

PENERAPAN ALGORITMA CLUSTERING DALAM SEGMENTASI PASAR BERDASARKAN DATA DEMOGRAFIS UNTUK STRATEGI PEMASARAN YANG LEBIH EFEKTIF

Arridho Ramadhan Firdaus¹, Syafika Dhea Taqiya Ramadhani²

¹ Program Studi Teknologi Informasi, Universitas Aisyiyah Yogyakarta

² Program Studi Manajemen, Universitas Aisyiyah Yogyakarta

Jl. Siliwangi (Ring Road Barat) No. 63 Nogotirto, Gamping, Sleman, Yogyakarta. 55292

arridhoramadhanfirdaus@unisayogya.ac.id

ABSTRAK

Perubahan perilaku konsumen menuntut perusahaan menerapkan pemasaran berbasis data (*data-driven marketing*) untuk memahami pola pembelian berdasarkan demografi. Permasalahan sering kali muncul ketika data pelanggan menjadi semakin kompleks dan terus bertambah, sehingga proses segmentasi secara manual tidak lagi memadai dan membutuhkan pengelompokan secara otomatis. Penelitian ini bertujuan untuk mengembangkan model segmentasi pelanggan menggunakan algoritma K-Means secara terfokus serta menyusun strategi pemasaran langsung dari hasil segmentasi tersebut. Pendekatan metode kuantitatif digunakan dengan memanfaatkan dataset *Mall Customer Segmentation Data* dari platform Kaggle. Algoritma K-Means diterapkan sebagai metode komputasi utama mulai dari penentuan jumlah kelompok optimal menggunakan metode *Elbow* hingga tahap pemetaan akhir. Kualitas hasil pembentukan lima *cluster* pelanggan tersebut dievaluasi secara matematis menggunakan metrik *Silhouette Score*. Hasil perhitungan metrik tersebut mencatatkan skor sebesar 0,3575963460506624. Penelitian ini memetakan profil dan karakteristik khas pada setiap kelompok pelanggan yang kemudian menjadi landasan valid untuk merumuskan rekomendasi strategi pemasaran yang lebih relevan, konkret, dan tepat sasaran.

Kata kunci : K-Means; Segmentasi_Pelanggan; Strategi_Pemasaran; Data-Driven_Marketing

1. PENDAHULUAN

Dunia pemasaran mengalami perubahan yang sangat cepat dalam beberapa tahun terakhir. Konsumen tidak lagi bisa dikelompokkan hanya dengan asumsi sederhana seperti usia atau jenis kelamin. Perilaku konsumen dipengaruhi oleh banyak faktor beragam, mulai dari tingkat pendapatan hingga pola belanja. Perusahaan mulai beralih dari pendekatan tradisional menuju strategi pemasaran berbasis data atau *data-driven marketing* untuk menghadapi perubahan tersebut [1].

Segmentasi pelanggan menjadi salah satu pendekatan yang paling sering digunakan dalam strategi pemasaran ini. Tujuan utama segmentasi ini adalah membagi konsumen ke dalam kelompok tertentu agar penargetan audiens menjadi lebih relevan [2].

Data demografis seperti umur, jenis kelamin, pendapatan, dan tingkat pengeluaran sering dijadikan acuan utama karena mudah diperoleh dan cukup representatif untuk memprediksi perilaku belanja konsumen.

Permasalahan sering kali muncul ketika data yang ada menjadi semakin kompleks seiring dengan bertambahnya jumlah pelanggan. Proses segmentasi secara manual jelas tidak lagi memadai untuk menangani basis data berskala besar tersebut. Metode analisis yang mampu mengelompokkan pelanggan secara otomatis, konsisten, dan mudah ditafsirkan sangat dibutuhkan oleh perusahaan saat ini [3].

Algoritma clustering K-Means hadir sebagai salah satu solusi yang paling efektif dan banyak

digunakan, metode ini tidak hanya mampu memproses data tanpa label kelompok sebelumnya, tetapi juga dapat memberikan gambaran yang lebih jelas mengenai karakteristik unik dari setiap tipe pelanggan [4], [5].

