

Segmentasi Pelanggan E-Commerce Berbasis Perilaku Belanja Menggunakan Algoritma K-Means Clustering dan Evaluasi Silhouette Score

Muhammad Yusran¹, Rizkah Fadillah^{2*}, Mesran³, Raymond Shawn²

¹Falkutas Ilmu Komputer dan Teknologi Informasi, Prodi Teknik Informatika, Universitas Budi Darma, Medan, Indonesia

²Prodi Teknologi Rekayasa Komputer Grafis, Politeknik Cendana, Medan, Indonesia

³Prodi Manajemen, Sekolah Tinggi Ilmu Manajemen Sukma Medan, Medan, Indonesia

Email: ¹myusran804@gmail.com, ^{2,*}fadillahrizkah@gmail.com, ³mesran.skompom@gmail.com, ⁴raymonds6@gmail.com

Email Penulis Korespondensi: fadillahrizkah@gmail.com

Abstrak—Dalam era persaingan E-Commerce yang ketat, pemahaman mendalam mengenai perilaku konsumen menjadi kunci utama keberhasilan strategi pemasaran. Penelitian ini bertujuan untuk melakukan segmentasi pelanggan toko online menggunakan algoritma K-Means Clustering berbasis perilaku belanja. Menggunakan dataset Mall Customer Segmentation, penelitian ini memproses fitur utama berupa Pendapatan Tahunan dan Skor Pengeluaran yang dinormalisasi. Evaluasi model dilakukan menggunakan metode Elbow dan Silhouette Score untuk menentukan jumlah kelompok optimal. Hasil penelitian menunjukkan bahwa pembagian pelanggan menjadi 5 cluster adalah yang paling optimal dengan nilai Silhouette Score sebesar 0.5547. Kelima segmen yang terbentuk meliputi Middle Class (40,5%), VIP/Big Spenders (19,5%), Hemat Mapan (17,5%), Ekonomis (11,5%), dan Impulsif (11%). Temuan ini memberikan wawasan strategis bagi pelaku bisnis untuk merancang kampanye pemasaran yang terpersonalisasi, seperti program loyalitas eksklusif untuk segmen VIP dan penawaran diskon intensif untuk segmen Ekonomis, guna meningkatkan retensi dan profitabilitas.

Kata Kunci : Data Mining; E-Commerce; K-Means Clustering; Segmentasi Pelanggan; Silhouette Score

Abstract—In the era of intense e-commerce competition, a deep understanding of consumer behavior is key to a successful marketing strategy. This study aims to segment online store customers using the K-Means Clustering algorithm based on shopping behavior. Using the Mall Customer Segmentation dataset, this study processes key features in the form of normalized Annual Income and Spending Score. Model evaluation was conducted using the Elbow and Silhouette Score methods to determine the optimal number of clusters. The results showed that dividing customers into five clusters was the most optimal, with a Silhouette Score of 0.5547. The five segments formed include Middle Class (40.5%), VIP/Big Spenders (19.5%), Thrifty Established (17.5%), Economical (11.5%), and Impulsive (11%). These findings provide strategic insights for businesses to design personalized marketing campaigns, such as exclusive loyalty programs for the VIP segment and intensive discount offers for the Economic segment, to increase retention and profitability.

Keywords: Data Mining; E-Commerce; K-Means Clustering; Customer Segmentation; Silhouette Score

1. PENDAHULUAN

Teknologi informasi telah mengubah lanskap perdagangan global, menggeser pola transaksi konvensional menuju perdagangan elektronik atau E-Commerce. Pertumbuhan ini menghasilkan akumulasi data transaksi dalam jumlah yang sangat besar (Big Data), yang menyimpan potensi wawasan strategis bagi keberlangsungan bisnis. Namun, tantangan utama yang dihadapi oleh para pelaku bisnis E-Commerce saat ini bukanlah ketersediaan data, melainkan kemampuan untuk mengolah data mentah tersebut menjadi informasi yang dapat ditindaklanjuti [1],[2]. Dalam basis pelanggan yang sangat heterogen, menerapkan strategi pemasaran yang seragam (one-size-fits-all) terbukti tidak efektif dan seringkali memboroskan anggaran, karena setiap pelanggan memiliki preferensi, daya beli, dan pola perilaku yang berbeda-beda.

