif. Hasil dari tahap ini selanjutnya digunakan dalam implementasi algoritma K-Means Clustering, yang akan mengelompokkan pelanggan berdasarkan pola pembelian dan karakteristik mereka

Tahap selanjutnya adalah evaluasi hasil klasterisasi, yang dilakukan dengan menggunakan metrik seperti Davies-Bouldin Index atau Dunn Index untuk mengukur kualitas klaster yang terbentuk. Hasil klasterisasi kemudian divisualisasikan menggunakan scatter plot atau heatmap untuk mempermudah interpretasi.

Tahap akhir penelitian ini adalah penyusunan kesimpulan dan rekomendasi strategi pemasaran berdasarkan hasil segmentasi pelanggan yang diperoleh. Selain itu, diberikan pula saran pengembangan lebih lanjut untuk meningkatkan efektivitas segmentasi pelanggan dalam konteks pemasaran berbasis data.

3.2. Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dari dataset publik yang tersedia di platform Kaggle dan sumber lain yang relevan dengan segmentasi pelanggan. Pemilihan dataset dilakukan berdasarkan beberapa kriteria, seperti keberadaan fitur transaksi pelanggan, perilaku belanja, dan informasi demografi. Sebelum dilakukan segmentasi, dataset yang diperoleh perlu diproses lebih lanjut untuk memastikan kualitas data. Tahap pra-pemrosesan dilakukan untuk menangani nilai yang hilang, menghilangkan outlier, serta menormalkan data agar lebih sesuai dengan algoritma K-Means Clustering.

3.2.1. Pemeriksaan Data Hilang (Missing Values)

Dataset diperiksa untuk mengidentifikasi apakah terdapat nilai yang hilang. Jika ditemukan, dilakukan penanganan dengan metode berikut:

- Mean Imputation (mengganti dengan rata-rata) untuk variabel numerik seperti pendapatan tahunan.
- Mode Imputation (mengganti dengan nilai paling sering muncul) untuk variabel kategorikal seperti gender.

3.2.2. Deteksi dan Penanganan Outlier

Outlier dapat mempengaruhi hasil clustering, sehingga perlu dilakukan deteksi menggunakan boxplot dan z-score. Jika ditemukan nilai yang sangat ekstrem, maka dilakukan penghapusan atau transformasi data.

3.2.3. Normalisasi Data

Variabel dengan skala yang berbeda akan dinormalisasi menggunakan Min-Max Scaling atau Z-score Standardization agar semua fitur memiliki bobot yang setara dalam proses clustering.

3.2.4. Feature Selection (Pemilihan Fitur yang Relevan)

Beberapa fitur yang tidak memberikan kontribusi signifikan akan dihapus. Fitur yang dipertahankan adalah variabel yang relevan dalam segmentasi pelanggan, seperti Age, Annual Income, Spending Score. Menggunakan perintah `df_selected`.

```
df_selected = df[['Age', 'Annual Income (k$)', 'Spending Score (1-100)']]

print(df_selected.head(5))
```

	Age	Annual Income (k\$)	Spending Score (1-100)
0	19	15	39
1	21	15	81
2	20	16	6
3	23	16	77
4	31	17	40

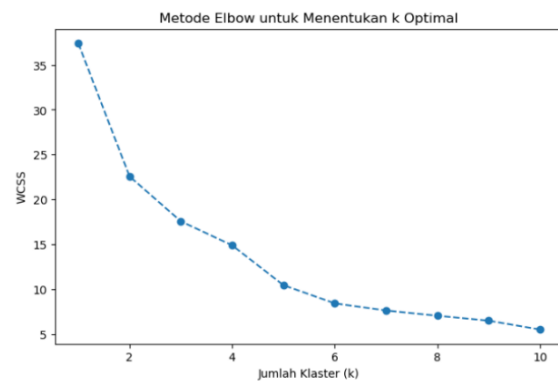
Gambar 2. Hasil Dari Feature Selection

3.3. Penentuan Jumlah Klaster Optimal

Sebelum menerapkan algoritma K-Means, perlu ditentukan jumlah klaster yang optimal agar segmentasi pelanggan dapat dilakukan secara efektif. Dalam penelitian ini, digunakan dua metode utama untuk menentukan jumlah klaster:

