

SEGMENTASI PELANGGAN PADA TOKO XYZ BERDASARKAN POLA PEMBELIAN MENGGUNAKAN ALGORITMA K-MODES

Amanda Rizky Putri Wibowo, Zahra Humaira Salsabila, Fathia Bilqish Wulandari, Harun Al Rosyid

Pendidikan Teknologi Informasi, Universitas Negeri Surabaya

Kampus Unesa 1, Jalan Ketintang, Surabaya 60231

amandarputriw@gmail.com

ABSTRAK

Perkembangan teknologi informasi dan meningkatnya persaingan bisnis ritel menuntut pelaku usaha untuk memahami perilaku pelanggan melalui analisis pola pembelian berbasis data. Toko XYZ masih menghadapi kendala karena segmentasi pelanggan yang diterapkan belum berbasis analisis data, sehingga strategi pemasaran yang dijalankan kurang spesifik dan optimal. Penelitian ini bertujuan untuk membentuk segmentasi pelanggan berdasarkan pola pembelian dengan menerapkan algoritma K-Modes pada data transaksi Toko XYZ. Penelitian ini menggunakan pendekatan kuantitatif dengan metode *unsupervised learning* terhadap 1.448 pelanggan yang mencakup variabel Order Substatus, Provinsi, Metode Pembayaran, Kategori Produk, serta nilai transaksi yang telah dikonversi ke dalam kategori rentang harga. Tahapan penelitian meliputi *pre-processing* data, penentuan jumlah kluster optimal menggunakan *Elbow Method*, penerapan algoritma K-Modes, serta evaluasi hasil segmentasi melalui analisis deskriptif. Hasil penelitian menunjukkan terbentuknya enam kluster pelanggan dengan karakteristik yang berbeda, yaitu *High-Value Selective Buyer* (25,00%), *Regular Mid-Value Buyer* (21,20%), *Occasional Moderate Buyer* (11,46%), *Price-Sensitive Economy Buyer* (17,27%), *Budget-Conscious Value Seeker* (21,06%), dan *Low-Frequency Planned Buyer* (4,01%). Setiap kluster mencerminkan perbedaan pola pembelian, metode pembayaran, dan rentang harga transaksi. Temuan ini dapat dimanfaatkan sebagai dasar perumusan strategi pemasaran, pengelolaan persediaan, dan peningkatan layanan pelanggan yang lebih terarah.

Kata kunci : data mining, segmentasi pelanggan, pola pembelian, algoritma k-modes

1. PENDAHULUAN

Perkembangan teknologi informasi telah mendorong perubahan signifikan dalam pola konsumsi masyarakat serta strategi bisnis pada sektor ritel. Pemanfaatan teknologi digital, seperti e-commerce dan analitika data, memungkinkan pelaku usaha meningkatkan efisiensi operasional sekaligus memahami kebutuhan pelanggan secara lebih mendalam [1]. Persaingan yang semakin ketat menuntut perusahaan tidak hanya berfokus pada kualitas produk, tetapi juga pada kemampuan mengelola dan menganalisis data pelanggan agar strategi pemasaran yang diterapkan lebih tepat sasaran dan berkelanjutan [2]. Analisis data transaksi pelanggan menjadi sumber informasi penting untuk mengidentifikasi pola pembelian, preferensi, serta karakteristik perilaku konsumen dalam konteks bisnis ritel [3].

Salah satu pendekatan yang banyak digunakan untuk memahami perilaku pelanggan adalah teknik segmentasi, yaitu metode *unsupervised learning* yang bertujuan mengelompokkan data berdasarkan kesamaan karakteristik tertentu [4]. Penerapan segmentasi dalam segmentasi pelanggan memungkinkan pelaku usaha memperoleh gambaran yang lebih objektif mengenai struktur pelanggan, sehingga dapat mendukung perumusan strategi pemasaran, promosi, dan pengelolaan hubungan pelanggan secara lebih terarah [5]. Meskipun demikian, sebagian pelaku usaha ritel masih menghadapi kendala karena segmentasi pelanggan yang dilakukan belum sepenuhnya berbasis analisis

data, melainkan masih mengandalkan pendekatan umum yang kurang mencerminkan perbedaan perilaku pelanggan secara nyata.

Permasalahan tersebut juga terjadi pada Toko XYZ, di mana data transaksi pelanggan telah tersedia namun belum dimanfaatkan secara optimal untuk membentuk segmentasi pelanggan berbasis pola pembelian. Sebagian besar atribut pada data transaksi Toko XYZ bersifat kategorikal, seperti kategori produk, metode pembayaran, dan status pesanan, sehingga memerlukan pendekatan segmentasi yang sesuai dengan karakteristik data tersebut. Penggunaan algoritma segmentasi numerik berpotensi menghasilkan segmentasi yang kurang representatif apabila diterapkan langsung pada data kategorikal [6].