Berbagai penelitian sebelumnya telah menguji efektivitas algoritma clustering untuk segmentasi pelanggan, khususnya di sektor ritel maupun e-commerce [6].

Beberapa penelitian berupaya mengembangkan model clustering dengan menggabungkan algoritma K-Means beserta optimisasi parameternya. Penelitian mengenai segmentasi pelanggan menggunakan clustering juga cukup berkembang pesat di Indonesia. Penerapan algoritma K-Means telah banyak dieksplorasi pada data pelanggan telekomunikasi, ritel modern, hingga aplikasi belanja daring [7], [8], [9].

Sebagian besar penelitian terdahulu masih memiliki keterbatasan berupa fokus murni pada ranah teknis keilmuan data semata. Peneliti terdahulu sering kali belum sampai pada tahap menghubungkan hasil segmentasi komputasi tersebut dengan rekomendasi strategi pemasaran yang konkret secara manajerial.

Penelitian ini dilakukan untuk menjawab bagaimana algoritma K-Means dapat digunakan secara optimal dalam membentuk kelompok pelanggan berdasarkan data demografis. Analisis berfokus pada upaya menemukan ciri khas dari setiap kelompok pelanggan tersebut agar dapat dijadikan dasar pemetaan target pasar yang akurat. Evaluasi kuantitatif dari model machine learning dikombinasikan langsung

dengan analisis manajerial untuk menghasilkan riset yang terukur secara komersial [10].

Hasil pengelompokan K-Means diharapkan tidak hanya valid secara perhitungan statistik, tetapi juga mampu memberikan rekomendasi strategi pemasaran langsung yang bernilai tambah bagi operasional bisnis perusahaan.

2. TINJAUAN PUSTAKA

Berbagai penelitian sebelumnya telah menguji efektivitas algoritma clustering untuk segmentasi pelanggan di berbagai sektor bisnis. Mengintegrasikan algoritma evolusi diferensial berbasis Q-learning dengan K-Means untuk memaksimalkan segmentasi pelanggan dalam pemasaran digital [11]. Mengaplikasikan machine learning untuk memprediksi perilaku konsumen sebagai dasar penentuan strategi pemasaran yang lebih ditargetkan [12].

Pendekatan analitik bisnis dua tahap guna melakukan segmentasi pelanggan berdasarkan perilaku dan geografis pelanggan [13].

Penelitian [2] membuktikan efektivitas K-Means clustering secara khusus untuk memetakan pelanggan pada sektor e-commerce.

Pengembangan metode clustering terus dilakukan oleh para peneliti untuk mengatasi tantangan karakteristik data yang beragam. Pendekatan clustering berbasis puncak kepadatan adaptif untuk menangani sebaran data yang tidak terdistribusi secara merata telah dikembangkan [14].

Penelitian [15] mengusulkan optimisasi nilai K secara otomatis yang digabungkan dengan K-Means berbasis RFM guna meningkatkan kualitas segmentasi. S. Wang mencoba pendekatan segmentasi pelanggan dinamis dengan menggabungkan parameter LRFMS dan clustering deret waktu multivariat.

Penelitian mengenai segmentasi pelanggan menggunakan algoritma clustering juga berkembang sangat pesat di Indonesia. Algoritma clustering untuk membagi kelompok pelanggan secara spesifik berdasarkan karakteristik transaksinya [16].

Penelitian [17] mengeksplorasi penggunaan algoritma Fuzzy C-Means sebagai alternatif untuk mengelompokkan pelanggan guna mendapatkan batas cluster yang lebih fleksibel.

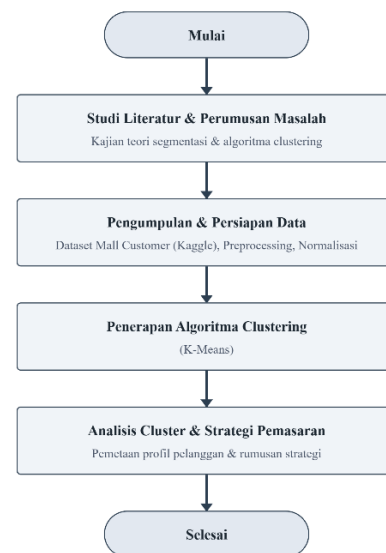
Penelitian [18] mengimplementasikan algoritma K-Means pada data pelanggan jaringan telekomunikasi guna mengoptimalkan strategi pemasaran perusahaan tersebut.

