

Komparasi Algoritma K-Means dan K-Medoids untuk Segmentasi Pelanggan E-Commerce Berbasis Pola Pembelian dan Lokasi Geografis

Comparison of K-Means and K-Medoids Algorithms for E-Commerce Customer Segmentation Based on Purchasing Patterns and Geographic Location

Jauza Kenar Indra Fawaid^{*1}, Aprilisa Arum Sari², Hanifah Permatasari³

^{1,2,3} Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Duta Bangsa Surakarta
e-mail: ^{*1}jauzafawaid831@gmail.com

Abstrak – E-commerce adalah transaksi jual beli secara elektronik yang di Indonesia kini sangat kompetitif karena hampir setiap pemilik usaha memiliki online shop sendiri, sehingga pemahaman pelanggan melalui pengelompokan perilaku transaksi menjadi penting. Penelitian ini membandingkan algoritma K-Means dan K-Medoids untuk menentukan algoritma segmentasi pelanggan e-commerce terbaik berbasis pola pembelian dan lokasi geografis. Data yang digunakan adalah data penjualan Shopee sebanyak 20.848 transaksi (Desember 2023–November 2025) dari 34 provinsi, dengan variabel total pembayaran, jumlah barang, dan ongkos kirim. Mengikuti tahapan CRISP-DM, data dipraproses melalui penyaringan sehingga diperoleh 17.706 data valid, yang kemudian digunakan untuk membandingkan kedua algoritma melalui RMSE, Skor Silhouette, dan Davies-Bouldin Index (DBI), dengan jumlah kluster ditentukan melalui Elbow Method dan Skor Silhouette. Hasil menunjukkan K-Means lebih unggul daripada K-Medoids (RMSE 1,14 berbanding 1,25; silhouette 0,40 berbanding 0,23; DBI 1,01 berbanding 1,14), sehingga dipilih sebagai algoritma terbaik dan membentuk 3 segmen, yaitu Pembeli Hemat, Pembeli Reguler, dan Pelanggan Grosir Premium.

Kata kunci – e-commerce; K-Means; K-Medoids; komparasi algoritma; segmentasi pelanggan

Abstract – E-commerce is electronic buying and selling that has become highly competitive in Indonesia, as nearly every business owner now runs an online shop, making customer understanding through transaction-based grouping essential. This study compares the K-Means and K-Medoids algorithms to determine the best e-commerce customer segmentation algorithm based on purchasing patterns and geographic location. The data used are 20,848 Shopee transaction records (December 2023–November 2025) from 34 provinces, with variables of total payment, number of items, and shipping cost. Following the CRISP-DM stages, the data were preprocessed through filtering, resulting in 17,706 valid records, which were then used to compare both algorithms using RMSE, Silhouette Score, and the Davies-Bouldin Index (DBI), with the number of clusters determined through the Elbow Method and Silhouette Score. The results show that K-Means outperforms K-Medoids (RMSE 1.14 versus 1.25; silhouette 0.40 versus 0.23; DBI 1.01 versus 1.14), and was therefore selected as the best algorithm, forming 3 segments: Budget Buyers, Regular Buyers, and Premium Wholesale Customers.

Keywords – algorithm comparison; customer segmentation; e-commerce; K-Means; K-Medoids

I. PENDAHULUAN

E-commerce di Indonesia berkembang sangat pesat. Sebagai gambaran, Badan Pusat Statistik mencatat jumlah usaha e-commerce mencapai 4,40 juta unit pada 2024, naik sekitar 15,3% dari tahun sebelumnya [19]. Sayangnya, data yang melimpah itu tidak berguna kalau hanya disimpan tanpa diolah. Di sinilah teknik *data mining* dan *machine learning* berperan menggali pola tersembunyi di balik data [1], salah satunya melalui segmentasi pelanggan, yaitu mengelompokkan pelanggan yang berkarakteristik serupa menggunakan clustering. Banyak penelitian sebelumnya memadukan K-Means dengan *model Recency, Frequency, Monetary* (RFM) untuk segmentasi pelanggan [2]-[15]. Sayangnya, kebanyakan berhenti pada perilaku belanja, perbandingan K-Means dengan algoritma lain masih jarang dilakukan [16], dan faktor lokasi geografis belum banyak dilibatkan sebagai pembeda antarsegmen.