Untuk mengatasi permasalahan tersebut, segmentasi pelanggan menjadi strategi krusial dalam manajemen pemasaran modern. Segmentasi bertujuan untuk mengelompokkan pelanggan ke dalam segmen-segmen yang homogen berdasarkan karakteristik tertentu, sehingga perusahaan dapat menerapkan strategi pemasaran yang lebih terarah dan terpersonalisasi [3]. Salah satu pendekatan yang paling efektif dalam memetakan profil pelanggan adalah dengan menganalisis perilaku belanja historis mereka, yang seringkali diukur melalui parameter Recency (rentang waktu sejak transaksi terakhir), Frequency (seberapa sering bertransaksi), dan Monetary (total nilai uang yang dikeluarkan). Analisis ini memungkinkan perusahaan untuk mengidentifikasi siapa pelanggan yang paling loyal, pelanggan yang boros (big spenders), atau pelanggan yang berisiko berhenti berlangganan (churn) [4],[5].

Mengingat besarnya volume data yang terlibat, segmentasi manual menjadi mustahil dilakukan. Oleh karena itu, pendekatan Data Mining menggunakan algoritma Machine Learning menjadi solusi yang relevan. Penelitian ini menerapkan algoritma K-Means Clustering, sebuah metode unsupervised learning yang populer karena efisiensinya dalam mengelompokkan data tanpa label ke dalam sejumlah cluster berdasarkan kemiripan fitur. Algoritma ini bekerja dengan meminimalkan jarak varians di dalam cluster untuk memastikan anggota kelompok memiliki karakteristik yang sangat mirip satu sama lain [6],[7],[8].

Meskipun K-Means merupakan algoritma yang handal, tantangan teknis dalam penggunaannya terletak pada penentuan jumlah cluster yang optimal dan validasi kualitas pengelompokan yang terbentuk. Pengelompokan yang buruk dapat menghasilkan wawasan yang bias dan menyesatkan strategi bisnis. Oleh karena itu, penelitian ini tidak hanya berfokus pada pembentukan segmen, tetapi juga menekankan pada evaluasi kinerja model menggunakan metode Silhouette Score.

Metrik ini digunakan untuk mengukur seberapa baik sebuah objek ditempatkan dalam cluster-nya sendiri dibandingkan dengan cluster lain, guna memastikan bahwa segmen yang terbentuk memiliki tingkat kepadatan dan pemisahan (separation) yang valid secara matematis [9],[10].

Tampubolon et al. pada tahun 2024 menerapkan metode RFM yang dipadukan dengan algoritma K-Means++ untuk melakukan segmentasi pelanggan pada platform E-Commerce. Penelitian tersebut menggunakan Silhouette Score sebagai indikator dalam menentukan jumlah cluster terbaik dan memperoleh tiga cluster optimal dengan nilai Silhouette sebesar 0,638. Hasil penelitian menunjukkan bahwa kombinasi RFM dan algoritma clustering mampu menghasilkan segmentasi pelanggan yang mudah diinterpretasikan dan mendukung pengambilan keputusan bisnis [11].

Selanjutnya, Wu pada tahun 2022 mengembangkan model RFM dengan algoritma clustering yang telah ditingkatkan untuk mengatasi kelemahan metode konvensional dalam menentukan jumlah cluster. Penelitian tersebut menunjukkan bahwa optimasi algoritma mampu meningkatkan akurasi segmentasi pelanggan E-Commerce sehingga informasi yang dihasilkan lebih representatif terhadap perilaku pelanggan [12].

Penelitian yang dilakukan oleh Lukas et al. pada 2025 mengintegrasikan pemodelan RFM dengan algoritma K-Means pada data penjualan online untuk menghasilkan segmentasi pelanggan sebagai dasar penyusunan strategi pemasaran. Hasil penelitian membuktikan bahwa pendekatan tersebut mampu mengidentifikasi kelompok pelanggan dengan karakteristik yang berbeda sehingga dapat dimanfaatkan dalam penyusunan program retensi maupun promosi yang lebih efektif [13].

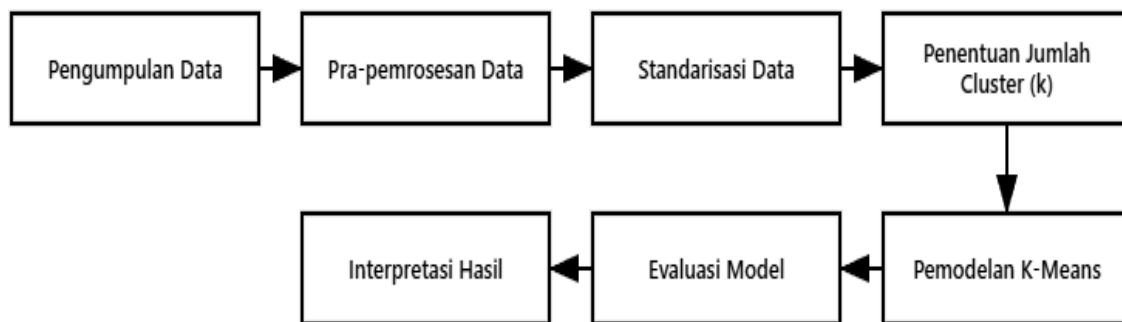
Berdasarkan uraian tersebut, penelitian ini bertujuan untuk melakukan segmentasi pelanggan pada dataset E-Commerce (bersumber dari repositori Kaggle) menggunakan algoritma K-Means Clustering. Melalui analisis ini, diharapkan dapat terbentuk profil-profil pelanggan yang jelas seperti kelompok pelanggan hemat, loyal, maupun impulsif yang divalidasi ketepatannya melalui Silhouette Score, sehingga dapat menjadi landasan pengambilan keputusan strategis bagi manajemen pemasaran toko online.