Sebagian besar penelitian terdahulu tersebut masih memiliki keterbatasan berupa fokus pada penggunaan satu jenis algoritma saja tanpa membandingkannya dengan algoritma alternatif. Peneliti terdahulu sering kali belum sampai pada tahap menghubungkan hasil segmentasi komputasi dengan rekomendasi strategi pemasaran yang konkret secara manajerial. Penelitian ini hadir untuk mengisi celah tersebut dengan mengoptimalkan penerapan algoritma K-Means secara spesifik untuk segmentasi pasar.

Algoritma K-Means dieksekusi secara terfokus mulai dari penentuan jumlah kelompok optimal hingga pemetaan karakteristik demografis pelanggan. Evaluasi kuantitatif dikombinasikan dengan analisis manajerial agar hasil pengelompokan tidak hanya valid secara statistik, tetapi juga sangat relevan dengan kebutuhan bisnis. Rekomendasi strategi pemasaran langsung disusun dari hasil pemetaan K-Means tersebut untuk memberikan nilai tambah nyata pada bidang manajemen pemasaran

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan algoritma *clustering* untuk melakukan segmentasi pelanggan berbasis data demografis. Dataset yang digunakan bersumber dari *Mall Customer Segmentation Data* pada platform Kaggle. Data tersebut memuat variabel umur, jenis kelamin, pendapatan tahunan, dan skor pengeluaran atau *spending score*. Metode *clustering* dipilih berdasarkan kemampuannya mengelompokkan data tanpa label pendahuluan. Pendekatan komputasi diterapkan secara spesifik menggunakan algoritma K-Means untuk memetakan pola konsumen secara optimal. Algoritma K-Means difungsikan sebagai metode utama sesuai tren penggunaannya yang luas dan terbukti efektif dalam berbagai studi segmentasi pelanggan ritel. Tingkat stabilitas dan kualitas *cluster* yang dihasilkan oleh algoritma tunggal ini dievaluasi secara kuantitatif untuk memastikan relevansinya terhadap karakteristik data demografis yang diolah. Berikut merupakan tahapan penelitian ditujukan pada Gambar 1.



Gambar 1. Tahapan Penelitian

3.1. Studi Literatur dan Perumusan Masalah

Tahap pertama difokuskan pada kegiatan studi literatur dan perumusan masalah. Pengkajian teori dilakukan terhadap konsep segmentasi pasar berbasis data, ragam algoritma *clustering*. Kerangka konseptual segmentasi pelanggan berbasis *clustering* dihasilkan

secara utuh dari tahapan ini. Asumsi dasar penelitian turut dirumuskan untuk memperjelas batasan masalah yang akan diselesaikan.

3.2. Pengumpulan dan Persiapan Data

Tahap kedua berpusat pada proses pengumpulan dan persiapan data utama. Dataset yang digunakan dalam penelitian ini adalah Mall Customer Segmentation Data yang bersumber dari platform Kaggle. Berkas data tersebut berformat CSV dengan nama file Mall_Customers.csv. Data ini memuat informasi demografis dan perilaku belanja pelanggan untuk keperluan analisis *clustering*.

Secara spesifik, dataset ini terdiri dari lima kolom atribut utama:

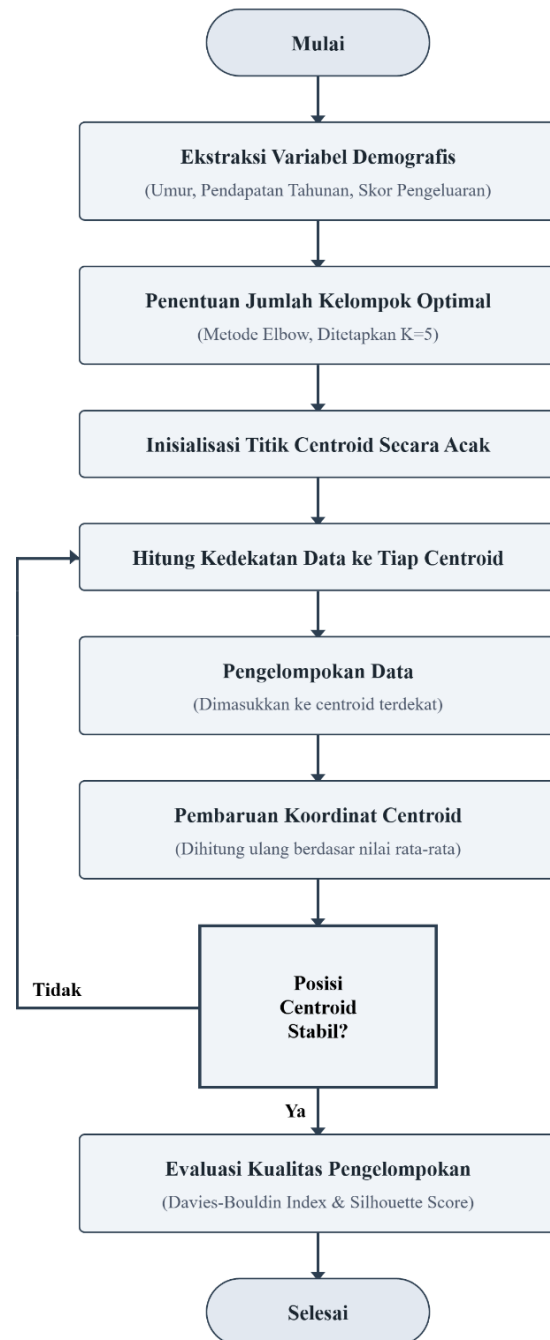
- CustomerID**: Nomor identitas unik untuk setiap pelanggan. Kolom ini berisi angka urut (seperti 1, 2, 3, dan seterusnya) untuk membedakan satu pelanggan dengan pelanggan lainnya.
- Gender** (Jenis Kelamin): Atribut berformat teks (kategorikal) yang berisi informasi jenis kelamin pelanggan, yaitu *Male* (Laki-laki) dan *Female* (Perempuan). Nilai teks ini kemudian diubah menjadi format angka menggunakan fungsi LabelEncoder pada tahap pra-pemrosesan agar dapat dihitung oleh algoritma.
- Age** (Umur): Usia pelanggan yang ditulis dalam bentuk angka bulat. Lima baris pertama data menunjukkan contoh rentang usia pelanggan yang bervariasi, seperti 19, 20, 21, 23, dan 31 tahun.
- Annual Income (k\$)** (Pendapatan Tahunan): Nilai pendapatan pelanggan selama satu tahun yang dinyatakan dalam satuan ribuan Dolar Amerika Serikat (USD). Angka "15" pada kolom ini memiliki arti bahwa pendapatan pelanggan tersebut adalah \$15.000 per tahun.
- Spending Score** (1-100) (Skor Pengeluaran): Nilai metrik evaluasi yang diberikan oleh pihak mal atau perusahaan ritel kepada pelanggan berdasarkan riwayat transaksi dan perilaku belanjanya. Nilai ini dibatasi pada rentang skala 1 hingga 100. Skor tinggi (misalnya 81 atau 77) menunjukkan pelanggan yang sangat konsumtif, sedangkan skor rendah (misalnya 6) menunjukkan pelanggan yang jarang berbelanja atau berhemat.

Proses komputasi algoritma K-Means pada penelitian ini tidak menggunakan seluruh kolom tersebut. Tiga variabel terakhir, yaitu umur (*Age*), pendapatan tahunan (*Annual Income*), dan skor pengeluaran (*Spending Score*), disaring secara khusus untuk menjadi input variabel utama dalam pembentukan kelompok sasaran pemasaran.