Pada data transaksi nyata, identitas pelanggan sering tidak tersedia, sehingga segmentasi lebih cocok dilakukan pada tingkat transaksi. Padahal faktor geografis seperti lokasi pengiriman bisa menjadi pembeda penting, apalagi di Indonesia yang wilayahnya sangat luas. Dari permasalahan tersebut, penelitian ini melakukan segmentasi pada tingkat transaksi, menambahkan sisi geografis lewat ongkos kirim, sekaligus membandingkan K-Means dengan K-Medoids agar pilihan algoritmanya benar-benar beralasan. Keseluruhan prosesnya mengikuti kerangka CRISP-DM [17], dengan jumlah klaster ditentukan melalui Metode Elbow [18].

Penelitian ini bertujuan membandingkan kinerja K-Means dan K-Medoids untuk segmentasi pelanggan e-commerce berdasarkan pola pembelian dan lokasi geografis, menentukan algoritma terbaik, lalu menyusun profil tiap segmen agar dapat dipakai secara praktis untuk strategi promosi berjenjang. Kontribusi penelitian ini terletak pada pemaduan ongkos kirim sebagai penanda lokasi dengan pola pembelian pada segmentasi tingkat transaksi tanpa identitas pelanggan, kombinasi yang masih jarang ditemui, serta penerapan model terbaik ke dalam sebuah dashboard yang dapat langsung menentukan segmen transaksi baru.

II. TINJAUAN TEORI

A. Data Mining

Data mining adalah proses menggali pola yang tersembunyi dari kumpulan data berukuran besar agar dapat memberikan nilai guna bagi pengambilan keputusan. Salah satu cabangnya adalah *machine learning*, yaitu pendekatan yang memungkinkan komputer mempelajari pola dari data tanpa diprogram secara eksplisit.

B. Clustering

Clustering adalah teknik unsupervised learning yang mengelompokkan data berdasarkan kemiripan karakteristik tanpa memerlukan label, sehingga data dalam satu klaster saling mirip sedangkan antar klaster saling berbeda. Pada penelitian ini, *clustering* digunakan untuk mengelompokkan transaksi berdasarkan pola pembelian dan lokasi geografis.

C. Algoritma K-Means

K-Means adalah algoritma *clustering* yang membagi data ke dalam K klaster menggunakan titik pusat (*centroid*), dengan menempatkan setiap data ke centroid terdekat lalu memperbarui posisinya menjadi rata-rata anggota klaster hingga tidak berubah lagi. Algoritma ini sederhana dan efisien, namun cukup sensitif terhadap data pencilan (*outlier*) karena pusat klasternya berupa nilai rata-rata.

D. Algoritma K-Medoids

K-Medoids mirip dengan K-Means, tetapi pusat klasternya berupa data nyata yang disebut *medoid*, bukan nilai rata-rata. Karena menggunakan data aktual sebagai pusat, K-Medoids secara teori lebih tahan terhadap outlier dibandingkan K-Means, meskipun perhitungannya lebih berat untuk data berjumlah besar.

E. Metode Elbow

Metode *Elbow* digunakan untuk menentukan jumlah klaster optimal dengan mengamati nilai *Within-Cluster Sum of Squares* (WCSS), yaitu total jarak kuadrat data terhadap pusat klasternya. Jumlah klaster yang dipilih adalah titik ketika penurunan WCSS mulai melandai dan membentuk pola menyerupai siku (*elbow*).

F. Metrik Evaluasi

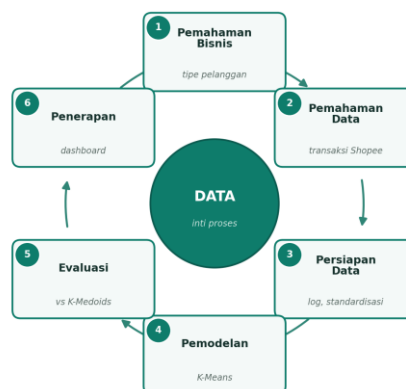
Kinerja hasil *clustering* diukur dengan tiga metrik. *Root Mean Square Error* (RMSE) menggambarkan kerapatan anggota klaster, semakin kecil semakin baik. *Skor Silhouette* menilai keterpisahan antar klaster dengan rentang -1 sampai 1, semakin besar semakin baik. *Davies-Bouldin Index* (DBI) mengukur rasio kemiripan antar klaster, semakin kecil semakin baik.

G. CRISP-DM

Cross-Industry Standard Process for Data Mining (CRISP-DM) adalah kerangka kerja standar untuk proyek data mining yang terdiri dari enam tahap, yaitu pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan penerapan. Kerangka ini dipilih karena sistematis dan tidak terikat pada satu algoritma tertentu.