2. METODOLOGI PENELITIAN

Proses penelitian dilakukan melalui beberapa tahapan yang saling berhubungan, dimulai dari pengolahan awal data suara hingga evaluasi model.

2.1 Tahapan Penelitian

Penelitian ini dilaksanakan mengikuti alur kerja Knowledge Discovery in Database (KDD) yang telah disesuaikan [14] dengan kebutuhan analisis clustering. Alur tahapan penelitian yang dilakukan disajikan pada Gambar 1 berikut.



Gambar 1. Tahapan Penelitian

Proses pada Gambar 1 diawali dengan pengumpulan data transaksi historis dari repositori publik (Kaggle), yang kemudian dilanjutkan ke tahap pra-pemrosesan data untuk menangani missing values dan ketidakkonsistenan format. Setelah fitur terbentuk, dilakukan standarisasi data guna menyeimbangkan skala antar variabel. Tahap ini melibatkan penerapan algoritma K-Means untuk membentuk segmen pelanggan secara iteratif. Proses penelitian diakhiri dengan evaluasi model menggunakan metode Elbow dan Silhouette Score untuk menentukan jumlah kelompok yang paling optimal serta validasi karakteristik dari setiap segmen yang terbentuk.

2.2 Algoritma K-Means Clustering

Algoritma K-Means digunakan untuk mempartisi himpunan data pelanggan ke dalam K kelompok yang berbeda, di mana setiap pengamatan termasuk dalam kelompok dengan rata-rata terdekat (pusat cluster atau centroid). Algoritma ini bekerja secara iteratif untuk meminimalkan varians intra-kelompok, yang dikenal sebagai Sum of Squared Errors (SSE) [15],[16].

Fungsi objektif (J) dari algoritma K-Means yang akan diminimalkan dalam penelitian ini didefinisikan sebagai jumlah kuadrat jarak Euclidean antara setiap titik data pelanggan (x_i) dengan pusat cluster terdekatnya (c_j). Rumus fungsi objektif tersebut adalah:

$$J = \sum_{j=1}^K \sum_{i=1}^n ||x_i^{(j)} - c_j||^2 \quad (1)$$

Dimana $||x_i^{(j)} - c_j||^2$ adalah jarak *Euclidean* antara titik data x_i dan *centroid* c_j . Proses iterasi akan terus berlangsung dengan memperbarui posisi centroid hingga posisi tersebut konvergen atau tidak ada perubahan signifikan pada keanggotaan cluster.

2.3 Silhouette Score

Untuk memvalidasi kualitas segmentasi dan menentukan jumlah cluster (K) yang paling optimal, penelitian ini menggunakan metrik Silhouette Score. Metode ini mengukur seberapa mirip suatu objek dengan cluster-nya sendiri (cohesion) dibandingkan dengan cluster lain (separation)[17][18].

Nilai Silhouette (s) untuk sebuah titik data tunggal i dihitung dengan mempertimbangkan rata-rata jarak intra-kuster (a(i)) dan rata-rata jarak ke cluster tetangga terdekat (b(i)). Formulasi matematis untuk Silhouette Coefficient adalah:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (2)$$

Nilai a(i) merepresentasikan rata-rata ketidakmiripan i dengan semua titik lain dalam cluster yang sama (semakin kecil semakin baik), sedangkan b(i) adalah ketidakmiripan terendah rata-rata i terhadap cluster lain (semakin besar semakin baik). Nilai Silhouette Score berkisar antara -1 hingga +1. Nilai yang mendekati +1 mengindikasikan bahwa pelanggan telah terkelompokkan dengan sangat baik dalam segmen yang tepat dan terpisah jauh dari segmen lainnya. Rata-rata dari s(i) untuk seluruh data digunakan sebagai penentu kualitas global model segmentasi yang terbentuk.

3. HASIL DAN PEMBAHASAN

Bab ini memaparkan hasil analisis data dan interpretasi mendalam mengenai segmen pelanggan yang terbentuk. Proses dimulai dari eksplorasi data awal, penentuan jumlah cluster optimal secara statistik, hingga pemprofilan karakteristik setiap kelompok pelanggan yang diidentifikasi oleh algoritma K-Means.