3.3. Penerapan Algoritma Clustering

Tahap ketiga memfokuskan kegiatan pada penerapan algoritma clustering menggunakan K-Means sebagai metode dasar atau baseline. Algoritma K-Means dipilih berkat kemampuannya yang terbukti

sederhana dan sangat efektif untuk mengelompokkan data pelanggan berkapasitas besar di berbagai konteks ritel [2]. Berikut merupakan alur algoritma dari K-Means ditujukan pada gambar 2.



Gambar 2. Tahapan Algoritma K-means

3.4. Ekstraksi Variabel Demografis

Tahap komputasi diawali dengan menyeleksi dan mengekstraksi atribut data yang relevan dari dataset Mall Customer Segmentation. Dalam penelitian ini, tiga variabel demografis utama yang digunakan sebagai sumbu analisis (ruang fitur) adalah Umur (*Age*), Pendapatan Tahunan (*Annual Income*), dan Skor Pengeluaran (*Spending Score*). Data yang telah diekstraksi dipastikan sudah melalui tahap

prapemrosesan (normalisasi) agar memiliki skala yang seragam sebelum masuk ke perhitungan jarak matematis.

3.5. Penentuan Jumlah Kelompok Optimal

Sebelum K-Means dijalankan secara penuh, jumlah kelompok atau cluster (K) yang paling optimal harus ditentukan. Proses ini dipandu menggunakan Metode Elbow, di mana algoritma akan menghitung nilai inersia (jarak kuadrat antar titik data dengan pusat cluster) untuk berbagai nilai K. Dengan melacak titik penurunan variansi yang drastis hingga membentuk patahan siku (batas pelandaian kurva grafik), parameter jumlah kelompok kemudian ditetapkan secara definitif sebanyak lima kelompok (K=5) [4].

3.6. Inisialisasi Titik *Centroid* Secara Acak

Setelah parameter K=5 ditetapkan, algoritma memulai proses dengan menginisialisasi atau menempatkan 5 buah titik pusat (*centroid*) awal. Penempatan titik-titik pusat ini dilakukan secara acak (random) ke dalam ruang data komputasi[16]. Titik acak ini akan bertindak sebagai "jangkar" awal untuk membentuk kelompok.

3.7. Proses Iterasi (Siklus Berulang)

Bagian ini merupakan inti dari algoritma K-Means ditandai dengan garis looping atau siklus berulang pada *flowchart*. Proses iterasi terdiri dari tiga langkah matematis [19]:

- **Hitung Kedekatan Data ke Tiap *Centroid*:** Algoritma mengevaluasi kedekatan setiap titik data pelanggan dengan kelima titik *centroid* yang ada menggunakan rumus perhitungan jarak (umumnya menggunakan *Euclidean Distance*).
- **Pengelompokan Data:** Berdasarkan hasil perhitungan jarak tersebut, setiap data konsumen secara otomatis dimasukkan (di-assign) ke dalam kelompok yang titik pusatnya (*centroid*) paling dekat dengan data tersebut.
- **Pembaruan Koordinat *Centroid*:** Setelah semua data memiliki kelompok, algoritma menghitung ulang nilai rata-rata (mean) dari seluruh atribut data yang ada di dalam masing-masing kelompok. Nilai rata-rata baru ini kemudian dijadikan sebagai titik koordinat *centroid* yang baru, sehingga titik pusat akan bergeser ke tengah kerumunan data.

3.8. Keputusan: Posisi *Centroid* Stabil?

Pada titik belah ketupat ini, algoritma melakukan pengecekan konvergensi. Apakah titik *centroid* yang baru saja dihitung mengalami pergeseran dari titik sebelumnya?

- **Jika Tidak (*Centroid* belum stabil):** Maka alur akan berulang naik ke atas (iterasi), di mana algoritma akan kembali menghitung jarak data ke *centroid* yang baru, lalu mengelompokkan ulang, dan memperbarui posisinya lagi.
- **Jika Ya (*Centroid* sudah stabil/konvergen):** Artinya tidak ada lagi data yang berpindah kelompok dan

posisi *centroid* sudah terkunci. Iterasi algoritma dihentikan dan lanjut ke tahap akhir.