III. METODE

Penelitian ini disusun mengikuti kerangka *Cross-Industry Standard Process for Data Mining* (CRISP-DM) karena sifatnya yang sistematis dan tidak terpaku pada satu algoritma tertentu.



Gambar 1. Tahapan penelitian berbasis CRISP-DM

A. Pemahaman Bisnis

Tahap ini menerjemahkan kebutuhan pelaku usaha menjadi tujuan penelitian. Persoalan utamanya, pemilik usaha belum mengenali kelompok-kelompok pelanggannya, baik dari sisi kebiasaan belanja maupun sebaran lokasinya, sehingga pemasaran menjadi kurang tepat sasaran. Karena itu, sasaran yang ingin dicapai adalah membentuk segmen pelanggan yang

bermakna, memilih algoritma pengelompokan yang paling cocok untuk data ini, dan menyediakan dasar bagi pemilik usaha untuk menyusun promosi yang berbeda di tiap segmen.

B. Pemahaman Data

Data yang dipakai berjumlah 20.848 catatan transaksi penjualan dari sebuah toko di platform Shopee, terhitung sejak Desember 2023 sampai November 2025. Transaksinya berasal dari 34 provinsi; empat provinsi sisanya dari total 38 provinsi di Indonesia tidak muncul bukan karena dihapus, melainkan karena memang tidak ada transaksi dari wilayah tersebut pada data toko ini. Yang perlu dicatat, data ini tidak memuat identitas pelanggan, sehingga satu baris mewakili satu transaksi atau pesanan, bukan satu orang pelanggan. Itulah alasan segmentasi pada penelitian ini diletakkan pada tingkat transaksi. Rincian variabel dan perannya dapat dilihat pada Tabel 1.

Tabel 1. Variabel penelitian dan perannya

Variabel	Kolom Sumber	Peran
Total pembayaran	Total Pembayaran	Pola pembelian (nilai belanja)
Jumlah barang	total_qty	Pola pembelian (kuantitas)
Ongkos kirim	Perkiraan Ongkos Kirim	Proksi lokasi geografis
Provinsi	Provinsi	Penjelas profil segmen

C. Persiapan Data

Sebelum diolah, data melewati beberapa langkah pembersihan agar layak dijadikan masukan model. Mula-mula, hanya transaksi berstatus selesai yang dipertahankan, karena transaksi yang batal atau dikembalikan tidak menggambarkan pembelian yang sungguh terjadi. Baris dengan nilai kosong dibuang supaya tidak mengganggu hasil pengelompokan. Setelah bersih, ketiga variabel dikenai transformasi logaritmik dengan log natural untuk meredam pengaruh nilai ekstrem tanpa mengubah urutan besar-kecilnya. Langkah terakhir adalah standarisasi, meski begitu, nilai asli tiap variabel tetap disimpan untuk keperluan membaca profil segmen nanti.

D. Pemodelan

Pada tahap ini, dua algoritma pengelompokan dijalankan lalu dibandingkan. K-Means bekerja dengan menaruh sejumlah K titik pusat (*centroid*), menempatkan setiap data ke *centroid* yang paling dekat, kemudian menggeser *centroid* ke posisi rata-rata anggotanya. Langkah itu diulang sampai posisi *centroid* tidak berubah lagi. Kedekatan antardata diukur dari jaraknya, sehingga data yang karakternya mirip akan masuk ke kelompok yang sama.

Jumlah klaster yang pas ditentukan dengan Metode *Elbow*. Pemilihan jumlah klaster ini diperkuat lagi dengan Skor *Silhouette*. Sebagai pembanding, dijalankan pula K-Medoids. Algoritma ini mirip K-Means, bedanya pusat tiap kelompok bukan rata-rata, melainkan salah satu data nyata (medoid). Karena pusatnya berupa data asli, K-Medoids secara teori lebih tahan terhadap nilai yang ekstrem, walaupun perhitungannya jadi lebih berat ketika datanya banyak.

E. Evaluasi

Kualitas kedua algoritma dinilai memakai tiga ukuran. Yang pertama adalah *Root Mean Square Error* (RMSE), ukuran kedua adalah Skor *Silhouette*, ukuran ketiga adalah Davies-Bouldin

Index (DBI). Kebalikan dari *silhouette*, nilai DBI yang kecil menandakan kelompok yang rapat sekaligus saling berjauhan. Dengan memakai tiga ukuran sekaligus, penilaiannya menjadi lebih adil dan tidak bergantung pada satu angka saja.

