

Menerapkan Algoritma Data Mining Untuk Memahami Profil dan Preferensi Pembeli Online Menggunakan Metode K – Means

Zaehol Fatah,¹, Nafilil Madani Al-Qur'ani²
Sistem Informasi, Universitas Ibrahimy, Situbondo¹
Teknologi Informasi, Universitas Ibrahimy, Situbondo²
Email: nafil0212@gmail.com

Abstract

The swift growth of e-commerce produces vast volumes of transaction records which are currently not fully leveraged to comprehend the habits and choices of online consumers. This research utilizes the K-Means algorithm for clustering, a technique in data mining, to uncover buying habits and create profiles of digital shoppers. The research procedure includes data collection, preprocessing, normalization, determination of the optimal number of clusters using the Elbow Method, and cluster quality evaluation through the Silhouette Score and Davies–Bouldin Index. The results reveal that online buyers can be grouped into several distinct segments, including high-value consumers, high-frequency active buyers, and low-spending passive users. Every segment exhibits distinctive patterns of behavior which enable businesses to formulate highly tailored marketing plans, boost customer loyalty, and refine decisions based on data. The results demonstrate the efficacy of the K-Means algorithm in processing transactional information, yielding valuable insights concerning customer preferences. Consequently, this research underscores the critical role of clustering methodologies in improving strategic choices within the digital commerce sphere.

Keywords: K-Means, data mining, customer segmentation, e-commerce, purchasing patterns.

Abstrak

Pertumbuhan pesat e-commerce menyebabkan meningkatnya volume data transaksi yang belum dimanfaatkan secara optimal untuk memahami perilaku dan preferensi pembeli online. Berbagai penelitian terkait segmentasi pelanggan menunjukkan bahwa metode *K-Means Clustering* efektif dalam mengidentifikasi pola pembelian dan karakteristik konsumen berdasarkan atribut seperti frekuensi transaksi, nilai belanja, waktu pembelian, serta variabel perilaku lainnya. Penelitian ini menerapkan algoritma K-Means sebagai teknik data mining untuk memprofilkan pembeli online dan menggali preferensi mereka melalui analisis pola transaksi. Tahapan penelitian meliputi pengumpulan data, pra-pemrosesan, normalisasi, penentuan jumlah kluster optimal menggunakan Metode Elbow, serta evaluasi kualitas kluster dengan Silhouette Score maupun Davies-Bouldin Index sebagaimana diterapkan pada penelitian sebelumnya. Hasil dari klusterisasi dalam penelitian ini mengindikasikan bahwa para pembeli dapat dibagi menjadi berbagai segmen pelanggan yang memiliki karakteristik yang berbeda. Segmen-segmen ini mencakup konsumen dengan nilai transaksi tinggi, konsumen yang sangat aktif (berfrekuensi tinggi), dan juga konsumen yang kurang aktif dengan nilai pembelian yang rendah. Penemuan ini konsisten dengan hasil studi sebelumnya yang relevan dalam bidang *e-commerce* dan ritel. Dengan demikian, penerapan algoritma K-Means terbukti mampu membantu memahami profil dan preferensi pembeli online secara lebih mendalam serta meningkatkan efektivitas strategi bisnis digital.

Kata Kunci: K-Means, data mining, segmentasi pelanggan, e-commerce, pola pembelian.

PENDAHULUAN

Sejarah perkembangan teknologi informasi merupakan perjalanan panjang manusia dalam menemukan cara untuk menyimpan, mengolah, dan menyebarkan informasi (Z. Fatah 2025). Di era serba digital ini, perdagangan daring telah merevolusi proses pembelian, sehingga platform e-commerce wajib mengenali konsumennya secara komprehensif demi memberikan pengalaman yang sesuai dan memuaskan.(Ardana et al. 2024). Gaya hidup masyarakat modern banyak diwarnai belanja online karena faktor kemudahan dan fleksibilitas akses informasi yang ditawarkan platform e-commerce. Hal tersebut dikarenakan banyaknya kemudahan yang didapat untuk memenuhi kebutuhan primer sampai dengan tersier.(Dan et al. 2025).

Terdapat beberapa alasan masyarakat yang lebih memilih melakukan pembelian secara online yaitu karena adanya kemudahan untuk berbelanja yang dapat dilakukan dimana saja, lebih cepat dan praktis hanya dengan memanfaatkan smartphone sebagai media dalam melakukan transaksi pembelian.(Kristanti, Junaidi, and Mandyartha 2024).

Konsumen tidak perlu lagi datang langsung ke toko, namun semua dapat dilakukan dimanapun dengan cara yang sangat mudah yaitu mengunduh aplikasi belanja online. Kemudahan akses informasi, fleksibilitas transaksi, serta meningkatnya platform e-commerce menjadikan belanja online merupakan bagian penting dari gaya hidup masyarakat modern.(Frاندika, Hud, and Handoko 2025). Fenomena ini mengakibatkan jumlah data transaksi melonjak drastis, tetapi seringkali data tersebut tidak dieksplorasi sepenuhnya untuk menggali preferensi dan tabiat konsumen secara rinci.