3.9. Evaluasi Kualitas Pengelompokan

Sebagai tahap final dari pemodelan K-Means, hasil segmentasi yang telah konvergen akan diuji secara kuantitatif. Evaluasi ini menggunakan *Davies-Bouldin Index* (DBI) untuk mengukur rasio penyebaran dalam kelompok (*intra-cluster*) dibandingkan pemisahan antar kelompok (*inter-cluster*), serta metrik *Silhouette Score* untuk memvalidasi seberapa padat dan tepat setiap data berada di kelompoknya masing-masing [18].

3.10. Analisis Cluster dan Strategi Pemasaran

Tahap kelima memfokuskan kegiatan pada analisis cluster secara mendalam beserta perumusan strategi pemasaran yang aplikatif. Profil pelanggan disusun secara terperinci menggunakan bantuan visualisasi grafik dua dimensi untuk memetakan sebaran data konsumen. Pemetaan visual ini mempermudah identifikasi ciri demografis unik dari setiap cluster yang terbentuk berdasarkan korelasi variabel umur, pendapatan tahunan, dan skor pengeluaran. Karakteristik spesifik dari suatu kelompok dipelajari secara komprehensif, seperti penemuan segmen pelanggan berpendapatan menengah yang tersebar di berbagai kelompok usia atau segmen pelanggan usia muda dengan daya beli sangat tinggi.

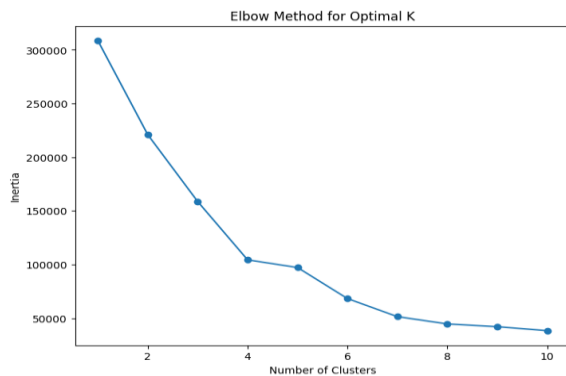
Pemahaman mendalam terhadap karakteristik setiap segmen tersebut sangat krusial dalam transisi manajemen menuju pendekatan pemasaran berbasis data atau *data-driven marketing*. Profil pelanggan yang telah teridentifikasi dimanfaatkan secara langsung sebagai landasan utama penyusunan kampanye pemasaran yang jauh lebih relevan dan efisien. Implementasi strategi pemasaran dapat disesuaikan secara presisi, misalnya melalui penawaran produk premium berdesain kekinian untuk segmen usia muda berpendapatan tinggi, atau pemberian program loyalitas khusus bagi segmen konsumen yang terbukti sangat konsumtif. Pendekatan analisis ini memastikan bahwa hasil riset komputasi tidak hanya valid secara perhitungan statistik semata, tetapi juga mampu memberikan nilai tambah secara langsung bagi operasional bisnis dan pengambilan keputusan manajemen pemasaran di perusahaan.

4. HASIL DAN PEMBAHASAN

Penelitian ini telah berhasil mengimplementasikan tahapan clustering menggunakan bahasa pemrograman Python. Dataset Mall Customer Segmentation diproses melalui serangkaian tahapan komprehensif mulai dari persiapan hingga evaluasi model. Tampilan awal data mentah memuat atribut identitas pelanggan, jenis kelamin, umur, pendapatan tahunan, dan skor pengeluaran. Proses pra-pemrosesan data diawali dengan mengubah atribut kategorikal menjadi format

numerik agar relevan dengan algoritma machine learning. Kolom jenis kelamin ditransformasikan menggunakan fungsi LabelEncoder dari pustaka scikit-learn. Tiga variabel utama dipilih secara spesifik untuk proses *clustering*, yaitu umur, pendapatan tahunan, dan skor pengeluaran. Tahapan ini memastikan seluruh data terstandarisasi dan siap untuk dianalisis lebih lanjut.