F. Penerapan

Pada tahap penerapan, model terbaik dikemas ke dalam sebuah *dashboard* interaktif supaya hasil penelitian bisa langsung dipakai oleh pelaku usaha. Model K-Means beserta penskala (scaler) dan penamaan segmennya disimpan dalam satu berkas lewat pickle, yaitu mekanisme penyimpanan objek di Python streamlit, sehingga model dapat dipanggil kembali tanpa harus dilatih ulang. Berkat cara ini, *dashboard* tinggal memuat model yang sudah tersimpan untuk menentukan segmen sebuah transaksi baru, dan prosesnya pun tetap konsisten serta cepat.

IV. HASIL DAN PEMBAHASAN

A. Hasil Praproses Data

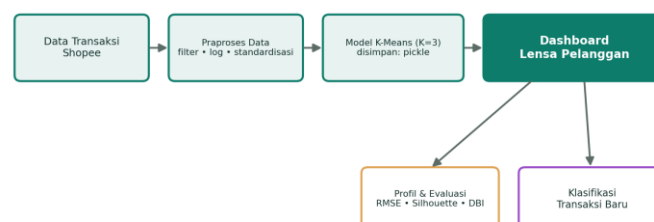
Data awal sebanyak 20.848 catatan transaksi diolah mengikuti tahap persiapan data. Penyaringan transaksi yang berstatus selesai dan pembuangan baris bermasalah membuat sebagian data gugur. Transaksi yang batal, dikembalikan, atau masih berjalan tidak ikut dihitung karena belum tentu menjadi pembelian yang benar-benar terjadi, sementara baris dengan pembayaran kosong atau bernilai nol dianggap tidak sah. Hasilnya, 3.142 transaksi tersaring keluar dan tersisa 17.706 transaksi valid, atau sekitar 84,9% dari data awal. Dari sinilah jumlah data menyusut dari 20.848 menjadi 17.706. Ringkasan statistik deskriptif ketiga variabel sebelum transformasi ditampilkan pada Tabel 2.

Tabel 2. Statistik deskriptif tiga variabel sebelum transformasi

Variabel	Minimum	Median	Rata-rata	Maksimum
Total pembayaran (Rp)	165	26.398	59.051	3.403.591
Jumlah barang	1	1	2	256
Ongkos kirim (Rp)	1	10.000	17.376	624.000

Dari Tabel 2 terlihat sebaran data yang sangat miring: nilai rata-ratanya jauh melampaui median, pertanda ada segelintir transaksi dengan nilai yang sangat besar. Pola seperti inilah yang melatarbelakangi penggunaan transformasi logaritmik pada tahap sebelumnya.

Setelah data bersih ini dipakai untuk membangun model, sistem kemudian dirancang agar bisa menentukan segmen sebuah transaksi baru secara langsung, tanpa perlu menghitung ulang seluruh pengelompokan. Alur kerjanya digambarkan pada Gambar 2: data transaksi melewati praproses, dilatih menjadi model K-Means, lalu disimpan sebagai berkas pickle yang nantinya dipanggil oleh *dashboard* untuk menentukan segmen data baru.



Gambar 2. Arsitektur penerapan model pada dashboard

Lewat masukan manual, pengguna cukup mengetik total pembayaran, jumlah barang, ongkos kirim, dan lokasi sebuah transaksi, lalu sistem langsung memberi tahu segmennya. Bila datanya banyak, pengguna tinggal mengunggah berkas dan seluruh barisnya akan disegmentasi sekaligus. Penentuan ini tetap bertumpu pada tiga variabel, yaitu total pembayaran, jumlah barang, dan ongkos kirim sebagai proksi lokasi. Sebagai gambaran, sepuluh transaksi baru dicoba dimasukkan ke sistem dan hasilnya disajikan pada Tabel 3, sementara penjelasan ketiga nama segmennya menyusul di bagian profil segmen.