3.3.1. Elbow Method

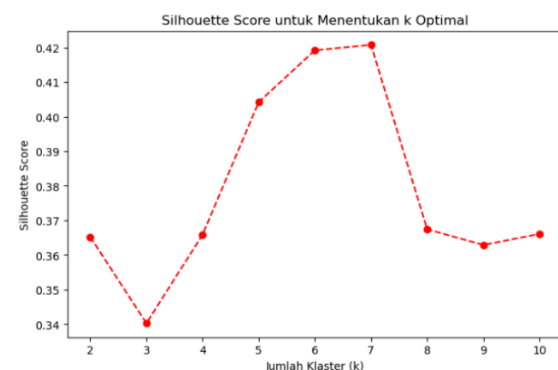
Metode ini mengevaluasi nilai Within-Cluster Sum of Squares (WCSS) untuk berbagai jumlah klaster. Jumlah klaster optimal ditentukan berdasarkan titik "elbow" pada grafik, di mana terjadi perubahan drastis dalam nilai WCSS.



Gambar 3. Penentuan jumlah klaster optimal menggunakan Elbow Method

3.3.2. Silhouette Score

Silhouette Score digunakan untuk mengukur seberapa baik objek dalam suatu klaster dikelompokkan. Nilai yang mendekati 1 menunjukkan bahwa klaster yang dibentuk lebih baik. Hasil dari kedua metode ini digunakan untuk menentukan jumlah klaster terbaik yang akan digunakan dalam algoritma K-Means.



Gambar 4. Penentuan jumlah klaster optimal menggunakan Silhouette Score

3.4. Implementasi K-Means Clustering

Algoritma K-Means Clustering bekerja dengan mengelompokkan data ke dalam sejumlah cluster berdasarkan kemiripan karakteristik. Persamaan dasar yang digunakan dalam perhitungan jarak antara titik data dengan centroid cluster adalah sebagai berikut:

$$d(x, c) = \sqrt{\sum_{i=1}^n (x_i - c_i)^2}$$

- $d(x, c)$ adalah jarak antara data dan centroid cluster
- x_i adalah nilai atribut dari data ke- i
- c_i adalah nilai centroid dari atribut ke- i
- n adalah jumlah atribut dalam dataset

Proses iteratif dilakukan untuk mengoptimalkan posisi centroid hingga tidak ada perubahan signifikan dalam klusterisasi.

3.5. EVALUASI HASIL CLUSTERING

Evaluasi dilakukan menggunakan Silhouette Score, yang mengukur bagaimana suatu titik dalam cluster sebanding dengan cluster lain. Nilai yang lebih tinggi menunjukkan segmentasi yang lebih baik.

3.6. VISUALISASI HASIL CLUSTERING

Hasil clustering divisualisasikan menggunakan *scatter plot* dengan pustaka *matplotlib* dan *seaborn*. Visualisasi ini bertujuan untuk:

- Memahami distribusi cluster.
- Mengidentifikasi pola pelanggan berdasarkan pendapatan dan skor belanja.
- Menunjukkan letak centroid dari masing-masing cluster.

Dengan metode ini, segmentasi pelanggan dapat dilakukan secara lebih akurat dan dapat digunakan untuk merancang strategi pemasaran yang lebih efektif.

4. HASIL DAN PEMBAHASAN

4.1. Proses K-Means Clustering

Proses K-Means Clustering pada penelitian ini dilakukan dengan tahapan sebagai berikut:

- Inisialisasi: Menentukan jumlah cluster (k) berdasarkan metode Elbow.
- Penentuan Centroid Awal: Memilih k titik data sebagai centroid awal secara acak.
- Perhitungan Jarak: Menghitung jarak setiap data ke centroid menggunakan rumus Euclidean Distance.
- Pengelompokan Data: Menempatkan setiap data ke dalam cluster dengan centroid terdekat.
- Pembaruan Centroid: Menghitung ulang centroid dari masing-masing cluster.
- Iterasi: Mengulang proses dari langkah 3 hingga 5 hingga nilai centroid tidak berubah signifikan atau jumlah iterasi maksimum tercapai.
- Evaluasi Hasil: Mengevaluasi hasil clustering dengan melihat distribusi data dalam masing-masing cluster.