3.1 Pre-Processing Data

Tahap awal penelitian melibatkan pemeriksaan integritas dataset Mall Customer Segmentation. Berdasarkan hasil pemeriksaan struktur data, dataset terdiri dari 200 entri transaksi tanpa adanya missing values pada seluruh variabel (CustomerID, Gender, Age, Annual Income, dan Spending Score), sehingga teknik imputasi data tidak diperlukan dalam penelitian ini (**Lihat Gambar 2**).

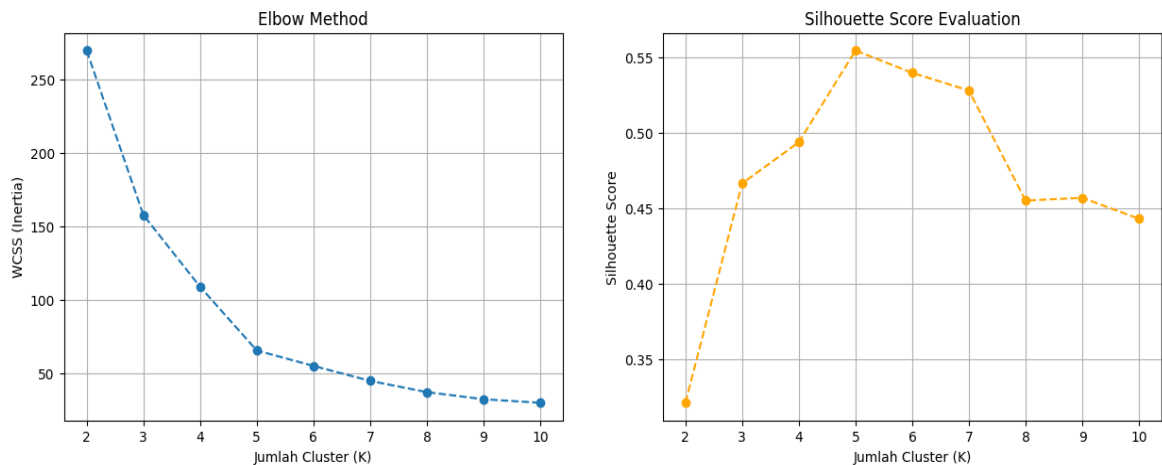
Info Dataset:				Jumlah Missing Values:	
<class 'pandas.core.frame.DataFrame'>				CustomerID	0
RangeIndex: 200 entries, 0 to 199				Gender	0
Data columns (total 5 columns):				Age	0
#	Column	Non-Null Count	Dtype	Annual Income (k\$)	0
0	CustomerID	200 non-null	int64	Spending Score (1-100)	0
1	Gender	200 non-null	object		
2	Age	200 non-null	int64		
3	Annual Income (k\$)	200 non-null	int64		
4	Spending Score (1-100)	200 non-null	int64		
dtypes: int64(4), object(1)					
memory usage: 7.9+ KB					

Gambar 2. Info Dataset dan Pengecekan Missing Values

Berdasarkan Gambar 2, fokus analisis diarahkan pada dua variabel kunci yang merepresentasikan daya beli dan perilaku konsumsi, yaitu Annual Income (k\$) dan Spending Score (1-100). Sebelum proses pemodelan, kedua fitur ini dinormalisasi menggunakan teknik StandardScaler. Langkah ini terbukti krusial karena data asli memiliki rentang yang timpang, di mana Annual Income memiliki rentang hingga ratusan ribu dolar sementara Spending Score terbatas pada skala 1 hingga 100. Standarisasi berhasil mentransformasi data ke dalam skala distribusi normal (Z-score) di sekitar angka nol, memastikan bahwa algoritma K-Means tidak bias terhadap variabel dengan nominal yang lebih besar.