Berdasarkan kondisi tersebut, penelitian ini bertujuan untuk melakukan segmentasi pelanggan Toko XYZ berdasarkan pola pembelian menggunakan algoritma K-Modes. Algoritma K-Modes dipilih karena dirancang khusus untuk menangani data kategorikal dengan menggunakan modus sebagai pusat kluster dan ukuran ketidaksamaan berbasis kecocokan kategori [7]. Melalui penerapan algoritma ini, diharapkan dapat dihasilkan kelompok pelanggan dengan karakteristik yang jelas dan mudah diinterpretasikan, sehingga dapat dimanfaatkan sebagai dasar pengambilan keputusan strategis.

Sebagai upaya penyelesaian permasalahan, penelitian ini dilakukan melalui tahapan pengolahan data transaksi pelanggan, penentuan jumlah kluster optimal, penerapan algoritma K-Modes, serta analisis hasil segmentasi. Pendekatan ini diharapkan mampu

memberikan kontribusi praktis bagi Toko XYZ dalam menyusun strategi pemasaran dan pengelolaan pelanggan yang lebih efektif, sekaligus memperkaya kajian penerapan segmentasi data kategorikal pada konteks bisnis ritel [8].

Selain itu, penelitian terbaru menunjukkan bahwa integrasi algoritma K-Modes dengan teknik pembelajaran lanjutan dapat memperbaiki kualitas segmentasi dengan menangkap pola perilaku pelanggan yang unik dan heterogen. Misalnya, penerapan K-Modes yang dikombinasikan dengan pembelajaran ensemble meningkatkan akurasi segmentasi dibandingkan dengan pendekatan tunggal pada data profil pelanggan yang kompleks [9]. Hal ini relevan dengan *Segmentasi Pelanggan Toko XYZ* karena pemanfaatan atribut non-numerik yang dominan dalam dataset menuntut metode segmentasi yang mempertahankan informasi kategorikal secara penuh tanpa distorsi numerik. Dengan demikian, penggunaan K-Modes dalam penelitian ini tidak hanya meningkatkan pemahaman tentang pola pembelian tetapi juga menyediakan dasar empiris untuk strategi pemasaran dan pengelolaan operasional berdasarkan karakteristik masing-masing segmen pelanggan [10].

2. TINJAUAN PUSTAKA

2.1. Penelitian Terkait

Beberapa penelitian sebelumnya telah menerapkan metode segmentasi untuk segmentasi pelanggan dan analisis pola data kategorikal. [11] melakukan segmentasi pelanggan agen *beauty local brand* menggunakan algoritma K-Modes berbasis *Python* dan berhasil membentuk beberapa segmen pelanggan berdasarkan perilaku pembelian serta karakteristik transaksi. Penelitian tersebut menunjukkan bahwa K-Modes efektif dalam mengelompokkan data kategorikal dan memberikan wawasan praktis bagi pemilik bisnis untuk menyusun strategi pemasaran yang lebih tepat sasaran.

Selain itu, [3] menerapkan algoritma K-Modes untuk menentukan strategi pemasaran di sebuah perguruan tinggi dengan mengelompokkan calon mahasiswa berdasarkan data kategorikal hasil pendaftaran. Hasil penelitian mereka menunjukkan bahwa penggunaan K-Modes memudahkan pihak manajemen dalam memetakan segmen calon mahasiswa dan merancang pendekatan promosi yang berbeda untuk setiap segmen. Penelitian ini menguatkan bahwa K-Modes relevan digunakan pada data kategorikal dalam konteks pengambilan keputusan strategis.

Berbagai penelitian menegaskan pentingnya penggunaan algoritma segmentasi yang dirancang khusus untuk menangani data kategorikal. [12] memperkenalkan algoritma K-Modes sebagai pengembangan dari K-Means dengan mengganti pusat kluster menjadi modus serta menggunakan ukuran ketidaksamaan berbasis kecocokan kategori, sehingga mampu menghasilkan kluster yang lebih stabil dan representatif pada data nominal. Pengembangan lebih

lanjut dilakukan melalui pendekatan *fuzzy k-modes*, yang memungkinkan satu objek memiliki derajat keanggotaan pada lebih dari satu kluster dan terbukti meningkatkan fleksibilitas serta kualitas kluster pada data kategorikal yang heterogen [12]. Selain itu, studi mengenai *multivariate fuzzy weighted k-modes* menunjukkan bahwa pemberian bobot atribut dan penggunaan jarak probabilistik dapat meningkatkan akurasi segmentasi, terutama ketika atribut memiliki tingkat kepentingan yang berbeda-beda [13].