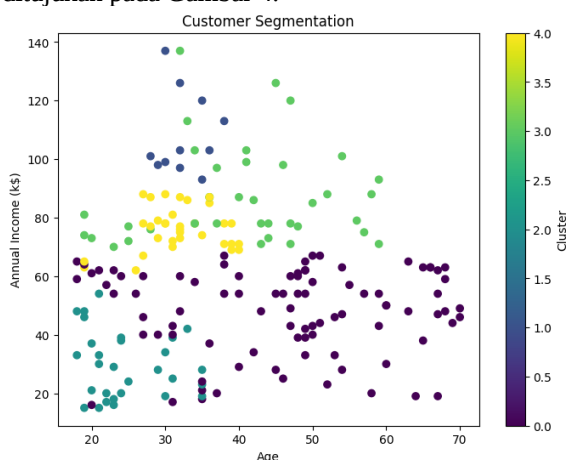
Penentuan jumlah cluster optimal dilakukan menggunakan bantuan grafik metode Elbow ditujukan pada Gambar 3.



Gambar 3. grafik metode Elbow

Algoritma K-Means dijalankan secara iteratif dengan nilai K dari satu hingga sepuluh. Nilai inersia dari setiap iterasi dihitung untuk melihat variansi di dalam setiap kelompok. Titik patahan pada grafik Elbow menunjukkan penurunan drastis nilai inersia dari K=1 hingga K=5. Kurva inersia mulai melandai secara bertahap pada K=6 dan seterusnya. Lima cluster ditetapkan sebagai parameter paling ideal berdasarkan observasi visual terhadap grafik tersebut. Algoritma K-Means diterapkan secara definitif menggunakan parameter lima cluster. Label kelompok pelanggan dari proses prediksi ditambahkan sebagai kolom baru bernama 'Cluster' ke dalam dataset.

Kualitas hasil pengelompokan dievaluasi secara matematis menggunakan metrik Silhouette Score ditujukan pada Gambar 4.



Gambar 4. metrik Silhouette Score

Hasil perhitungan metrik tersebut mencatatkan skor sebesar 0,3575963460506624. Angka tersebut mengonfirmasi validitas batasan pemisah antar kelima kelompok pelanggan. Visualisasi scatter plot memetakan persebaran kelima kelompok pelanggan secara dua dimensi. Grafik tersebut menampilkan korelasi antara variabel umur pada sumbu X dan pendapatan tahunan pada sumbu Y. Pemetaan warna menggunakan colormap viridis membedakan kelima cluster pelanggan secara visual. Persebaran titik pada grafik menunjukkan adanya lima kelompok utama yang tersebar berdasarkan pola pendapatan dan usia. Kelompok pelanggan usia muda dengan pendapatan menengah hingga tinggi direkomendasikan untuk ditargetkan dengan produk premium berskala tren masa kini. Kelompok pelanggan lansia dengan pendapatan bervariasi dapat dijangkau menggunakan strategi promosi yang lebih konservatif. Temuan profil demografis ini membuktikan efektivitas segmentasi K-Means dalam memfasilitasi perumusan strategi pemasaran yang lebih relevan dan tepat sasaran.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil menerapkan algoritma K-Means untuk menyegmentasi pelanggan berdasarkan data demografis dari dataset Mall Customer Segmentation. Metode Elbow secara efektif menetapkan lima cluster sebagai jumlah kelompok paling optimal untuk mengkategorikan konsumen. Evaluasi kualitas model mencatatkan nilai Silhouette Score sebesar 0,3575963460506624 yang mengonfirmasi keberhasilan pemisahan pelanggan berdasarkan variabel umur, pendapatan tahunan, dan skor pengeluaran. Profil unik dari setiap segmen tersebut direkomendasikan sebagai landasan utama bagi manajemen ritel dalam merancang kampanye pemasaran berskala personal secara spesifik. Peneliti selanjutnya disarankan memperluas ruang lingkup riset dengan menambahkan variabel perilaku transaksi historis untuk memperkaya dimensi segmentasi. Penggunaan dataset berukuran lebih besar sangat diperlukan guna menguji tingkat kestabilan algoritma pada studi mendatang.