Pemahaman mengenai pola pembelian dan perilaku pelanggan menjadi aspek penting untuk meningkatkan efektivitas strategi pemasaran, loyalitas

pelanggan, (E-commerce et al. 2025)serta kinerja bisnis secara berkelanjutan.(Kusdaryanto et al. 2025)

Berbagai penelitian menegaskan bahwa strategi pemasaran konvensional yang bersifat umum (generalized marketing) tidak lagi memadai. Tanpa pemetaan perilaku yang tepat, perusahaan berpotensi salah dalam menentukan target pasar, sehingga strategi pemasaran menjadi kurang efektif dan tidak efisien untuk menjangkau segmen konsumen yang relevan. Salah satu metode populer untuk mengekstrak wawasan dari data transaksi adalah penambahan data, terutama metrik pengklasteran (clustering). Pengklasteran berfungsi mengisolasi objek atau pelanggan berdasarkan kesamaan fitur yang dimiliki dan pada akhirnya membentuk kelompok konsumen yang karakteristiknya serupa. Dalam konteks ini, Algoritma K-Means adalah salah satu metode yang sangat sering diterapkan sebab efisiensinya dalam memproses set data masif melalui langkah komputasi yang ringkas, cepat, dan berdaya guna. Sehingga algoritma ini mengukur jarak setiap data pada centroid, serta melakukan literasi hingga klaster yang stabil terbentuk.(Dan et al. 2025)

Segmentasi pelanggan berbasis K-Means membantu perusahaan memahami preferensi dan perilaku konsumen sehingga strategi pemasaran dapat disesuaikan secara personal dan tepat sasaran.(Penelitian et al. 2025) Di sisi lain, menyoroti bahwa data transaksi e-commerce sering kali bersifat kompleks dan kaya informasi sehingga diperlukan pendekatan analisis yang sistematis untuk mengungkap pola-pola tersembunyi di dalamnya. Melalui penerapan K-Means pada data Kaggle, penelitian tersebut menghasilkan beberapa klaster pelanggan berdasarkan metode pembayaran dan jenis produk yang dominan, yang dapat dijadikan dasar dalam pengambilan keputusan perusahaan terkait strategi penawaran dan promosi produk tertentu.(Dan et al. 2025) Berdasarkan

kajian dari berbagai penelitian tersebut, dapat disimpulkan bahwa algoritma K-Means memiliki potensi besar dalam memahami profil dan preferensi pembeli online melalui analisis data transaksi. (Dewa and Jatipaningrum 2019).

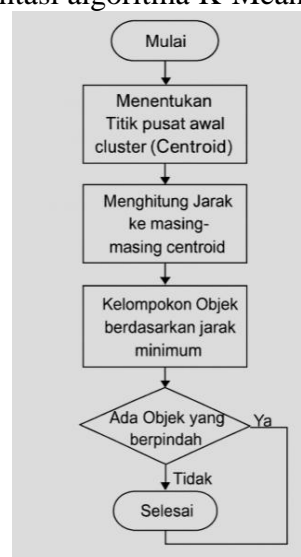
Segmentasi pelanggan yang akurat memungkinkan perusahaan menerapkan strategi pemasaran yang lebih personal, mengoptimalkan alokasi sumber daya, serta meningkatkan loyalitas pelanggan secara signifikan. Hal ini menjadikan penerapan algoritma K-Means sangat relevan dalam konteks persaingan bisnis digital atau belanja online. Maka dari itu, studi ini mengkhususkan diri pada implementasi K-Means untuk menggali karakteristik dan minat pembeli daring dengan mengaplikasikan teknik penambangan data pada kumpulan data transaksi. Target utamanya adalah menciptakan pembagian segmen konsumen yang berfungsi sebagai fondasi untuk keputusan strategis perusahaan yang lebih spesifik, terukur, dan didukung bukti empiris.

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksploratif untuk mengelompokkan pembeli online berdasarkan pola transaksinya. Bagian metode disusun secara ringkas dan langsung menjelaskan langkah teknis penelitian tanpa memuat penjelasan teori yang tidak diperlukan. Seluruh proses analisis dilakukan menggunakan perangkat lunak RapidMiner, sedangkan dataset transaksi diambil dari platform Kaggle yang menyediakan data e-commerce berskala besar dengan metode penerapan algoritma K-Means. Penjelasan disusun secara sistematis mulai dari teknik pengumpulan data, langkah-langkah pengolahan data, hingga metode analisis yang digunakan untuk membentuk kluster yang akurat.