Dalam konteks segmentasi pelanggan, penerapan K-Modes juga terbukti efektif dalam mengelompokkan pelanggan berdasarkan atribut non-numerik. [10] menerapkan algoritma K-Modes pada data pengguna Netflix dan berhasil mengidentifikasi segmen pelanggan yang memiliki karakteristik perilaku berbeda berdasarkan preferensi layanan dan pola penggunaan. Temuan tersebut relevan dengan penelitian pada Toko XYZ, di mana sebagian besar atribut yang digunakan, seperti kategori produk, metode pembayaran, status pesanan, dan rentang harga transaksi, bersifat kategorikal. Oleh karena itu, pemilihan algoritma K-Modes dalam penelitian ini dinilai lebih tepat dibandingkan algoritma segmentasi numerik, karena mampu menangkap pola pembelian pelanggan secara lebih akurat dan interpretatif.

2.2. Segmentasi Pelanggan

Segmentasi pelanggan adalah proses membagi pelanggan ke dalam beberapa kelompok yang relatif homogen berdasarkan kesamaan karakteristik atau perilaku pembelian sehingga setiap segmen memiliki pola dan kebutuhan yang berbeda [14]. Dalam konteks ritel dan *e-commerce*, segmentasi pelanggan umumnya dilakukan dengan teknik clustering untuk mendukung perancangan strategi pemasaran, program promosi, dan layanan yang lebih terarah sesuai nilai dan perilaku tiap segmen [5]. Menurut [15], segmentasi pelanggan berbasis data transaksi, misalnya menggunakan atribut frekuensi, nilai pembelian, atau kategori produk, terbukti mampu meningkatkan efektivitas strategi pemasaran dan pengelolaan hubungan pelanggan karena tiap kelompok dapat diperlakukan secara berbeda sesuai karakteristiknya. Pada Toko XYZ, segmentasi pelanggan dilakukan dengan mengelompokkan pelanggan berdasarkan kombinasi rentang harga pembelian, kategori produk, metode pembayaran, dan status pesanan, sehingga dapat diidentifikasi segmen pelanggan premium, menengah, hingga pelanggan dengan daya beli lebih rendah sebagai dasar perumusan strategi pemasaran yang lebih spesifik..

2.3. Pola Pembelian Pelanggan

Pola pembelian pelanggan merefleksikan kebiasaan dan preferensi konsumen ketika melakukan transaksi, meliputi jenis produk yang dipilih, waktu dan frekuensi pembelian, nominal transaksi, serta metode pembayaran yang digunakan [16]. Analisis pola pembelian memungkinkan perusahaan membaca

perilaku konsumen secara lebih mendalam dan menyusun strategi penawaran yang relevan, misalnya menentukan produk yang perlu diprioritaskan, merancang bundling, atau menetapkan skema diskon tertentu [4]. Dalam penelitian ini, pola pembelian pelanggan Toko XYZ direpresentasikan melalui variabel *Product Category*, *Payment Method*, *Order Substatus*, *Province*, dan *SKU Subtotal After Discount* yang telah dikonversi ke dalam kategori rentang harga, sehingga dapat dianalisis menggunakan teknik segmentasi data kategorikal.

2.4. Algoritma K-Modes

Algoritma K-Modes merupakan pengembangan dari K-Means yang dirancang khusus untuk menangani data kategorikal dengan mengganti pusat kluster dari nilai rata-rata (mean) menjadi modus (mode) dan menggunakan ukuran ketidaksamaan berbasis kecocokan kategori [3]. Pendekatan ini menghindari penggunaan jarak *Euclidean* yang tidak tepat untuk data kategorikal, sehingga lebih sesuai untuk kasus di mana seluruh atau sebagian besar atribut berbentuk label, seperti kategori produk, metode pembayaran, dan kelas rentang harga.

Secara umum, K-Modes mengubah K-Means dalam beberapa aspek berikut:

1. Menggunakan ukuran kecocokan kategori sederhana pada data bertipe kategori.
2. Mengganti pusat kluster dengan modus (nilai kategori yang paling sering muncul)
3. Memperbarui pusat kluster berdasarkan frekuensi setiap kategori pada masing-masing atribut

Secara matematis, ukuran ketidaksamaan (*dissimilarity measure*) pada algoritma K-Modes didefinisikan sebagai berikut. Misalkan $X = (x_1, x_2, \dots, x_m)$ adalah suatu objek data dengan m atribut kategorikal, dan $Z = (z_1, z_2, \dots, z_m)$ adalah pusat kluster (modus). Fungsi ketidaksamaan antara X dan Z dirumuskan sebagai:

$$d(X, Z) = \sum_{j=1}^m \delta(x_j, z_j) \quad (1)$$

dengan fungsi indikator $\delta(\cdot)$ didefinisikan sebagai:

$$\delta(x_j, z_j) = \begin{cases} 0, & \text{jika } x_j = z_j \\ 1, & \text{jika } x_j \neq z_j \end{cases} \quad (2)$$

Rumus ini menunjukkan bahwa ketidaksamaan dihitung berdasarkan jumlah atribut yang tidak cocok antara objek data dan pusat kluster. Semakin kecil nilai $d(X, Z)$, maka semakin tinggi tingkat kemiripan objek terhadap kluster tersebut.