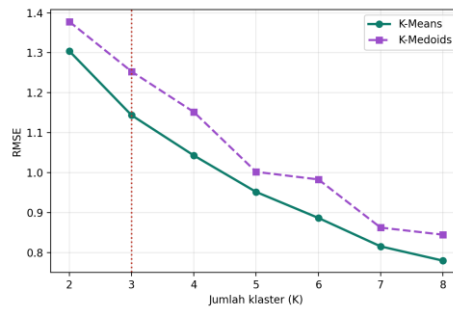
Tabel 3. Penentuan segmen sepuluh transaksi baru menggunakan model tersimpan

No.	Total Pembayaran (Rp)	Barang	Ongkir (Rp)	Provinsi	Segmen
1	15.000	1	8.000	Jawa Barat	Pembeli Hemat
2	22.000	1	9.000	DKI Jakarta	Pembeli Hemat
3	9.000	1	7.000	Banten	Pembeli Hemat
4	30.000	1	12.000	Jawa Tengah	Pembeli Hemat
5	48.000	2	18.000	Jawa Timur	Pembeli Reguler
6	65.000	3	25.000	Sumatera Utara	Pembeli Reguler
7	55.000	2	30.000	Sulawesi Selatan	Pembeli Reguler
8	120.000	7	32.000	Bali	Pelanggan Grosir Premium
9	180.000	10	38.000	Kalimantan Timur	Pelanggan Grosir Premium
10	250.000	15	40.000	Papua	Pelanggan Grosir Premium

Mengacu pada Tabel 3, dari sepuluh transaksi uji, empat transaksi bernilai kecil dengan satu barang masuk ke Pembeli Hemat, tiga transaksi bernilai menengah dengan dua sampai tiga barang masuk ke Pembeli Reguler, dan tiga transaksi bernilai besar dengan barang yang banyak tergolong Pelanggan Grosir Premium. Artinya, model cukup andal menempatkan tiap transaksi baru ke segmen yang sesuai dengan karakternya, sehingga pemilik usaha bisa langsung menindaklanjutinya dengan promosi yang tepat tanpa harus mengolah ulang seluruh data.

B. Perbandingan K-Means dan K-Medoids

Sebelum segmen akhir dibentuk, kedua algoritma diadu lebih dulu untuk melihat mana yang lebih cocok pada data ini. Pengujian dilakukan pada rentang jumlah kluster K dari 2 sampai 8 dengan tiga metrik sekaligus, yaitu RMSE, Skor *Silhouette*, dan DBI. Gambar 3 memperlihatkan tren RMSE keduanya di setiap nilai K, sedangkan Tabel 4 merangkum ketiga metrik pada tiga kluster, yakni jumlah yang akhirnya dipakai untuk model akhir.



Gambar 3. Perbandingan RMSE K-Means dan K-Medoids

Tabel 4. Perbandingan kinerja K-Means dan K-Medoids pada tiga klaster

Metrik Evaluasi	K-Means	K-Medoids	Keterangan
RMSE	1,14	1,25	lebih kecil lebih baik
Skor Silhouette	0,40	0,23	lebih besar lebih baik
Davies-Bouldin Index	1,01	1,14	lebih kecil lebih baik

Tabel 4 memperlihatkan K-Means unggul pada ketiga metrik sekaligus: RMSE-nya lebih kecil (1,14 berbanding 1,25), Skor *Silhouette*-nya lebih tinggi (0,40 berbanding 0,23), dan DBI-nya lebih kecil (1,01 berbanding 1,14). Keunggulan ini bukan kebetulan pada tiga klaster saja, sebab konsisten sepanjang K=2 hingga 8, dengan K-Means menang pada 20 dari 21 perbandingan metrik. Kecenderungan serupa juga ditemukan pada penelitian terdahulu yang membandingkan K-Means dan K-Medoids, di mana K-Means memperoleh DBI yang sedikit lebih kecil (0,976 berbanding 0,981) [20], sehingga keunggulan K-Means pada penelitian ini sejalan dengan temuan sebelumnya. Atas dasar itu, K-Means ditetapkan sebagai algoritma terbaik untuk data ini.

Meski begitu, perlu jujur diakui bahwa RMSE secara matematis memang sedikit menguntungkan K-Means. Pusat K-Means berupa rata-rata yang justru dirancang untuk memperkecil jarak kuadrat ke pusat, sehingga RMSE seolah menilai hal yang persis menjadi tujuan internal K-Means itu sendiri. Karena itulah RMSE tidak dijadikan satu-satunya patokan. Skor *Silhouette* dan DBI menilai pemisahan antarklaster secara mandiri tanpa memihak metode mana pun, dan keduanya pun tetap memenangkan K-Means, sehingga kesimpulannya bisa dipegang. Keunggulan K-Medoids dalam menahan nilai ekstrem juga menjadi kurang berpengaruh di sini karena efek nilai ekstrem sudah diredam lewat transformasi logaritmik dan standardisasi. Sebagai gambaran tambahan, pada data terstandar nilai RMSE tanpa pengelompokan sekitar 1,73; begitu dibagi menjadi tiga klaster, angkanya turun ke 1,14 atau kira-kira 66% dari kondisi awal, setara dengan menjelaskan sekitar 56% variasi data. Penurunan sebesar itu menandakan pengelompokan yang memang bermakna.