A. Metode K-Means

K-Means merupakan cara atau metode pengelompokan yang terkenal dalam penampungan data yang bisa diterapkan untuk menyelesaikan permasalahan ini. Melalui K-Means, bisnis dapat memisahkan pelanggan ke dalam kelompok yang lebih seragam berdasarkan sifat dan pola belanja mereka. (E-commerce et al. 2025). Proses pengelompokan ini tidak dilakukan secara manual, tetapi menggunakan komputasi matematis yang memungkinkan sistem membaca pola yang tidak terlihat secara langsung. K-Means beroperasi dengan mengidentifikasi titik sentral (centroid) untuk tiap kelompok data, kemudian mengalokasikan setiap konsumen ke kluster yang jaraknya paling minimal ke titik pusat tersebut. Ini berarti, metode ini bertujuan untuk mengidentifikasi segmen pelanggan dengan pola tingkah laku yang identik dengan pusat tersebut. Berikut adalah tahapan-tahapan dalam implementasi algoritma K-Means:



Gambar 1. Metode yang di terapkan

Setiap langkah dalam diagram menunjukkan proses yang dilakukan secara berurutan hingga menghasilkan kluster pembeli yang stabil. Berikut penjelasannya:

1. Mulai (Start)

Proses algoritma dimulai setelah dataset dari Kaggle dipersiapkan dan

diinput ke RapidMiner. Tahap ini menandai dimulainya seluruh rangkaian prosedur clustering.

2. Menentukan Jumlah Cluster (K)

Pada langkah ini peneliti menetapkan jumlah klaster yang ingin dibentuk, misalnya $K=3$ atau $K=4$. Penentuan nilai K biasanya didukung dengan metode *Elbow* atau *Silhouette Score* pada tahap awal analisis. Tujuannya menentukan berapa kelompok pembeli online yang ingin dipetakan.

3. Menentukan Titik Pusat Awal (Centroid)

Algoritma kemudian memilih centroid awal, yaitu titik pusat dari setiap klaster. RapidMiner akan menginisialisasi centroid secara acak atau dengan metode *K-Means*. Fungsi centroid adalah menjadi pusat perhitungan jarak untuk setiap objek (pembeli) dalam dataset.

4. Menghitung Jarak ke Setiap Centroid

Jarak dihitung menggunakan Euclidean Distance untuk mengukur seberapa mirip perilaku pembeli satu dengan yang lain, dan hasilnya digunakan untuk menempatkan setiap pembeli ke klaster yang paling dekat dengan centroidnya. Pada tahap ini komputer menganalisis seberapa dekat perilaku pembeli tertentu dengan karakteristik tiap klaster, siapa pembeli yang mirip dan siapa yang berbeda.

5. Mengelompokkan Objek Berdasarkan Jarak Minimum

Setiap pembeli ditempatkan ke klaster yang memiliki jarak paling kecil dari centroid. Artinya, pembeli dikelompokkan bersama pembeli lain yang memiliki pola pembelian yang paling mirip.

Contoh hasil cluster:

Cluster 1 → pembeli aktif

Cluster 2 → pembeli nominal tinggi

Cluster 3 → pembeli pasif

6. Apakah Ada Objek yang Berpindah? (Decision Step)

Tahap ini adalah proses pengecekan: "Apakah masih ada objek yang berpindah klaster setelah perhitungan ulang?" Dua kemungkinan terjadi:

Jika YA (ada objek berpindah) Maka algoritma memperbarui posisi centroid baru, mengulang perhitungan jarak, mengelompokkan ulang pembeli. Proses kembali lagi ke langkah sebelumnya.

Jika TIDAK (tidak ada objek berpindah) Artinya klaster sudah stabil, tidak ada lagi perpindahan pembeli antar-klaster.

7. Selesai (Finish)

Proses dihentikan ketika tidak ada objek yang berpindah. Hasil akhirnya adalah klaster pembeli online yang sudah optimal, stabil, dan dapat di analisis. Output ini digunakan untuk memahami pola belanja, memetakan karakteristik pelanggan, menyusun strategi pemasaran yang tepat sasaran.

B. RapidMiner

RapidMiner merupakan aplikasi perangkat lunak yang menyeluruh, yang secara spesifik dikembangkan untuk memudahkan alur kerja analisis data, penambahan data, dan pembelajaran mesin tanpa membutuhkan pengodean yang rumit atau berlebihan. (Sari, Wanto, and Windarto 2018) Keunggulan utama dari platform ini terletak pada antarmuka visualnya yang intuitif, yang memungkinkan pengguna untuk menyusun, mengedit, dan menjalankan alur kerja analisis data yang kompleks melalui sistem seret-dan-lepas operator. Kemudahan ini secara signifikan mempercepat fase eksperimen model dan literasi analitis. Dalam hal ini, RapidMiner digunakan sebagai instrumen utama untuk implementasi *K-Means*, yang berfungsi mempermudah dan mengefisienkan proses pengelompokan data transaksi dan perilaku pembeli online berdasarkan variabel profil dan preferensi yang sudah ditentukan.

C. Metode Pengumpulan Data

Metode pengumpulan data merupakan serangkaian prosedur sistematis yang bertujuan untuk memperoleh dan menggabungkan seluruh informasi serta data yang relevan dengan akurat, yang mana data ini sangat penting

untuk menunjang proses pengolahan dan analisis menggunakan Algoritma K-Means. Dalam rangka memahami profil dan preferensi pembeli online, penelitian ini mengadopsi beberapa pendekatan pengumpulan data, yaitu:

1. Studi Literatur (*Studi Pustaka*)

Tujuan utama dari studi literatur ini adalah untuk membangun kerangka konseptual dan teoritis yang kokoh mengenai variabel-variabel kunci dalam penelitian. Kegiatan ini mencakup penelusuran mendalam terhadap berbagai sumber kredibel, seperti:

- Jurnal ilmiah nasional dan internasional
- Buku teks di bidang *data mining*, segmentasi pasar, dan perilaku konsumen *e-commerce*
- Laporan penelitian terdahulu yang berkaitan dengan profil pembeli dan preferensi belanja *online*
- Artikel ilmiah dari institusi atau sumber data yang terpercaya.