Tujuan utama algoritma K-Modes adalah meminimalkan fungsi objektif total, yaitu jumlah ketidaksamaan seluruh data terhadap pusat kluster masing-masing. Fungsi objektif tersebut dirumuskan sebagai:

$$J = \sum_{i=1}^n \sum_{k=1}^K \omega_{ik} d(X_i, Z_k) \quad (3)$$

di mana:

- n adalah jumlah objek data,
- K adalah jumlah kluster,
- ω_{ik} adalah variabel keanggotaan biner yang bernilai 1 jika objek ke- i termasuk kluster ke- k , dan 0 jika tidak,
- Z_k adalah pusat kluster ke- k .

Pembaruan pusat kluster dilakukan dengan memilih nilai kategori yang memiliki frekuensi tertinggi pada setiap atribut di dalam kluster tersebut. Secara formal, modus untuk atribut ke- j pada kluster ke- k ditentukan sebagai:

$$z_{kj} = \arg \max_{c \in C_j} f(c | \text{kluster}_k) \quad (4)$$

dengan C_j Adalah himpunan kategori pada atribut ke- j , dan $f(c | \text{kluster}_k)$ Adalah frekuensi kemunculan kategori c pada kluster ke- k

Dengan mekanisme tersebut, algoritma K-Modes mengubah K-Means dalam beberapa aspek utama, yaitu:

1. Menggunakan ukuran ketidaksamaan berbasis kecocokan kategori, bukan jarak *Euclidean*.
2. Mengganti pusat kluster dari nilai rata-rata menjadi modus kategori.
3. Memperbarui pusat kluster berdasarkan frekuensi kategori pada setiap atribut di dalam kluster.

Sejumlah penelitian terkini menunjukkan bahwa algoritma K-Modes efektif digunakan dalam segmentasi pelanggan berbasis data kategorikal, baik pada konteks layanan digital maupun ritel. Hasil segmentasi yang diperoleh cenderung lebih interpretatif karena pusat kluster direpresentasikan dalam bentuk kategori aktual, sehingga memudahkan pembentukan persona pelanggan dan penerjemahannya ke dalam strategi pemasaran yang lebih spesifik dan berbasis perilaku [17].

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode *unsupervised learning* melalui penerapan algoritma K-Modes untuk melakukan segmentasi pelanggan berdasarkan pola pembelian. Tahapan penelitian disusun secara sistematis untuk memastikan proses analisis data berjalan terstruktur dan menghasilkan segmentasi pelanggan yang valid serta dapat diinterpretasikan secara operasional [11]. Secara umum, metode penelitian terdiri atas lima tahapan utama, yaitu pengumpulan data, *Pre-Processing* data, penentuan jumlah kluster optimal, implementasi algoritma K-Modes, serta evaluasi dan interpretasi hasil segmentasi.

3.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini berupa data transaksi pelanggan yang diperoleh dari platform *e-commerce* toko. Dataset mentah terdiri atas 1.448 transaksi dengan lima variabel utama yang mencerminkan karakteristik pola pembelian pelanggan, yaitu Order ID, Product Category, Payment Method, SKU Subtotal After Discount, dan Order Substatus. Data dikumpulkan dalam format *Microsoft Excel* (.xlsx) dan selanjutnya diolah menggunakan *Python* melalui *Google Colaboratory* untuk memudahkan proses analisis [18].

3.2. Pre-Processing Data

Tahap *Pre-Processing* data dilakukan untuk mempersiapkan dataset agar sesuai dengan kebutuhan algoritma K-Modes yang hanya menerima data bertipe kategorikal. *Pre-processing* merupakan tahap penting dalam *data mining* untuk meningkatkan kualitas data dengan menghilangkan data yang tidak lengkap, tidak relevan, dan tidak konsisten [19]. Proses ini meliputi beberapa langkah sebagai berikut:

3.3. Seleksi fitur dan penanganan Missing Value

Dari dataset awal, dilakukan seleksi terhadap variabel yang relevan dengan pola pembelian pelanggan. Variabel yang dipilih adalah Order ID, Product Category, Payment Method, SKU Subtotal After Discount, dan Order Substatus. Variabel Order ID digunakan sebagai identitas unik transaksi, sedangkan Order Substatus digunakan sebagai label keputusan yang tidak diikutsertakan dalam proses segmentasi. Data yang mengandung nilai kosong (*missing value*) dihapus untuk menjaga stabilitas proses segmentasi menggunakan fungsi *dropna()* pada *library* Pandas. Langkah ini penting karena algoritma K-Modes tidak dapat menangani nilai kosong secara langsung [3].