3.2 Penentuan Jumlah Cluster Optimal

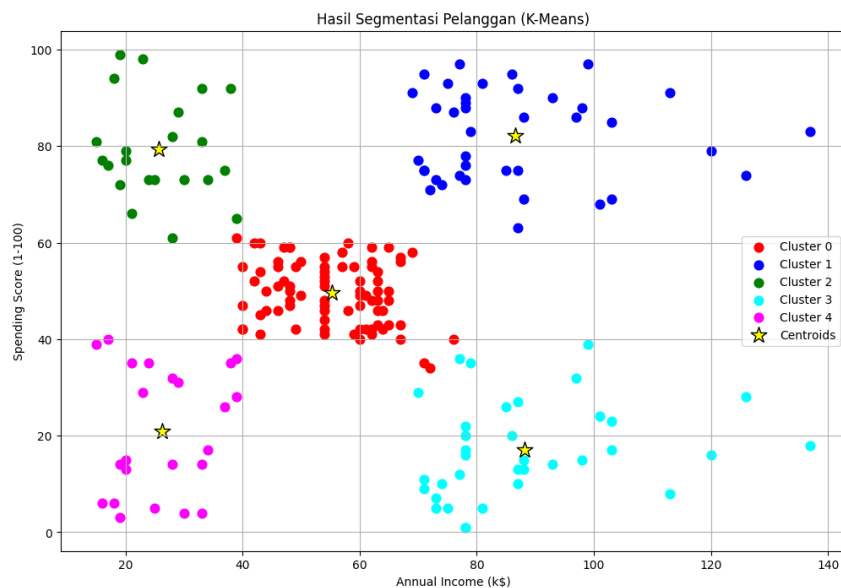
Salah satu tantangan utama dalam Unsupervised Learning adalah menentukan jumlah kelompok yang paling representatif. Penelitian ini menggunakan pendekatan ganda, yaitu Elbow Method untuk melihat inersia dan Silhouette Score untuk validasi kepadatan cluster yang disajikan pada Gambar 3 berikut.



Gambar 3. Evaluasi Jumlah Cluster - Elbow Method dan Silhouette Score

3.3 Analisis Profil Segmen Pelanggan

Penerapan algoritma K-Means dengan $K=5$ menghasilkan pembagian pelanggan yang tervisualisasi dengan jelas dalam ruang dua dimensi. Sebaran data menunjukkan adanya lima kelompok yang terpisah tanpa banyak tumpang tindih (overlap), membuktikan efektivitas model dalam membedakan pola belanja (Lihat Gambar 4)[19].



Gambar 4. Visualisasi Scatter Plot Hasil Segmentasi

Untuk memperoleh gambaran karakteristik masing-masing cluster, dilakukan proses profiling dengan menghitung nilai centroid berupa rata-rata Annual Income dan Spending Score pada setiap kelompok pelanggan. Selain itu, hasil evaluasi menggunakan Silhouette Score juga ditampilkan untuk menunjukkan kualitas pengelompokan yang dihasilkan oleh algoritma K-Means. Hasil profiling dan evaluasi model ditunjukkan pada Gambar 5.

Silhouette Score Final untuk K=5: 0.5547

=== Profiling Cluster (Rata-rata per Kelompok) ===

Cluster	Jumlah Pelanggan	Rata-rata Income	Rata-rata Spend Score
0	81	55.296296	49.518519
1	39	86.538462	82.128205
2	22	25.727273	79.363636
3	35	88.200000	17.114286
4	23	26.304348	20.913043

Gambar 5. Statistik Profiling Cluster dan Skor Silhouette Akhir

Berdasarkan Gambar 5, algoritma K-Means berhasil membentuk lima cluster pelanggan dengan karakteristik yang berbeda. Perbedaan tersebut terlihat dari jumlah anggota, nilai centroid pada variabel Annual Income dan Spending Score, serta interpretasi bisnis masing-masing cluster. Hasil ini menunjukkan bahwa pelanggan pada dataset memiliki pola perilaku yang beragam sehingga memerlukan strategi pemasaran yang berbeda untuk setiap segmen.

1. Cluster 0: Segmen Menengah (Middle Class)

Kelompok ini merupakan segmen terbesar dengan jumlah anggota sebanyak 81 pelanggan. Secara statistik, kelompok ini memiliki rata-rata Pendapatan Tahunan sebesar 55.3k\$ dan Skor Pengeluaran rata-rata 49.5. Posisi mereka berada tepat di tengah grafik scatter plot (titik berwarna merah). Profil ini merepresentasikan pelanggan "standar" yang berbelanja sesuai dengan kemampuan finansial mereka—tidak terlalu hemat, namun juga tidak boros. Stabilitas kelompok ini menjadikannya target yang aman untuk promosi produk-produk mainstream.

2. Cluster 1: Segmen Target Utama (VIP/Big Spenders)

Cluster 1 (titik berwarna biru) adalah kelompok yang paling bernilai bagi perusahaan. Terdiri dari 39 pelanggan, segmen ini memiliki kombinasi ideal: rata-rata Pendapatan Tahunan yang tinggi sebesar 86.5k\$ dan Skor Pengeluaran yang sangat tinggi mencapai 82.1. Karakteristik ini menunjukkan bahwa mereka adalah pelanggan mapan yang loyal dan tidak ragu mengeluarkan uang dalam jumlah besar. Strategi pemasaran untuk kelompok ini harus berfokus pada layanan eksklusif dan produk premium untuk menjaga retensi[20].