4.2. Hasil Clustering

Berdasarkan implementasi algoritma K-Means pada dataset pelanggan, diperoleh hasil clustering yang menunjukkan adanya lima kluster optimal. Penentuan jumlah kluster optimal. Dilakukan melalui metode Elbow dan dikonfirmasi menggunakan Silhouette Score. Adapun distribusi pelanggan dalam masing-masing kluster adalah sebagai berikut:

- Kluster 0: 81 pelanggan
- Kluster 1: 39 pelanggan
- Kluster 2: 22 pelanggan
- Kluster 3: 35 pelanggan
- Kluster 4: 23 pelanggan

```
import numpy as np
import pandas as pd

# Menghitung jumlah anggota di setiap cluster
cluster_counts = pd.Series(kmeans.labels_.value_counts().sort_index)

# Menampilkan jumlah anggota per cluster
print(cluster_counts)
```

0	81
1	39
2	22
3	35
4	23

Name: count, dtype: int64

Gambar 5. Jumlah Anggota Pada Setiap Cluster

4.3. Evaluasi Hasil Clustering

```
from sklearn.metrics import silhouette_score

silhouette_avg = silhouette_score(X, kmeans.labels_)
print(f"Silhouette Score: {silhouette_avg}")
```

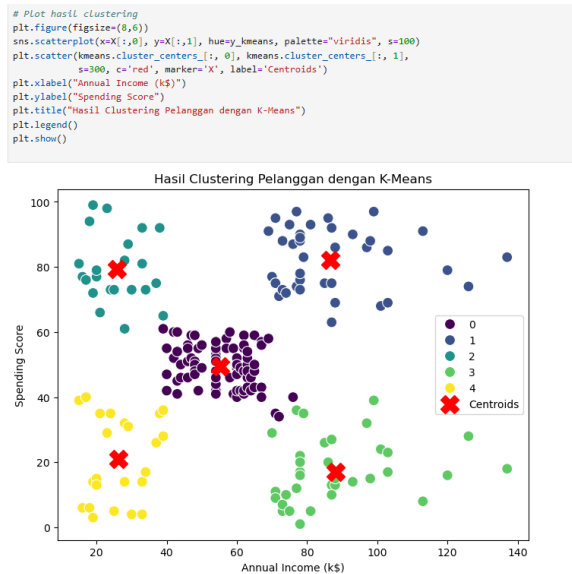
Silhouette Score: 0.553931997444648

Gambar 6. Evaluasi Hasil Clustering

Evaluasi terhadap hasil clustering dilakukan dengan menggunakan nilai Silhouette Score untuk mengukur kualitas pemisahan antar cluster. Berdasarkan analisis yang telah dilakukan, diperoleh nilai Silhouette Score sebesar 0.553. Nilai ini menunjukkan bahwa cluster yang terbentuk memiliki pemisahan yang cukup baik, meskipun masih terdapat ruang untuk peningkatan kualitas segmentasi melalui pemilihan metode clustering lain atau optimalisasi parameter yang lebih lanjut.

4.4. Visualisasi Hasil Clustering

Visualisasi hasil clustering menggunakan metode K-Means menghasilkan lima cluster pelanggan yang divisualisasikan dalam scatter plot berdasarkan pendapatan tahunan (Annual Income) dan skor pengeluaran (Spending Score). Setiap cluster direpresentasikan dengan warna berbeda, sementara centroid masing-masing cluster ditandai dengan simbol merah (X), menunjukkan titik pusat dari setiap kelompok pelanggan yang terbentuk.