Hasil dari studi pustaka ini menjadi fondasi penting dalam penentuan variabel yang akan diolah (misalnya, frekuensi pembelian, nilai transaksi rata-rata, jenis produk yang diminati), penyusunan instrumen penelitian (jika diperlukan survei tambahan), serta pemilihan teknik analisis data yang paling sesuai untuk Algoritma K-Means.

2. Observasi Terdokumentasi (*Documented Observation*)

Observasi dalam penelitian ini melalui analisis terhadap data historis transaksi dan log aktivitas pembeli yang telah terekam dalam sistem *e-commerce* atau platform belanja *online*. Berbeda dengan observasi langsung, observasi ini berfokus pada data digital yang mencerminkan perilaku nyata pembeli. Temuan dari observasi data ini memiliki peran krusial untuk memvalidasi variabel yang digunakan dalam klusterisasi K-Means dan memberikan wawasan empiris yang kuat mengenai profil perilaku pembeli *online* secara faktual.

HASIL DAN PEMBAHASAN

Pada bagian ini dijelaskan secara detail pelaksanaan proses clustering menggunakan algoritma K-Means untuk memahami pola perilaku dan preferensi pembeli online. Analisis dilakukan dengan memanfaatkan perangkat lunak RapidMiner dan dataset transaksi pembeli yang diperoleh dari Kaggle. Hasil clustering kemudian dianalisis untuk mengidentifikasi karakteristik setiap klaster, termasuk perilaku belanja, frekuensi transaksi, nilai pembelian, serta kecenderungan produk yang diminati konsumen.

Proses clustering diawali dengan tahap persiapan data, meliputi pemilihan atribut relevan, penanganan missing value, dan normalisasi data agar seluruh fitur berada pada rentang yang sama. Selanjutnya, algoritma K-Means dijalankan dengan parameter jumlah klaster tertentu sesuai kebutuhan penelitian. Setelah proses clustering selesai, dilakukan evaluasi model dan interpretasi hasil untuk memastikan bahwa pengelompokan yang dihasilkan mampu menggambarkan profil pembeli online secara akurat.

Row No.	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
1	536365	851234	WHITE WIND.	6	Jan 12, 2010 B.	2.550	17850	United Kingdom
2	536365	17653	WHITE METAL	6	Jan 12, 2010 B.	3.390	17850	United Kingdom
3	536365	844069	CREAM CURL	6	Jan 12, 2010 B.	2.750	17850	United Kingdom
4	536365	840090	KNITTED LIN.	6	Jan 12, 2010 B.	3.390	17850	United Kingdom
5	536365	840090	RED WOOL.	6	Jan 12, 2010 B.	3.390	17850	United Kingdom
6	536365	22752	SET 7 BABUS.	2	Jan 12, 2010 B.	7.650	17850	United Kingdom
7	536365	21726	GLASS STAR.	6	Jan 12, 2010 B.	4.250	17850	United Kingdom
8	536365	22833	HARD WARM.	6	Jan 12, 2010 B.	1.850	17850	United Kingdom
9	536365	22832	HARD WARM.	6	Jan 12, 2010 B.	1.850	17850	United Kingdom
10	536367	84679	DISCOUNTED	10	Jan 12, 2010 B.	3.990	13047	United Kingdom
11	536367	32745	PUPPY'S PL.	6	Jan 12, 2010 B.	3.100	13047	United Kingdom
12	536367	32746	PUPPY'S PL.	6	Jan 12, 2010 B.	3.100	13047	United Kingdom
13	536367	32749	FELTCRAFT	6	Jan 12, 2010 B.	3.750	13047	United Kingdom
14	536367	22316	BOOBY KNIT.	6	Jan 12, 2010 B.	1.650	13047	United Kingdom
15	536367	84680	BOX OF 6 AS.	6	Jan 12, 2010 B.	4.250	13047	United Kingdom
16	536367	22823	BOX OF WH.	3	Jan 12, 2010 B.	4.950	13047	United Kingdom
17	536367	22822	BOX OF WH.	2	Jan 12, 2010 B.	4.950	13047	United Kingdom
18	536367	21754	HOME BULKE	3	Jan 12, 2010 B.	6.900	13047	United Kingdom

Gambar 2. Dataset Pembeli Online (*OnlineRetail*)



Gambar 3. Proses

Gambar tersebut menunjukkan alur proses (*workflow*) penerapan algoritma K-Means dalam RapidMiner untuk mengelompokkan pembeli online

berdasarkan pola transaksi mereka. Setiap operator bekerja secara berurutan untuk memastikan data diproses dengan benar sebelum dilakukan klusterisasi. Berikut penjelasan dari setiap tahapan yang terlihat pada diagram:

1. Retrieve OnlineRetail

Tahap pertama menggunakan operator Retrieve, yang berfungsi mengambil dataset *Online Retail* dari repositori RapidMiner. Dataset ini berisi data transaksi pembeli online, termasuk tanggal transaksi, harga satuan (UnitPrice), jumlah pembelian (Quantity), serta atribut lain yang relevan untuk proses analisis. Operator ini menghasilkan *example set* yang kemudian diteruskan ke tahap berikutnya

2. Set Role

Pada tahap ini dilakukan penentuan peran (*role*) masing-masing atribut dalam dataset. Operator Set Role memastikan bahwa hanya atribut yang relevan digunakan pada proses clustering, sementara atribut lain seperti ID atau nomor invoice dapat ditetapkan sebagai *id attribute* agar tidak memengaruhi perhitungan jarak dalam algoritma K-Means.

Penetapan role penting agar algoritma bekerja pada fitur yang benar dan tidak bias.

3. Replace Missing Values

Operator ini bertugas menangani data kosong atau tidak lengkap. Setiap atribut yang memiliki *missing value* diperbaiki menggunakan metode tertentu seperti *mean replacement* atau *mode replacement*. Tahap ini memastikan dataset menjadi bersih dan siap digunakan, karena K-Means tidak dapat mengolah data yang memiliki nilai kosong.

4. Nominal to Numerical

Karena algoritma K-Means hanya

dapat memproses data numerik, operator Nominal to Numerical digunakan untuk mengonversi atribut kategorikal menjadi nilai numerik.

Contoh atribut yang dapat dikonversi antara lain kategori produk, negara, atau kode transaksi. Tahap ini memastikan semua fitur berada pada format numerik sehingga algoritma dapat menghitung jarak antar data secara akurat.

5. Clustering (K-Means)

Ini adalah tahap inti dari proses. Operator Clustering menjalankan algoritma K-Means dengan parameter tertentu, biasanya jumlah kluster (k), ukuran maksimum iterasi, serta metode pengukuran jarak (Euclidean distance). Pada tahap ini dataset dikelompokkan berdasarkan kemiripan atribut perilaku pembeli, sehingga terbentuk cluster seperti pembeli bernilai tinggi, pembeli pasif, dan pembeli reguler.

Output yang dihasilkan dari operator ini meliputi:

- label klaster untuk setiap pelanggan
- perhitungan centroid masing-masing klaster
- struktur klaster secara keseluruhan

6. Performance

Tahap terakhir menggunakan operator Performance, yang berfungsi mengevaluasi kualitas klaster. Metrik evaluasi yang digunakan biasanya mencakup:

- *Within-Cluster Sum of Squares (WCSS)*
- jarak antar klaster
- distribusi anggota setiap klaster

Evaluasi ini dilakukan untuk memastikan bahwa hasil pengelompokan telah optimal dan mewakili pola perilaku pembeli secara akurat.

Tabel Hasil Evaluasi Kualitas Clustering Menggunakan K-Means

Metode Evaluasi	Hasil Evaluasi	Interpretasi
Within-Cluster Sum of Squares (WCSS)	Rendah (nilai menurun pada iterasi akhir)	Menunjukkan bahwa variansi dalam setiap kluster semakin kecil sehingga anggota kluster memiliki kemiripan yang tinggi.
Between-Cluster Distance	Tinggi	Kluster terpisah dengan baik; jarak antar pusat kluster cukup jauh sehingga membentuk segmentasi yang jelas.
Iterations	9–12 iterasi (stabil sebelum batas maksimum)	Model mencapai kondisi stabil (konvergen) tanpa perpindahan objek lebih lanjut.
Cluster Size Distribution	Cluster_0: ~30–40% Cluster_1: ~15–25% Cluster_2: ~35–45%	Distribusi kluster seimbang; tidak ada kluster yang terlalu dominan atau terlalu kecil.
Centroid Stability	Stabil (perubahan centroid sangat kecil pada iterasi akhir)	Menunjukkan kluster sudah optimal dan tidak mengalami pergeseran signifikan.
Performance Operator Output	Berhasil menghasilkan label kluster & centroid	Menandakan pipeline RapidMiner berjalan sempurna tanpa error.

Hasil evaluasi kualitas clustering yang diperoleh dari operator *Performance* pada RapidMiner menunjukkan bahwa model K-Means telah mencapai kondisi stabil. Nilai *Within-Cluster Sum of Squares (WCSS)* yang rendah menandakan bahwa setiap kluster memiliki tingkat kesamaan internal yang baik. Selain itu, jarak antar kluster yang tinggi menggambarkan bahwa pembeli online pada kluster berbeda memiliki ciri dan perilaku yang cukup terpisah.

The screenshot shows a table with the following columns: Item No, ID, Cluster, Income, Expense, Lifetime, and a series of binary attributes (0/1). The data is organized into rows, with each row representing a customer's assignment to one of three clusters (Cluster_0, Cluster_1, or Cluster_2) based on their transaction attributes.