3. Cluster 2: Segmen Impulsif (Carefree)

Segmen ini menarik karena anomali perilakunya. Meskipun memiliki rata-rata Pendapatan Tahunan terendah di antara semua kelompok, yaitu hanya 25.7k\$, mereka memiliki Skor Pengeluaran yang sangat tinggi, mencapai rata-rata 79.4. Dengan jumlah anggota 22 pelanggan, kelompok ini (titik hijau) kemungkinan besar didominasi oleh kalangan muda atau pelajar yang mementingkan gaya hidup di atas kemampuan finansial jangka panjang. Mereka sangat responsif terhadap tren, namun memiliki risiko kredit yang lebih tinggi.

4. Cluster 3: Segmen Hemat Mapan (Frugal)

Berbanding terbalik dengan Cluster 1, Cluster 3 (titik cyan) yang terdiri dari 35 pelanggan memiliki daya beli tinggi dengan rata-rata pendapatan 88.2k\$, namun sangat menahan pengeluaran mereka, terbukti dari Skor Pengeluaran yang rendah di angka 17.1. Kelompok ini merepresentasikan pelanggan yang sangat rasional, berhati-hati, atau mungkin kurang puas dengan produk yang ditawarkan saat ini. Pendekatan untuk segmen ini memerlukan strategi "value for money" atau penawaran investasi yang meyakinkan untuk memicu spending mereka.

5. Cluster 4: Segmen Ekonomis (Budget Conscious)

Kelompok terakhir (titik ungu/pink) terdiri dari 23 pelanggan dengan karakteristik Pendapatan Tahunan rendah (26.3k\$) dan Skor Pengeluaran yang juga rendah (20.9). Perilaku belanja mereka sangat terbatas oleh batasan finansial. Dalam konteks pemasaran, segmen ini bukan prioritas utama untuk produk premium, namun bisa menjadi target yang efektif untuk program diskon, flash sale, atau produk entry-level.

3.4 Pembahasan

Hasil segmentasi menunjukkan bahwa algoritma K-Means mampu mengelompokkan pelanggan berdasarkan kemiripan karakteristik pendapatan dan perilaku pengeluaran sehingga tujuan penelitian untuk memperoleh profil pelanggan berhasil dicapai. Setiap cluster memiliki karakteristik yang berbeda sehingga memerlukan pendekatan pemasaran yang berbeda pula. Segmentasi ini dapat membantu perusahaan dalam mengalokasikan sumber daya pemasaran secara lebih efektif dibandingkan menerapkan strategi yang sama kepada seluruh pelanggan.

Dari perspektif bisnis e-commerce, pelanggan pada Cluster 1 dapat dijadikan prioritas dalam program loyalitas, pemberian voucher eksklusif, maupun penawaran produk premium karena memiliki nilai transaksi yang tinggi. Sebaliknya, Cluster 3 menunjukkan adanya pelanggan dengan kemampuan finansial tinggi namun tingkat pembelian yang rendah. Kelompok ini berpotensi ditingkatkan nilai transaksinya melalui strategi personalisasi promosi, rekomendasi produk, maupun bundling produk. Sementara itu, pelanggan pada Cluster 2 lebih responsif terhadap tren sehingga strategi flash sale, limited offer, dan promosi media sosial diperkirakan lebih efektif. Adapun Cluster 4 lebih sesuai diberikan promosi berbasis harga,

seperti diskon atau cashback, sedangkan Cluster 0 dapat dipertahankan melalui promosi rutin untuk menjaga loyalitas pelanggan.

Hasil penelitian ini menunjukkan bahwa algoritma K-Means berhasil membentuk lima cluster pelanggan dengan nilai Silhouette Score sebesar 0,5547, yang menunjukkan bahwa kualitas cluster berada pada kategori cukup baik karena memiliki tingkat kohesi dan separasi yang memadai. Hasil tersebut membuktikan bahwa algoritma K-Means mampu mengidentifikasi kelompok pelanggan berdasarkan kemiripan karakteristik pendapatan dan perilaku belanja sehingga dapat digunakan sebagai dasar penyusunan strategi pemasaran yang lebih tepat sasaran.

Temuan penelitian ini sejalan dengan penelitian Tampubolon et al. (2024) [11] yang menerapkan metode RFM dan K-Means++ untuk segmentasi pelanggan e-commerce. Penelitian tersebut menunjukkan bahwa algoritma clustering mampu menghasilkan kelompok pelanggan dengan karakteristik yang berbeda dan memperoleh nilai Silhouette Score sebesar 0,638, yang menunjukkan kualitas cluster yang baik. Kesamaan hasil tersebut menunjukkan bahwa algoritma berbasis K-Means merupakan metode yang efektif dalam melakukan segmentasi pelanggan.