Gambar 7. Visualisasi Hasil Clustering

Dari hasil visualisasi, terlihat bahwa kluster pelanggan terbagi menjadi beberapa kelompok utama dengan karakteristik yang berbeda:

- Pelanggan dengan pendapatan tinggi dan skor pengeluaran tinggi – Menunjukkan segmen pelanggan premium yang aktif dalam konsumsi.
- Pelanggan dengan pendapatan tinggi dan skor pengeluaran rendah – Mengindikasikan potensi pasar yang dapat dioptimalkan dengan strategi pemasaran lebih agresif.
- Pelanggan dengan pendapatan menengah dan skor pengeluaran sedang – Mewakili kelompok pelanggan stabil yang dapat ditingkatkan loyalitasnya.
- Pelanggan dengan pendapatan rendah dan skor pengeluaran tinggi – Menunjukkan pelanggan dengan kecenderungan belanja tinggi meskipun memiliki keterbatasan finansial.
- Pelanggan dengan pendapatan rendah dan skor pengeluaran rendah – Kelompok ini memiliki potensi konsumsi rendah, sehingga memerlukan pendekatan pemasaran yang lebih spesifik.

Visualisasi ini memberikan wawasan yang lebih jelas mengenai pola konsumsi pelanggan dan dapat digunakan sebagai dasar untuk merancang strategi pemasaran berbasis data yang lebih efektif. Dengan memahami distribusi pelanggan dalam kluster yang terbentuk, perusahaan dapat mengoptimalkan alokasi sumber daya untuk meningkatkan keterlibatan dan retensi pelanggan di masing-masing segmen.

4.5. Pembahasan

Berdasarkan hasil clustering yang diperoleh, dapat disimpulkan bahwa pelanggan dapat dikelompokkan berdasarkan pola belanja mereka. Kluster yang terbentuk menunjukkan variasi pola pengeluaran dan pendapatan pelanggan, sehingga dapat digunakan sebagai dasar dalam perancangan strategi pemasaran yang lebih efektif. Beberapa

implikasi hasil clustering terhadap strategi pemasaran adalah sebagai berikut:

- Pelanggan dengan Pendapatan Tinggi dan Skor Pengeluaran Tinggi
 - Strategi: Fokus pada strategi retensi pelanggan premium dengan memberikan program loyalitas eksklusif, diskon khusus, atau layanan premium.
 - Contoh Implementasi: Menawarkan membership VIP, layanan personal shopper, atau pengalaman belanja eksklusif untuk meningkatkan engagement dan retensi.
- Pelanggan dengan Pendapatan Tinggi dan Skor Pengeluaran Rendah
 - Strategi: Meningkatkan insentif untuk meningkatkan pembelanjaan, misalnya dengan promosi berbasis personalisasi.
 - Contoh Implementasi: Menggunakan email marketing dengan rekomendasi produk premium atau strategi bundling untuk menarik perhatian mereka.
- Pelanggan dengan Pendapatan Menengah dan Skor Pengeluaran Sedang
 - Strategi: Fokus pada peningkatan loyalitas pelanggan dengan promosi reguler dan program diskon.
 - Contoh Implementasi: Menawarkan reward points atau cashback pada pembelian tertentu agar mereka tetap setia pada brand.
- Pelanggan dengan Pendapatan Rendah dan Skor Pengeluaran Tinggi
 - Strategi: Mengoptimalkan strategi promosi harga dan memberikan opsi pembayaran fleksibel.
 - Contoh Implementasi: Menawarkan diskon khusus, program cicilan, atau promo cashback agar pelanggan tetap tertarik berbelanja meskipun memiliki keterbatasan finansial.
- Pelanggan dengan Pendapatan Rendah dan Skor Pengeluaran Rendah
 - Strategi: Menargetkan mereka dengan promosi harga menarik dan paket hemat untuk meningkatkan daya beli mereka.
 - Contoh Implementasi: Memberikan promo berbasis kebutuhan dasar, kupon diskon berkala, atau program referral untuk mendorong pelanggan baru.