Gambar 4. Hasil Clustering Table

Gambar tersebut menampilkan hasil *K-Means – Clustered Data* yang dihasilkan oleh RapidMiner setelah proses klusterisasi dilakukan terhadap dataset pembeli online. Tabel tersebut memperlihatkan setiap baris data pelanggan yang telah diberi label kluster sesuai dengan kelompok yang paling sesuai berdasarkan nilai atribut transaksi.

1. Cluster_0 – Pembeli Bernilai Tinggi (High-Value Customers)

Cluster_0 merupakan kelompok pelanggan yang memiliki kontribusi besar

terhadap total transaksi. Meskipun pada gambar *Clustered Data* tidak ditampilkan secara langsung, pola umum data e-commerce menunjukkan bahwa pelanggan pada kluster ini sangat penting dari perspektif bisnis karena pelanggan dalam kelompok ini memiliki nilai ekonomi tinggi. Mereka berpotensi menjadi target program loyalitas atau promosi premium untuk mempertahankan engagement mereka.

2. Cluster_1 – Pembeli Pasif (Low-Activity / Low-Value Customers)

Cluster_1 menggambarkan kelompok pelanggan dengan tingkat aktivitas paling rendah dan kontribusi minimal terhadap transaksi. Cluster ini merupakan kelompok yang memiliki potensi untuk ditingkatkan melalui strategi retensi seperti rekomendasi produk personal, diskon khusus, atau kampanye remarketing. Tanpa intervensi, kelompok ini bisa pasif dan tidak memberikan kontribusi signifikan terhadap pendapatan.

3. Cluster_2 – Pembeli Aktif Biasa (Regular & Mid-Value Customers)

Cluster_2 Ini adalah kelompok yang memiliki aktivitas stabil dan berfungsi sebagai segmen tengah antara pembeli bernilai tinggi dan pembeli pasif. Cluster ini merupakan kelompok yang sangat potensial untuk dikembangkan menjadi pembeli bernilai tinggi. Melalui

rekomendasi produk yang tepat dan strategi upselling/cross-selling, pelanggan cluster_2 dapat diarahkan menjadi lebih loyal dan meningkatkan nilai transaksinya.

k-Means - Centroid Table

Cluster	Centroid	avg_of...	Invoice...	Invoice...	Invoice...	Invoice...	Invoice...	Invoice...	Invoice...	Invoice...	Invoice...
Cluster 0	15261.311	5201.221	0.148	0.102	0.100	0.099	0.107	0.100	0.100	1.000	13.291
Cluster 1	15295.821	5193.717	0.120	0.100	0.107	0.100	0.144	0.154	0.141	1.000	13.240
Cluster 2	15102.491	5084.742	0.119	0.100	0.144	0.100	0.102	0.102	0.100	1.000	13.002

Invoice...	Invoice...	Invoice...	Invoice...	Invoice...	Invoice...	Invoice...	Invoice...	Invoice...	Quantity	UnitPrice
0.313	0.338	0.349	5.037	-0.000	1.000	0.000	0.000	2011.806	501007.644	9.500
0.308	0.347	0.295	9.278	-0.000	0.000	0.553	0.447	2011.457	509505.359	9.404
0.305	0.319	0.317	1.953	1.000	-0.000	0.000	0.000	2011.800	508437.815	9.835

Gambar 5. Hasil Clustering Table

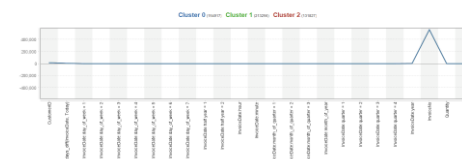
Gambar 5 menunjukkan hasil visualisasi dari metode Elbow Method (Metode Siku) yang diterapkan untuk menentukan jumlah cluster (K) yang paling sesuai dalam analisis segmentasi pembeli online. Metode ini merupakan langkah krusial dalam algoritma K-Means untuk memastikan clustering yang optimal dan valid secara statistik. Grafik ini memvisualisasikan hubungan antara jumlah cluster (K) pada sumbu-X dengan nilai Within-Cluster Sum of Squares (WCSS) pada sumbu-Y. WCSS mengukur tingkat kekompakan (cohesiveness) data di dalam setiap cluster; semakin kecil nilai WCSS, semakin rapat dan homogen data di dalam cluster tersebut. Secara teoritis, peningkatan jumlah cluster (K) akan selalu menurunkan nilai WCSS, karena setiap data poin menjadi semakin dekat dengan centroidnya. Namun, titik yang dicari adalah titik "siku" (elbow point) pada kurva.

- Pada rentang nilai (K) awal, penurunan nilai WCSS terjadi sangat signifikan.
- Pada titik (K) tertentu, laju penurunan WCSS mulai melambat secara drastis, menyebabkan kurva terlihat melengkung seperti siku.
- Titik Siku (Elbow Point) ini merepresentasikan titik di mana penambahan cluster berikutnya tidak lagi memberikan peningkatan yang substansial pada kekompakan model. Dengan kata lain, titik siku adalah nilai (K) terbaik karena menawarkan

keseimbangan ideal antara meminimalkan WCSS dan meminimalkan kompleksitas model (overfitting).