Hasil penelitian ini juga mendukung penelitian Wu (2022) [12] yang menyatakan bahwa teknik clustering mampu mengidentifikasi pola perilaku pelanggan sehingga perusahaan dapat menerapkan strategi pemasaran yang lebih personal. Dalam penelitian Wu, fokus utama diarahkan pada pengembangan algoritma clustering untuk meningkatkan kualitas segmentasi, sedangkan penelitian ini lebih menitikberatkan pada interpretasi bisnis dari setiap cluster yang dihasilkan.

Selain itu, penelitian ini memiliki kesamaan dengan penelitian Lukas et al. (2025) [13] yang memanfaatkan algoritma K-Means untuk mengelompokkan pelanggan berdasarkan karakteristik transaksi. Kedua penelitian sama-sama menghasilkan beberapa kelompok pelanggan yang memiliki perilaku pembelian berbeda sehingga dapat dimanfaatkan sebagai dasar penyusunan strategi pemasaran. Perbedaannya, penelitian ini memberikan interpretasi yang lebih rinci terhadap setiap cluster berdasarkan kombinasi pendapatan dan tingkat pengeluaran sehingga rekomendasi strategi pemasaran yang dihasilkan menjadi lebih spesifik.

Secara keseluruhan, penelitian ini memperkuat temuan-temuan sebelumnya bahwa algoritma K-Means merupakan metode yang efektif untuk segmentasi pelanggan. Kontribusi penelitian ini terletak pada penggunaan evaluasi Silhouette Score sebagai dasar penentuan jumlah cluster optimal serta penyajian interpretasi bisnis yang lebih komprehensif terhadap setiap segmen pelanggan sehingga hasil segmentasi dapat langsung dimanfaatkan dalam pengambilan keputusan pemasaran pada platform e-commerce.

4. KESIMPULAN

Penerapan algoritma K-Means Clustering pada dataset pelanggan berhasil mengungkap pola segmentasi pelanggan secara efektif dan berbasis data. Dengan dukungan pra-pemrosesan berupa standarisasi fitur, algoritma ini mampu menghasilkan cluster yang terpisah jelas dan mudah diinterpretasikan. Evaluasi menggunakan metode Elbow dan Silhouette Score menunjukkan jumlah cluster optimal sebanyak lima, dengan nilai Silhouette Score sebesar 0,5547 yang menandakan kualitas pengelompokan berada pada tingkat menengah. Lima cluster yang terbentuk merepresentasikan karakteristik pelanggan yang berbeda, mulai dari pelanggan kelas menengah, pelanggan bernilai tinggi (VIP), pelanggan impulsif, pelanggan berpendapatan tinggi namun hemat, hingga pelanggan ekonomis yang sensitif terhadap harga. Hasil segmentasi ini memberikan dasar yang kuat bagi perusahaan E-Commerce untuk menerapkan strategi pemasaran terarah, sehingga promosi dan penawaran dapat disesuaikan dengan profil masing-masing segmen. Pendekatan ini diharapkan mampu meningkatkan efektivitas pemasaran, kepuasan pelanggan, serta loyalitas jangka panjang.

REFERENCES

- [1] L. Wang, T. R. A. L. Pertheban, T. Li, and L. Zhao, "Application of business intelligence based on big data in E-commerce data evaluation," *Heliyon*, vol. 10, no. 21, p. e38768, 2024.
- [2] V. M. Reddy and L. N. Nalla, "Leveraging Big Data Analytics to Enhance Customer Experience in E-commerce," *Rev. Esp. Doc. Cient.*, vol. 18, no. 02, pp. 295–324, 2024.
- [3] J. M. John, O. Shobayo, and B. Ogunleye, "An Exploration of Clustering Algorithms for Customer Segmentation in the UK Retail Market," *Analytics*, vol. 2, no. 4, pp. 1–16, 2023, doi: 10.3390/analytics2040042.
- [4] W. Liang, Y. Zhang, M. Codreanu, J. Wang, H. Cao, and J. Zou, "The widespread adoption of large language model-assisted writing across society," *Patterns*, vol. 6, no. 12, pp. 1–23, 2025, doi: 10.1016/j.patter.2025.101366.
- [5] A. Y. N. Rumapea, D. Pratiwi, and S. Sari, "Analisis segmentasi pelanggan ritel online menggunakan K-Means clustering berdasarkan model Recency, Frequency, Monetary (RFM)," *J. Sains dan Teknol.*, vol. 6, no. 3, pp. 292–299, 2024.
- [6] K. Tabianan, S. Velu, and V. Ravi, "K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data," *Sustainability*, vol. 14, no. 12, p. 7243, 2022, doi: 10.3390/su14127243.
- [7] A. Putera, U. Siahaan, N. A. Harahap, and R. Y. Simanullang, "Analysis of Inpatient Data Using Cluster Analysis on Simulation Dataset," *J. BIT (Bulletin Inf. Technol.)*, vol. 6, no. 1, pp. 33–39, 2025, doi: 10.47065/bit.v5i2.1830.
- [8] L. Erlistya, P. S. Sumarna, and ..., "Implementasi Algoritma K-Means Clustering pada Studi Kasus Seleksi Mahasiswa Berprestasi