3.4. Transformasi Data Numerik menjadi Kategorikal

Variabel SKU Subtotal After Discount yang bersifat numerik diubah menjadi data kategorikal dengan membagi nilainya ke dalam empat rentang harga menggunakan metode *quantile-based discretization* (*pd.qcut*). Empat kategori yang dihasilkan adalah "Sangat Rendah", "Rendah", "Sedang", dan "Tinggi". Setelah transformasi, kolom SKU Subtotal After Discount yang masih numerik dihapus dari dataset

3.5. Agregasi Data Pelanggan

Karena satu pelanggan dapat melakukan beberapa transaksi, dilakukan agregasi data berdasarkan Order ID menggunakan fungsi *groupby()* dengan mengambil modus (*mode*) dari setiap variabel kategorikal. Modus dipilih sebagai representasi nilai yang paling sering muncul untuk setiap pelanggan pada atribut Product Category, Payment Method, *price_range*, dan Order Substatus.

Penggunaan modus sebagai metode agregasi data kategorikal telah terbukti efektif dalam penelitian segmentasi sebelumnya [3].

3.6. Penentuan Jumlah Kluster Optimal

Penentuan jumlah kluster optimal dilakukan menggunakan metode *Elbow Method* yang mengukur nilai *cost function* dari algoritma K-Modes. Nilai *cost* dihitung untuk rentang jumlah kluster (*k*) dari 2 hingga 7, kemudian divisualisasikan dalam bentuk grafik. Titik *elbow* pada grafik menunjukkan jumlah kluster yang optimal, yaitu titik di mana penambahan jumlah kluster tidak lagi memberikan penurunan *cost* yang signifikan. Berdasarkan hasil analisis, jumlah kluster optimal yang dipilih adalah *k* = 6.

3.7. Implementasi Algoritma K-Modes

Algoritma K-Modes diimplementasikan menggunakan *library* *kmodes* pada *Python* dengan parameter sebagai berikut:

- n_clusters*: 6 (jumlah kluster optimal berdasarkan *Elbow Method*)
- init*: 'Huang' (metode inisialisasi pusat kluster menggunakan algoritma Huang)
- n_init*: 10 (jumlah iterasi inisialisasi untuk meningkatkan stabilitas hasil)
- verbose*: 1 (menampilkan proses iterasi segmentasi)

Proses segmentasi dilakukan dengan memanggil fungsi *fit_predict()* pada matriks fitur *X* yang terdiri atas variabel Product Category, Payment Method, dan *price_range*. Hasil segmentasi berupa label kluster untuk setiap pelanggan disimpan dalam kolom baru bernama 'cluster' pada dataframe pelanggan.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Pre-Processing Data

Proses *Pre-Processing* data dilakukan terhadap dataset transaksi pelanggan yang terdiri atas 1.448 transaksi dengan lima variabel utama, yaitu Order ID, Product Category, Payment Method, SKU Subtotal After Discount, dan Order Substatus.

Order ID	0
Product Category	0
Payment Method	0
SKU Subtotal After Discount	0
Order Substatus	0

Gambar 1. Fitur *Missing value*

Gambar 1 adalah penanganan *missing value* menggunakan *Google Colaboration* dengan menghapus seluruh baris data yang memiliki nilai kosong, sehingga dataset yang digunakan menjadi bersih dan siap untuk diproses lebih lanjut.

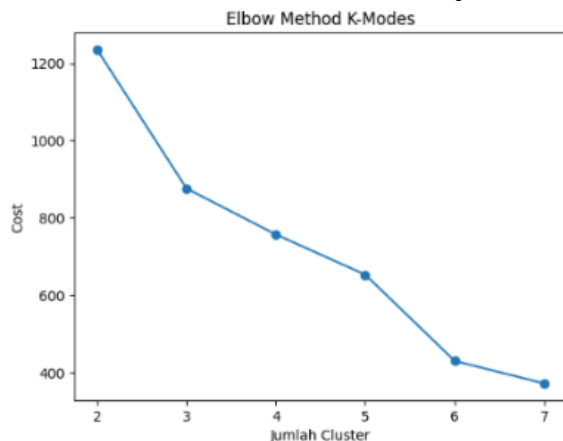
	Order ID	Product Category	Payment Method	price_range	Order Substatus
0	580206442277209236	Sarung	Kartu kredit/debit	Tinggi	Dibatalkan
1	5802064637568148	Sarung	Kartu kredit/debit	Tinggi	Selesai
2	580211456736659114	Sarung	Bayar di tempat	Sangat Rendah	Selesai
3	580213544508753838	Sarung	Bayar di tempat	Rendah	Selesai
4	580214253675054505	Sarung	Transfer bank	Sangat Rendah	Dibatalkan

Gambar 2. Transformasi data

Gambar 2 adalah variabel *price_range* yang bersifat numerik diubah menjadi kategorikal dengan membagi nilainya ke dalam empat rentang harga menggunakan metode *quantile-based discretization*, yaitu "Sangat Rendah", "Rendah", "Sedang", dan "Tinggi". Transformasi ini dilakukan agar seluruh variabel memiliki tipe kategorikal yang sesuai dengan karakteristik algoritma K-Modes.