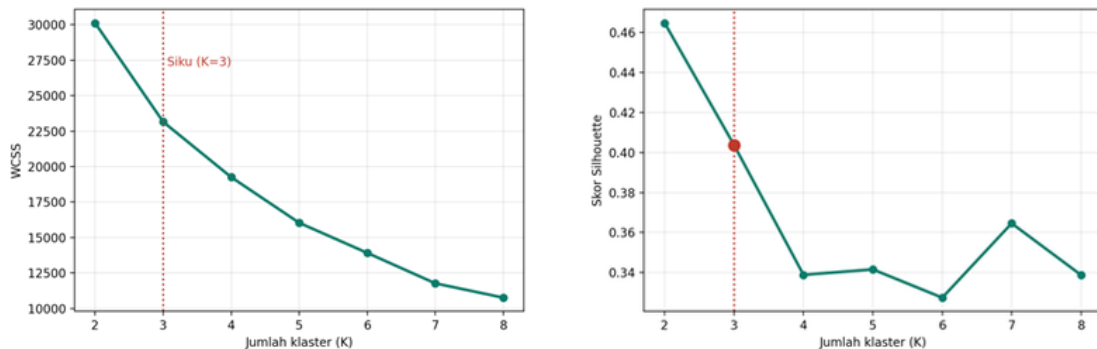
C. Penentuan Jumlah Klaster

Begitu K-Means terpilih sebagai algoritma terbaik, langkah selanjutnya adalah mencari jumlah klaster yang paling pas khusus untuk K-Means. Caranya dengan mengamati nilai WCSS lewat Metode *Elbow* sekaligus nilai Skor *Silhouette* untuk K dari 2 hingga 8. Rangkumannya ada pada Tabel 5, sedangkan grafiknya tampak pada Gambar 4.

Tabel 5. WCSS, skor silhouette, dan RMSE K-Means untuk tiap K

Jumlah Klaster (K)	WCSS	Skor Silhouette	RMSE
2	30.117	0,46	1,30

Jumlah Klaster (K)	WCSS	Skor Silhouette	RMSE
3	23.154	0,40	1,14
4	19.254	0,34	1,04
5	16.033	0,34	0,95
6	13.915	0,33	0,89
7	11.768	0,36	0,82
8	10.760	0,34	0,78



Gambar 4. Metode Elbow (WCSS) K-Means dan Skor Silhouette K-Means

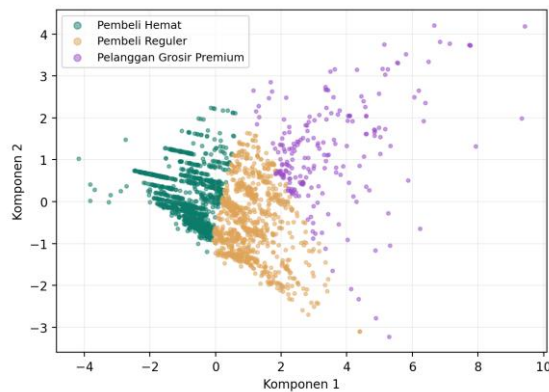
Pada Tabel 5, nilai WCSS dan RMSE memang terus turun setiap kali K ditambah, jadi keduanya tidak bisa berdiri sendiri sebagai penentu; yang diperhatikan justru di titik mana kurvanya mulai melandai. Kurva WCSS pada Gambar 4 mulai mendatar sejak K=3 yang menjadi titik sikunya, dan Skor *Silhouette* pun tergolong tinggi di K=3, yakni 0,40. Karena titik siku dan *Silhouette* tertinggi pada rentang awal sama-sama mengarah ke K=3, maka jumlah klaster yang dipakai adalah tiga.

D. Profil Segmen Pelanggan

Dengan tiga klaster, K-Means menghasilkan tiga segmen pelanggan yang penamaannya mengikuti urutan nilai belanja. Rinciannya dapat dilihat pada Tabel 6 dan gambarannya pada Gambar 6.