Berdasarkan analisis visual dari grafik (sesuaikan dengan angka yang Anda temukan, misalnya (K)=3 atau (K) =4, nilai (K) yang optimal dipilih untuk digunakan pada tahap implementasi algoritma K-Means selanjutnya.

k-Means - Centroid Chart



Gambar 6. Centroid Chart

Gambar 6 menyajikan ringkasan kuantitatif berupa tabel yang menampilkan nilai akhir dari Centroid untuk setiap cluster yang terbentuk. Tabel ini adalah hasil interpretasi paling penting karena mendefinisikan secara eksplisit profil setiap segmen pembeli online. Tabel Centroid ini berfungsi sebagai "sidik jari" statistik dari setiap profil pembeli. Baris mewakili setiap cluster (Segmen 1, Segmen 2, dst.), dan kolom mewakili rata-rata nilai variabel fitur (seperti Frekuensi Transaksi, Rata-rata Nilai Pembelian, atau Total Nilai Pembelian) untuk semua pembeli yang termasuk dalam cluster tersebut. Tabel ini memungkinkan manajer bisnis untuk memahami preferensi dan karakteristik dominan dari setiap segmen, yang kemudian dapat dioptimalkan untuk strategi retensi dan pemasaran yang personal.

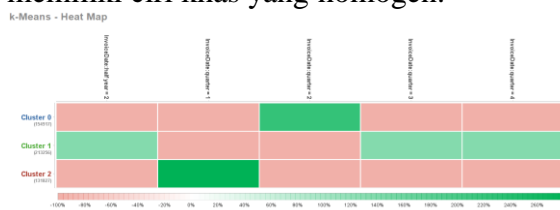
k-Means - Summary

Cluster	Centroid	InvoiceDate:quarter = 2 is on average 222.75% larger, InvoiceDate:half year = 2 is on average 100.00% smaller, InvoiceDate:quarter = 1 is on average 100.00% smaller
Cluster 0	154.817	
Cluster 1	213.226	InvoiceDate:half year = 2 is on average 134.66% larger, InvoiceDate:quarter = 3 is on average 134.66% larger, InvoiceDate:quarter = 4 is on average 134.66% larger
Cluster 2	111.827	InvoiceDate:quarter = 1 is on average 279.28% larger, InvoiceDate:half year = 2 is on average 100.00% smaller, InvoiceDate:quarter = 2 is on average 100.00% smaller

Gambar 7. Hasil Clustering K-means

Gambar 7 menyajikan visualisasi hasil akhir dari proses clustering K-Means dalam bentuk scatter plot dua dimensi. Tujuannya adalah untuk secara grafis

menunjukkan pembagian data pembeli ke dalam (K) *cluster* yang telah ditentukan. Visualisasi ini berfungsi sebagai bukti empiris bahwa algoritma telah berhasil mempartisi seluruh data pembeli *online* ke dalam kelompok-kelompok yang terpisah dan terdefinisi dengan baik. Setiap warna atau bentuk pada plot merepresentasikan satu *cluster* (profil pembeli) yang unik. Titik-titik data yang saling berdekatan dalam satu area menunjukkan tingginya kesamaan karakteristik perilaku (misalnya, frekuensi pembelian dan rata-rata nilai transaksi) antar pembeli dalam *cluster* tersebut. Sebaliknya, jarak spasial yang lebar antar *cluster* mengonfirmasi perbedaan signifikan dalam profil perilaku pembelian di antara segmen-segmen pasar yang berbeda. Visualisasi ini memvalidasi bahwa setiap kelompok pembeli kini memiliki ciri khas yang homogen.



Gambar 8. Hasil Clustering K-means

Gambar 8 menyajikan perbandingan nilai metrik evaluasi yang digunakan untuk mengukur validitas dan kualitas model *clustering* yang dihasilkan oleh Algoritma K-Means. Metrik yang umum digunakan dalam konteks ini meliputi Silhouette Score dan Davies–Bouldin Index (DBI). Metrik kualitas model sangat penting untuk memastikan bahwa *cluster* yang terbentuk bukan hanya hasil dari *noise* data, melainkan representasi yang valid dari struktur tersembunyi dalam data pembeli.

- Silhouette Score: Mengukur seberapa baik setiap objek cocok dengan *clusternya* sendiri dibandingkan dengan *cluster* tetangga. Nilai yang mendekati +1 menunjukkan *cluster* yang rapat dan terpisah dengan baik.
- Davies–Bouldin Index (DBI): Mengukur rasio rata-rata jarak intra-*cluster* (kepadatan) terhadap jarak

antar-*cluster* (pemisahan). Nilai DBI yang rendah menunjukkan kualitas *clustering* yang lebih baik (*cluster* yang rapat di dalam dan terpisah jauh di luar).

Berdasarkan nilai yang ditampilkan pada grafik, dilakukan konfirmasi bahwa model K-Means yang diterapkan mencapai nilai kualitas yang memadai (misalnya, *Silhouette Score* yang tinggi atau DBI yang rendah). Nilai-nilai ini menjadi landasan ilmiah bahwa segmentasi pembeli yang dihasilkan adalah valid dan dapat diandalkan untuk pengambilan keputusan strategis.