Proses agregasi data dilakukan dengan mengelompokkan transaksi berdasarkan Order ID, kemudian mengambil modus dari setiap variabel kategorikal untuk merepresentasikan karakteristik pelanggan. Modus dipilih sebagai representasi karena mencerminkan kategori yang paling sering muncul pada setiap pelanggan, sehingga pola pembelian dapat digambarkan secara lebih konsisten. Hasil agregasi menghasilkan dataset pelanggan dengan empat variabel kategorikal utama, yaitu Product Category, Payment Method, price_range, dan Order Substatus, di mana variabel Order Substatus dijadikan sebagai label keputusan dan tidak diikutsertakan dalam proses segmentasi.

4.2. Hasil Penentuan Jumlah Kluster Optimal



Gambar 3. Grafik elbow method

Gambar 3 menunjukkan penentuan jumlah kluster optimal yang dilakukan menggunakan metode *Elbow Method* dengan menghitung nilai *cost function* dari algoritma K-Modes untuk rentang jumlah kluster (k) dari 2 hingga 7. Nilai *cost* menunjukkan total ketidaksamaan (*dissimilarity*) antara data dengan pusat kluster yang terbentuk. Grafik *Elbow Method* menunjukkan penurunan nilai *cost* yang signifikan hingga $k = 6$, setelah itu penurunan cenderung melambat dan membentuk pola siku (*elbow*), yang mengindikasikan

bahwa $k = 6$ merupakan jumlah kluster yang optimal untuk dataset ini.

Pemilihan $k = 6$ juga didukung oleh pertimbangan interpretasi bisnis, di mana jumlah kluster yang terlalu sedikit akan menghasilkan segmentasi yang terlalu umum, sedangkan jumlah kluster yang terlalu banyak akan menyulitkan dalam penyusunan strategi pemasaran yang spesifik. Dengan enam kluster, segmentasi pelanggan dapat dilakukan secara lebih terinci namun tetap dapat diinterpretasikan dengan baik untuk kepentingan pengambilan keputusan strategis.

4.3. Hasil Segmentasi Pelanggan

Implementasi algoritma K-Modes dengan parameter $n_clusters = 6$, $init = 'Huang'$, dan $n_init = 10$ menghasilkan 6 kluster pelanggan dengan distribusi yang bervariasi. Hasil segmentasi menunjukkan bahwa kluster yang terbentuk memiliki karakteristik yang berbeda berdasarkan pola pembelian, metode pembayaran, dan rentang harga transaksi. Tabel 1 berikut menyajikan ringkasan distribusi pelanggan pada setiap kluster:

Tabel 1. Distribusi pelanggan tiap kluster

Kluster	Pelanggan	Persentase	Interpretasi
0	362	25,00%	High-Value Selective Buyer
1	307	21,20%	Regular Mid-Value Buyer
2	166	11,46%	Occasional Moderate Buyer
3	250	17,27%	Price-Sensitive Economy Buyer
4	305	21,06%	Budget-Conscious Value Seeker
5	58	4,01%	Low-Frequency Planned Buyer
Total	1.448	100,00%	-

4.4. Hasil Profil Karakteristik

Tabel 2 dibawah ini menyajikan ringkasan karakteristik utama setiap kluster berdasarkan empat dimensi penting, yaitu rentang harga dominan, metode pembayaran utama, kategori produk yang dibeli, dan tingkat penyelesaian pesanan. Analisis terhadap tabel ini menunjukkan pola yang jelas dalam segmentasi pelanggan, di mana setiap kluster memiliki kombinasi unik antara daya beli, preferensi pembayaran, dan perilaku pembelian. Rentang harga merupakan dimensi paling membedakan antar kluster, dengan variasi dari tinggi pada kluster 0 hingga sangat rendah pada kluster 4. Metode pembayaran menunjukkan polarisasi antara pembayaran tunai (bayar di tempat) pada kluster 0, 1, 3, dan 4, serta pembayaran digital (transfer bank) pada kluster 2 dan 5. Tingkat penyelesaian pesanan yang tinggi pada semua kluster (berkisar 77,07% hingga 87,93%) mengindikasikan bahwa algoritma K-Modes berhasil mengidentifikasi segmen-segmen dengan perilaku pembelian yang

konsisten dan dapat diprediksi. Penjelasan detail mengenai karakteristik setiap klaster disajikan pada bagian berikutnya