Tabel 6. Profil tiga segmen pelanggan (model K-Means, K=3)

Segmen	Jumlah	Persen	Median Belanja	Barang	Ongkir
Pembeli Hemat	11.291	63,8%	Rp18.900	1	Rp8.000
Pembeli Reguler	5.180	29,3%	Rp54.000	2	Rp22.500
Pelanggan Grosir Premium	1.235	7,0%	Rp128.764	8	Rp34.000



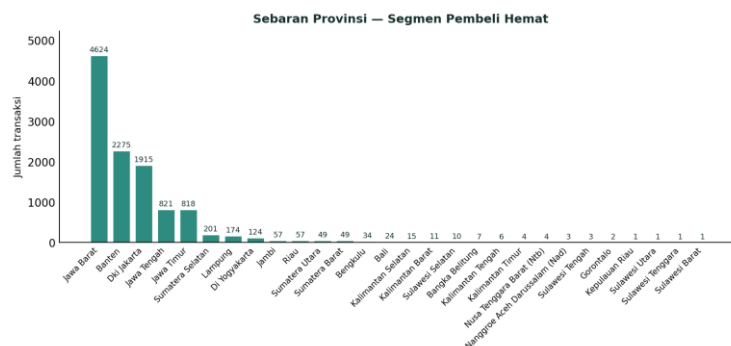
Gambar 6. Visualisasi tiga segmen pelanggan (proyeksi dua komponen)

Pembeli Hemat adalah kelompok paling besar (63,8%) dengan nilai belanja kecil dan rata-rata hanya satu barang per transaksi; kelompok ini paling pas didorong lewat promosi paket atau poin kecil supaya belanjanya bertambah. Pembeli Reguler (29,3%) berada di tengah dengan sekitar dua barang per transaksi, dan cocok diberi imbalan tingkat menengah seperti gratis ongkos kirim bersyarat. Sementara itu, Pelanggan Grosir Premium memang paling sedikit jumlahnya (7,0%), tetapi nilainya paling tinggi, dengan median belanja Rp128.764 dan barang yang jauh lebih banyak; merekalah yang paling layak mendapat poin atau imbalan terbesar sekaligus program loyalitas agar tetap bertahan. Dengan perbedaan semacam ini, pemilik usaha dapat menjalankan promosi berjenjang yang disesuaikan dengan nilai tiap segmen.

E. Sebaran Geografis Segmen

Dari sisi geografis, perbedaan ongkos kirimnya cukup mencolok: median ongkos kirim di dalam Pulau Jawa hanya sekitar Rp9.500, sementara di luar Jawa bisa mencapai sekitar Rp32.300. Secara umum, mayoritas transaksi (sekitar 80,9%) memang berasal dari Pulau Jawa.

Selain proporsi luar Jawa di atas, sebaran tiap segmen dapat dilihat lebih rinci per provinsi. Gambar 7, Gambar 8, dan Gambar 9 menampilkan jumlah transaksi tiap segmen di masing-masing provinsi. Ketiga segmen sama-sama terpusat di provinsi-provinsi Pulau Jawa seperti Jawa Barat, Banten, dan DKI Jakarta, tetapi pola sebarannya berbeda. Pembeli Hemat sangat menumpuk di Jawa Barat, sedangkan Pembeli Reguler dan Pelanggan Grosir Premium relatif lebih menyebar hingga ke provinsi-provinsi luar Jawa.



Gambar 7. Sebaran transaksi segmen Pembeli Hemat per provinsi

dashboard yang dapat menentukan segmen transaksi baru secara langsung, sehingga pemilik usaha bisa memakainya untuk merancang promosi atau program loyalitas yang berbeda bagi tiap segmen, dan ke depannya perbandingan masih bisa diperluas dengan algoritma lain atau data yang memuat identitas pelanggan.