Dengan demikian, penggunaan algoritma K-Means dalam segmentasi pelanggan terbukti efektif dalam mengidentifikasi kelompok pelanggan yang memiliki karakteristik serupa. Hasil penelitian ini dapat dimanfaatkan oleh perusahaan dalam menyusun strategi pemasaran yang lebih terarah, sehingga meningkatkan efisiensi dan efektivitas dalam menarik serta mempertahankan pelanggan.

5. KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa algoritma K-Means dapat digunakan secara efektif dalam segmentasi pelanggan untuk strategi pemasaran yang

lebih terarah. Dari hasil pengujian, diperoleh jumlah cluster optimal sebanyak 4 dengan distribusi anggota: Cluster 0 terdiri dari 81 pelanggan, Cluster 1 sebanyak 39 pelanggan, Cluster 2 sebanyak 22 pelanggan, Cluster 3 sebanyak 35 pelanggan, dan Cluster 4 sebanyak 23 pelanggan. Dengan adanya segmentasi ini, perusahaan dapat menyesuaikan strategi pemasaran berdasarkan karakteristik pelanggan di masing-masing cluster. Sebagai saran, penelitian selanjutnya dapat mengeksplorasi metode lain seperti DBSCAN atau Hierarchical Clustering untuk membandingkan efektivitas segmentasi. Kluster 0: 81 pelanggan

DAFTAR PUSTAKA

- [1] Awalina, E. F. L., & Rahayu, W. I. (2023). Optimalisasi strategi pemasaran dengan segmentasi pelanggan menggunakan penerapan K-means clustering pada transaksi online retail. *Jurnal Teknologi Dan Informasi*, 13(2), 122-137.
- [2] Ariati, I., Norsa, R. N., Akhsan, L., & Heikal, J. (2023). Segmentasi Pelanggan Menggunakan K-Means Clustering Studi Kasus Pelanggan Uht Milk Greenfield. *Cerdika: Jurnal Ilmiah Indonesia*, 3(7), 729-743.
- [3] Achmad, S. L., Fauzi, A., Rahmat, R., & Indra, J. (2024). SEGMENTASI PELANGGAN MENGGUNAKAN K-MEANS CLUSTERING DI TOKO RETAIL. *Jurnal Tekinkom (Teknik Informasi dan Komputer)*, 7(2), 736-747.
- [4] Alkhairi, P., & Windarto, A. P. (2019). Penerapan K-Means Cluster Pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara. *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, 762-767.
- [5] Hafidz Ardana, C., Khoyum, A. A. A. A. A., & Faisal, M. (2024). Segmentasi pelanggan penjualan online menggunakan Metode K-Means Clustering. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 9(1), 1-9.
- [6] Kurniawan, S. D., Hermanto, D. M. C., Widiastuti, R. Y., Pipin, S. J., Mukhlis, I. R., Surianto, D. F., ... & Judijanto, L. (2024). *Big Data (Menenal Big Data & Implementasinya di Berbagai Bidang)*. PT. Sonpedia Publishing Indonesia.
- [7] Sulistyowati, S. (2018, September). Analisa Segmentasi Konsumen Menggunakan Algoritma K-Means Clustering. In *Prosiding Seminar Nasional Sains dan Teknologi Terapan* (pp. 51-58).
- [8] Permana, S. A., Florentin, S. F., & Santoso, A. (2022). Analisis Segmentasi Pelanggan Menggunakan K-Means Clustering Studi Kasus Aplikasi Alfagift. *Sebatik*, 26 (2), 446-457.
- [9] Ramadhan, A. G. (2023). Data mining untuk segmentasi pelanggan dengan algoritma K-means: Studi kasus pada data pelanggan di toko retail. *Syntax Literate; Jurnal Ilmiah Indonesia*, 8(10), 5698-5715.
- [10] P. W. Rahayu, I. G. I. Sudipa, Suryani, A. Surachman, A. Ridwan, I. G. M. Darmawiguna, M. N. Sutoyo, I. Slamet, S. Harlina, dan I. M. D. Maysanjaya, *Buku Ajar Data Mining*. Jambi: Sonpedia Publishing, 2024.