Tabel 2. Karakteristik utama tiap klaster

Klaster	Range Harga Dominan	Metode Pembayaran Utama	Kategori Produk	Status Selesai (%)
0	Tinggi (100%)	Bayar di tempat (59,94%), Transfer bank (20,44%)	Sarung	77,07%
1	Sedang (100%)	Bayar di tempat (80,78%)	Sarung	83,71%
2	Rendah (69,28%), Sangat Rendah (30,72%)	Transfer bank (59,64%)	Sarung, Peci	86,14%
3	Rendah (100%)	Bayar di tempat (100%)	Sarung	84,00%
4	Sangat Rendah (100%)	Bayar di tempat (82,62%)	Sarung, Kit Parfum, Peci	87,54%
5	Sedang (100%)	Transfer bank (100%)	Sarung	87,93%

4.5. Klaster 0: *High-Value Selective Buyer*

Klaster 0 terdiri atas 362 transaksi dengan karakteristik utama adalah rentang harga tinggi pada seluruh transaksi (100%). Kategori produk yang dominan adalah sarung, dengan metode pembayaran yang paling banyak digunakan adalah bayar di tempat (217 transaksi atau 59,94%) dan transfer bank (74 transaksi atau 20,44%). Selain itu, sebagian pelanggan juga menggunakan metode *PayLater* (35 transaksi atau 9,67%) dan QRIS (16 transaksi atau 4,42%). Dari segi status pesanan, sebagian besar transaksi pada klaster ini telah selesai (279 transaksi atau 77,07%), namun terdapat juga transaksi yang dibatalkan sebanyak 83 transaksi (22,93%). Klaster ini merepresentasikan pelanggan dengan daya beli tinggi yang cenderung melakukan transaksi dengan nominal besar, sehingga menjadi segmen yang sangat bernilai bagi toko.

4.6. Klaster 1: *Regular Mid-Value Buyer*

Klaster 1 memiliki 307 transaksi dengan rentang harga sedang (100%). Metode pembayaran yang dominan adalah bayar di tempat (248 transaksi atau 80,78%), diikuti oleh *PayLater* (33 transaksi atau 10,75%) dan QRIS (12 transaksi atau 3,91%). Semua transaksi pada klaster ini adalah produk kategori sarung. Status pesanan pada klaster ini menunjukkan bahwa mayoritas transaksi telah selesai (257 transaksi atau 83,71%), dengan tingkat pembatalan yang relatif rendah (49 transaksi atau 15,96%). Klaster ini menggambarkan pelanggan dengan daya beli menengah yang cenderung memilih metode pembayaran tunai dan memiliki tingkat konversi yang cukup baik

4.7. Klaster 2: *Occasional Moderate Buyer*

Klaster 2 terdiri atas 166 transaksi dengan rentang harga yang bervariasi, yaitu sangat rendah (51 transaksi atau 30,72%) dan rendah (115 transaksi atau 69,28%). Metode pembayaran yang paling banyak digunakan adalah transfer bank (99 transaksi atau 59,64%), yang mengindikasikan preferensi terhadap metode pembayaran non-tunai. Selain itu, terdapat juga penggunaan *PayLater* (33 transaksi atau 19,88%), QRIS (14 transaksi atau 8,43%), dan DANA

(9 transaksi atau 5,42%). Kategori produk yang dibeli mayoritas adalah sarung (161 transaksi atau 96,99%), dengan sebagian kecil adalah peci (5 transaksi atau 3,01%). Status pesanan menunjukkan bahwa sebagian besar transaksi telah selesai (143 transaksi atau 86,14%), dengan tingkat pembatalan sebesar 13,86% (23 transaksi). Klaster ini merepresentasikan pelanggan yang memiliki preferensi kuat terhadap pembayaran digital, meskipun dengan rentang harga yang lebih rendah.

4.8. Klaster 3: *Price-Sensitive Economy Buyer*

Klaster 3 memiliki 250 transaksi dengan rentang harga rendah (100%). Karakteristik utama dari klaster ini adalah penggunaan metode pembayaran bayar di tempat secara eksklusif (100%). Semua transaksi pada klaster ini adalah produk kategori sarung, dengan status pesanan yang telah selesai sebanyak 210 transaksi (84,00%) dan dibatalkan sebanyak 40 transaksi (16,00%). Klaster ini menggambarkan pelanggan dengan daya beli ekonomis yang sangat bergantung pada metode pembayaran tunai, kemungkinan karena keterbatasan akses terhadap metode pembayaran digital atau preferensi personal.