- menggunakan Visual Basic,” *Pros. Semin.*, vol. 4, no. 2721, 2021.
- [9] M. R. P. Pratama *et al.*, “Perbandingan Algoritma K-Means, K-Medoid, dan DbSCAN Untuk Clustering Kualitas Hidup Indonesia Dalam Perspektif Knowledge Management Dan Data Discovery,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 4, pp. 5903–5910, 2025, doi: 10.36040/jati.v9i4.13916.
 - [10] R. Y. Simanullang, P. Wanny, and S. M. Rambe, “Implementation of the K-Means Algorithm on Smart Systems in Grouping Gadget Accessory Purchase Patterns,” *JITCSE (Journal Inf. Technol. Comput. Sci. Electr. Eng.*, vol. 2, no. 3, pp. 22–31, 2026, doi: 0.61306/jitcse.
 - [11] A. L. M. Tampubolon, T. M. E. Y. B. Butar, and S. Rochimah, “Segmentasi Pelanggan Majalah Pada Situs Web E-Commerce Segmentation of Magazine Subscribers On E-Commerce Website using K-Means ++ and RFM Method,” *J. Teknol. Inf. dan ilmu Komput.*, vol. 11, no. 6, pp. 1243–1252, 2024, doi: 10.25126/jtiik.2024118208.
 - [12] Z. Wu, L. Jin, J. Zhao, L. Jing, and L. Chen, “Research on Segmenting E-Commerce Customer through an Improved K-Medoids Clustering Algorithm,” *Comput. Intell. Neurosci.*, vol. 2022, no. 1, p. 9930613, 2022, doi: 10.1155/2022/9930613.
 - [13] I. L. Pratama *et al.*, “Pemodelan RFM & K-Means Clustering untuk Segmentasi Pelanggan dalam Penjualan Online,” *J. Softw. Eng. Inf. Syst.*, vol. 5, no. 2, pp. 49–58, 2025, doi: 10.37859/seis.v5i2.9556.
 - [14] V. S. Kirubakaran and S. G., “Customer Segmentation using Clustering Techniques: Data-Driven Approach to Enhance Marketing Strategy,” in *2024 International Conference on Sustainable Communication Networks and Application (ICSCNA)*, 2024, pp. 1–6. doi: 10.1109/ICSCNA63714.2024.10864053.
 - [15] Y. Putri, D. Aldo, and W. Ilham, “Retail Marketing Strategy Optimization: Customer Segmentation with Artificial Intelligence Integration and K-Means Clustering,” *Sink. J. dan Penelit. Tek. Inform.*, vol. 8, no. 1, pp. 20–28, 2024, doi: 10.33395/sinkron.v8i4.14000.
 - [16] M. Risanasari, N. Aulia, and L. Cahyani, “Clustering Of Student Learning Styles in the Industry 4 . 0 Using KMeans Algorithm,” *J. Bina Komput.*, vol. 24, no. August, pp. 246–257, 2022.
 - [17] A. K. E. Pily and U. Rio, “Komparasi K-Means Clustering dengan Euclidean dan Cosine Similarity untuk Segmentasi dan Rekomendasi Produk pada Data E-Commerce,” *Indones. J. Comput. Sci.*, vol. 14, no. 2, 2025, doi: 10.33022/ijcs.v14i2.4713.
 - [18] R. D. Nugraha, D. D. Adelia, and D. Rivaldi, “Segmentasi Pelanggan Retail Berbasis Perilaku menggunakan Algoritma K-Means Clustering,” *Digit. Transform. Technol.*, vol. 5, no. 2, pp. 141–148, 2025, doi: 10.47709/digitech.v5i2.6340.
 - [19] A. Ghosh Chowdhury, R. Deora, R. Prathikantam, and S. Kumar, “Dynamic Clustering with Deep Learning for Customer Segmentation,” in *2025 7th International Conference on Intelligent Sustainable Systems (ICISS)*, 2025, pp. 450–456. doi: 10.1109/ICISS63372.2025.11076506.
 - [20] Z. Lei, “Advances in Regression and Classification Models for Customer Data Analysis in Retail Enterprises: A Comprehensive Review,” *Adv. Econ. Manag. Res.*, vol. 15, no. 1, pp. 200–210, 2025, doi: 10.56028/aemr.15.1.750.2025.