DAFTAR PUSTAKA

- [1] O. H. Olayinka, "Data driven customer segmentation and personalization strategies in modern business intelligence frameworks," *World J. Adv. Res. Rev.*, vol. 12, no. 03, pp. 711–726, 2021.
- [2] K. Tabianan, "K-Means clustering for e-commerce customer segmentation," *Sustainability*, vol. 14, no. 21, 2022.
- [3] N. Hidayat, D. P. Gevano, A. Putra, and R. Ilahi, "Segmentasi Pelanggan Menggunakan K-Means Clustering Berdasarkan Data Kepribadian dan Pola Konsumsi," *J. Tek. Inform.*, vol. 6, no. 5, pp. 3914–3924, 2025.
- [4] G. K. Siri and I. Technology, "CUSTOMER SEGMENTATION IN MALLS," *ijcspub*, vol.

- 13, no. 4, pp. 473–481, 2023.
- [5] J. M. John, “An exploration of clustering algorithms for customer segmentation in the UK retail market,” *Mathematics*, vol. 11, no. 22, 2023.
- [6] K. Tabianan and S. Velu, “K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data,” *Sustainability*, vol. 14, no. 12, pp. 1–15, 2022.
- [7] R. C. Classification, “SS symmetry,” *symmetry S Artic.*, vol. 13, no. 10, pp. 1–15, 2021.
- [8] A. Kester *et al.*, “Consumer preference analysis on the attributes of samgyeopsal Korean cuisine and its market segmentation: Integrating conjoint analysis and K-means clustering,” *PLoS One*, vol. 18, no. 2, pp. 1–23, 2023, doi: 10.1371/journal.pone.0281948.
- [9] M. Carvache-franco, T. Hassan, O. Carvache-franco, W. Carvache-, F. Id, and O. M. Id, “Demand segmentation and sociodemographic aspects of food festivals: A study in Bahrain,” *PLoS One*, vol. 18, no. 6, pp. 1–17, 2023, doi: 10.1371/journal.pone.0287113.
- [10] R. M. Fauzan and G. Alfian, “Segmentasi Pelanggan E-Commerce Menggunakan Fitur Recency, Frequency, Monetary (RFM) dan Algoritma Klasterisasi K-Means,” *JISKA*, vol. 9, no. 3, pp. 170–177, 2024.
- [11] 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. 3, 2025.
- [12] J. Lin, “Application of machine learning in predicting consumer behavior and precision marketing,” *PLoS One*, vol. 20, no. 5, pp. 1–12, 2025, doi: 10.1371/journal.pone.0321854.
- [13] A. Griva, “A two-stage business analytics approach to perform behavioural and geographic customer segmentation using e-commerce delivery data,” *J. Bus. Anal.*, vol. 7, no. 4, 2024.
- [14] W. Xingqiong, “Adaptive density peak clustering based on Delaunay graph,” *PLoS One*, vol. 20, no. 5, 2025.
- [15] G. Vo and D. Huynh, “Customer Segmentation: Automatic K-Optimization and RFM-Based K-Means Clustering,” *ACM*, vol. 1, no. 1, pp. 1–10, 2025.
- [16] M. F. Fadhillah, A. Lovely, A. Suyoso, and I. Puspitasari, “Customer Segmentation with Clustering Algorithm Based on Recency, Frequency, and Monetary (RFM) Attributes Segmentasi Pelanggan dengan Algoritma Clustering Berdasarkan Atribut Recency, Frequency dan Monetary (RFM),” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 5, no. 1, pp. 48–56, 2025.
- [17] A. Wicaksono, F. A. Bachtiar, and N. Y. Setiawan, “Segmentasi Pelanggan Menggunakan Fuzzy C-Means Clustering berdasarkan RFM Model pada E-Commerce (Studi Kasus: E-Commerce XYZ),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, pp. 1351–1360, 2021.
- [18] A. Julian, A. Faqih, and G. Dwilestari, “Segmentasi Pelanggan Menggunakan Algoritma K-Means Pada Jaringan Telekomunikasi Untuk Optimalisasi Strategi Pemasaran,” *J. Teknol. TERPADU*, vol. 13, no. 1, 2025.
- [19] L. Wicaksono, “Penerapan Metode K-Means Clustering untuk Menentukan Kepuasan Mahasiswa,” *Bridge*, vol. 3, no. 1, 2024.