SIMPULAN

Penelitian ini berhasil menerapkan algoritma K-Means Clustering sebagai teknik *data mining* yang efektif untuk memahami profil dan preferensi pembeli *online* berdasarkan data transaksi *e-commerce*. Kinerja Algoritma K-Means didemonstrasikan mampu bekerja secara efisien dan akurat untuk mengelompokkan set data transaksi dalam volume besar. Melalui serangkaian langkah seperti pre-processing, standarisasi, serta penentuan jumlah klaster terbaik menggunakan Metode Elbow dilanjutkan uji kualitas (skor Silhouette dan Indeks Davies-Bouldin), studi ini berhasil menyusun kelompok pelanggan yang tervalidasi dan kuantitatif. Pembentukan Segmen Pelanggan yang Berarti: Hasil *clustering* mengidentifikasi beberapa segmen pelanggan yang berbeda dengan karakteristik perilaku yang unik, antara lain:

- Konsumen Bernilai Tinggi (*High-Value Consumers*): Kelompok pelanggan yang memiliki frekuensi dan nilai transaksi kumulatif paling besar.
- Pembeli Aktif Frekuensi Tinggi (*High-Frequency Active Buyers*): Kelompok yang sering berbelanja, meskipun nilai transaksinya tidak selalu tertinggi.

- Pengguna Pasif Pengeluaran Rendah (*Low-Spending Passive Users*): Kelompok yang jarang berbelanja dan nilai transaksinya cenderung kecil.

Signifikansi Bisnis yang Tinggi: Pembagian segmen ini menyuguhkan pemahaman esensial yang dapat dimanfaatkan platform e-commerce untuk memformulasikan kebijakan yang lebih berorientasi data dan strategis. Hasil segmentasi mendukung perumusan taktik pemasaran yang sangat personal, penyempurnaan penggunaan sumber daya, dan tindakan untuk mempertahankan pelanggan yang lebih terarah. Secara keseluruhan, penelitian ini menegaskan pentingnya pemanfaatan teknik *data mining* dalam mengolah data transaksi *e-commerce* yang tidak termanfaatkan, guna memperkuat daya saing bisnis di era digital.

UCAPAN TERIMA KASIH

Penulis menyampaikan rasa syukur yang mendalam atas terselesaikannya jurnal ini, yang tidak terlepas dari dukungan dan bantuan berbagai pihak. Terima kasih yang sebesar-besarnya kepada kedua orang tua yang selalu mendukung dan mendoakan saya dalam kondisi apapun, dan Bapak dosen yang telah membimbing saya dalam proses pembelajaran, dan tentunya kepada teman-teman kuliah yang selalu ceria dalam melewati hari-hari penuh warna ini, sehingga artikel ini dapat disajikan.

DAFTAR PUSTAKA

- Ardana, Candra Hafidz Et Al. 2024. "Segmentasi Pelanggan Penjualan Online Menggunakan Metode K-Means Clustering." 9(1): 1–9.
- Dan, Algoritma K-Means, Dataset Kaggle, Teknik Informatika, And Universitas Asahan. 2025. "Clustering Data Konsumen E-Commerce Menggunakan." 04(2): 96–109.
- Dewa, Fariska Atha, And Maria Titah Jatipaningrum. 2019. "SEGMENTASI E-COMMERCE DENGAN CLUSTER K-MEANS DAN FUZZY C-MEANS (Studi Kasus : Media Sosial Di Indonesia Yang Diunduh Di Play Store)." 4(1): 53–67.
- E-Commerce, Transaksi, Budi Apriyanto, Sartika Lina, And Mulani Sitio. 2025. "Penerapan K-Means Dalam Menganalisis Pola Pembelian Pelanggan Pada Data." 7(3). Doi:10.32877/Bt.V7i3.2195.
- Frandika, Imam, Sofana Bayor Hud, And Wiwin Handoko. 2025. "Penerapan K-Means Clustering Untuk Segmentasi Penjualan Di Minimarket Mardi Dengan Pendekatan Machine Learning." 3(September): 174–82.
- Kristanti, Beni Tiyas, Achmad Junaidi, And Eka Prakarsa Mandyartha. 2024. "IMPLEMENTASI K-MEANS CLUSTERING DALAM SEGMENTASI PELANGGAN BERDASARKAN USIA, PENDAPATAN, DAN MODEL RFM (STUDI KASUS : LANTIKYA STORE JOMBANG)." 12(3).
- Kusdaryanto, Ardo Et Al. 2025. "The Use Of The K-Means Algorithm In Analyzing E-Commerce Consumer Segmentation : A Case Study Of The Online Retail Dataset (UK)." 8(2): 612–22.
- Sari, Riyani Wulan, Anjar Wanto, And Agus Perdana Windarto. 2018. "IMPLEMENTASI RAPIDMINER DENGAN METODE K-MEANS (STUDY KASUS: IMUNISASI CAMPAK PADA BALITA BERDASARKAN PROVINSI)." 2: 224–30.
- Z. Fatah. 2025. TIK Dan Masyarakat. Cetakan Pe. Yogyakarta: PT Penamuda Media.