VI. DAFTAR PUSTAKA

- [1] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *SN Computer Science*, vol. 2, p. 160, 2021.
- [2] A. Febri, N. Ningsih, dan J. Lemantara, "Aplikasi analisis segmentasi pelanggan untuk menentukan strategi pemasaran menggunakan kombinasi metode K-Means dan model RFM," *SISTEMASI: Jurnal Sistem Informasi*, vol. 10, no. 1, pp. 139–151, 2021.
- [3] S. F. Siregar, B. I. Nasution, dan M. F. Syahputra, "Optimalisasi segmentasi pelanggan ritel dengan algoritma K-Means berbasis RFM untuk strategi pemasaran," *JURIKOM (Jurnal Riset Komputer)*, vol. 10, no. 6, 2023.
- [4] P. M. K. Mado dan H. Hendry, "Implementasi algoritma clustering K-Means untuk segmentasi pelanggan di e-commerce," *Jurnal Indonesia: Manajemen Informatika dan Komunikasi*, vol. 6, no. 3, pp. 1680–1686, 2025.
- [5] A. P. Pramudiansyah, "Segmentasi pelanggan menggunakan algoritma K-Means berdasarkan model recency frequency monetary," *Jurnal Ilmiah Ilmu Komputer*, vol. 7, no. 2, pp. 6–19, 2021.
- [6] K. Z. Wijaya, A. Djunaidi, dan F. Mahananto, "Segmentasi pelanggan menggunakan algoritma K-Means dan analisis RFM di Ova Gaming E-Sports Arena Kediri," *Jurnal Teknik ITS*, vol. 10, no. 2, 2021.
- [7] W. A. Silamantha dan K. Hadiono, "Analisis RFM dan K-Means clustering untuk segmentasi pelanggan pada PT. Sanutama Bumi Arto," *Kesatria: Jurnal Penerapan Sistem Informasi (Komputer dan Manajemen)*, vol. 5, no. 3, pp. 1297–1305, 2024.
- [8] B. T. Kristanti, A. Junaidi, dan E. P. Mandyartha, "Implementasi K-Means clustering dalam segmentasi pelanggan berdasarkan usia, pendapatan, dan model RFM (Studi kasus: Lantikya Store Jombang)," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, 2024.
- [9] R. Siagian, P. S. P. Sirait, dan A. Halima, "E-commerce customer segmentation using K-Means algorithm and length, recency, frequency, monetary model," *Journal of Informatics and Telecommunication Engineering*, vol. 5, no. 1, pp. 21–30, 2021.
- [10] R. B. Ardi, F. E. Nastiti, dan S. Sumarlinda, "Algoritma K-Means clustering untuk segmentasi pelanggan (Studi kasus: Fashion Viral Solo)," *INFOTECH Journal*, vol. 9, no. 1, pp. 124–131, 2023.
- [11] E. F. L. Awalina dan W. I. Rahayu, "Optimalisasi strategi pemasaran dengan segmentasi pelanggan menggunakan penerapan K-Means clustering pada transaksi online retail," *Jurnal Teknologi dan Informasi*, vol. 13, no. 2, pp. 122–137, 2023.
- [12] R. Fitriyani, A. L. Firmansyah, dan L. A. Nurfadillah, "Penerapan algoritma clustering untuk segmentasi pelanggan e-commerce berdasarkan data pembelian dan aktivitas," *Seminar Teknologi Majalengka (STIMA)*, vol. 8, pp. 372–379, 2024.

- [13] C. Hafidz Ardana, A. A. Khoyum, dan M. Faisal, "Segmentasi pelanggan penjualan online menggunakan metode K-Means clustering," *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 9, no. 1, pp. 1–9, 2024.
- [14] K. Tabianan, S. Velu, dan V. Ravi, "K-Means clustering approach for intelligent customer segmentation using customer purchase behavior data," *Sustainability*, vol. 14, no. 12, p. 7243, 2022.
- [15] I. Tazkiyah, A. A. Arifiyanti, dan A. R. E. Najaf, "Implementasi segmentasi pelanggan e-commerce menggunakan algoritma K-Means pada website," *Prosiding Seminar Nasional Teknologi dan Sistem Informasi (SITASI)*, vol. 3, no. 1, pp. 217–226, 2023.
- [16] S. D. K. Wardani, A. S. Ariyanto, M. Umroh, dan D. Rolliawati, "Perbandingan hasil metode clustering K-Means, DBSCAN & hierarchical untuk analisa segmentasi pasar," *JIKO (Jurnal Informatika dan Komputer)*, vol. 7, no. 2, p. 191, 2023.
- [17] C. Schröder, F. Kruse, dan J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," *Procedia Computer Science*, vol. 181, pp. 526–534, 2021.
- [18] R. Nainggolan, R. Perangin-angin, E. Simarmata, dan F. A. Tarigan, "Improved the performance of the K-Means cluster using the sum of squared error (SSE) optimized by using the Elbow method," *Journal of Physics: Conference Series*, vol. 1361, p. 012015, 2019.
- [19] Badan Pusat Statistik, *Statistik E-Commerce 2024*. Jakarta: Badan Pusat Statistik, 2024.
- [20] Y. Hasan, "Pengukuran silhouette score dan davies-bouldin index pada hasil cluster K-Means dan K-Medoids," *Jurnal Teknik Informatika*, 2024.