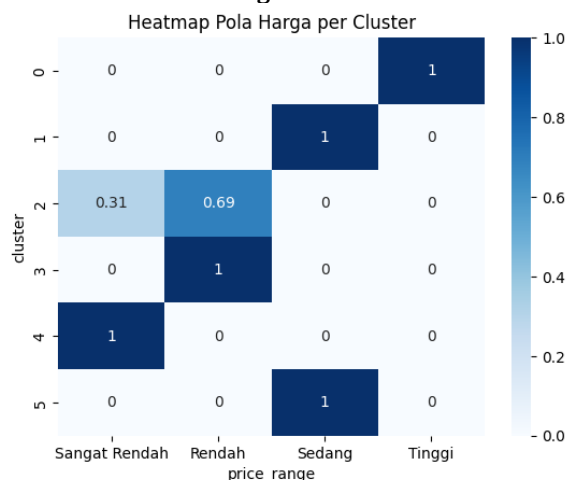
4.9. Klaster 4: *Budget-Conscious Value Seeker*

Klaster 4 terdiri atas 305 transaksi dengan rentang harga sangat rendah (100%). Metode pembayaran yang dominan adalah bayar di tempat (252 transaksi atau 82,62%), diikuti oleh *PayLater* (16 transaksi atau 5,25%), DANA (14 transaksi atau 4,59%), dan QRIS (12 transaksi atau 3,93%). Kategori produk pada klaster ini lebih beragam dibandingkan klaster lainnya, dengan mayoritas adalah sarung (277 transaksi atau 90,82%), kit parfum (16 transaksi atau 5,25%), dan peci (12 transaksi atau 3,93%). Status pesanan menunjukkan tingkat penyelesaian yang tinggi (267 transaksi atau 87,54%), dengan tingkat pembatalan sebesar 12,46% (38 transaksi). Klaster ini merepresentasikan pelanggan dengan daya beli sangat terbatas namun memiliki variasi produk yang lebih luas, sehingga memerlukan pendekatan promosi yang berfokus pada harga kompetitif dan bundling produk

4.10. Klaster 5: Low-Frequency Planned Buyer

Klaster 5 merupakan segmen terkecil dengan 58 transaksi, yang memiliki karakteristik unik berupa penggunaan metode transfer bank secara eksklusif (100%). Rentang harga pada klaster ini adalah sedang (100%), dengan kategori produk yang seluruhnya adalah sarung. Status pesanan menunjukkan bahwa mayoritas transaksi telah selesai (51 transaksi atau 87,93%), dengan tingkat pembatalan sebesar 12,07% (7 transaksi). Klaster ini mencerminkan pelanggan yang memiliki preferensi kuat terhadap metode pembayaran transfer bank, kemungkinan karena alasan keamanan, kebiasaan, atau keterbatasan akses terhadap metode pembayaran digital lainnya.

4.11. Analisis Pola Harga Per Klaster



Gambar 4. Heatmap pola harga per klaster

Visualisasi *heatmap* pola harga per klaster pada Gambar 4 menunjukkan distribusi proporsi pelanggan berdasarkan rentang harga pada setiap klaster. Hasil visualisasi menunjukkan bahwa klaster 0 memiliki konsentrasi 100% pada rentang harga tinggi, klaster 1 pada rentang harga sedang, klaster 3 pada rentang harga rendah, dan klaster 4 pada rentang harga sangat rendah. Sementara itu, klaster 2 menunjukkan distribusi campuran antara rentang harga sangat rendah dan rendah, sedangkan klaster 5 memiliki konsentrasi pada rentang harga sedang. Pola ini mengkonfirmasi bahwa algoritma K-Modes berhasil mengelompokkan pelanggan berdasarkan rentang harga sebagai salah satu dimensi utama dalam segmentasi.

4.12. Hubungan Klaster dengan Status Pesan

Gambar 5 yaitu Analisis hubungan antara klaster dengan status pesanan menunjukkan bahwa sebagian besar klaster memiliki tingkat penyelesaian pesanan yang tinggi, dengan proporsi status "Selesai" berkisar antara 77,07% hingga 87,93%. Klaster 0 memiliki tingkat pembatalan tertinggi (22,93%), yang kemungkinan disebabkan oleh nominal transaksi yang tinggi sehingga pelanggan lebih selektif atau

mengalami kendala dalam proses pembayaran. Sebaliknya, klaster 4 memiliki tingkat penyelesaian tertinggi (87,54%), yang mengindikasikan bahwa pelanggan dengan daya beli rendah cenderung lebih komitmen terhadap transaksi yang telah dilakukan. Temuan ini memberikan wawasan penting bagi toko untuk mengidentifikasi klaster yang memerlukan perhatian khusus dalam hal follow-up pesanan dan strategi retensi pelanggan.



Gambar 5. Stacked bar chart

5. KESIMPULAN DAN SARAN

Hasil penelitian menunjukkan bahwa penerapan algoritma K-Modes pada data transaksi Toko XYZ mampu menghasilkan enam segmen pelanggan yang memiliki karakteristik perilaku pembelian yang berbeda dan dapat diinterpretasikan secara operasional. Perbedaan antar klaster terutama dipengaruhi oleh variasi rentang harga transaksi dan metode pembayaran yang digunakan, sehingga menegaskan bahwa kedua variabel tersebut merupakan dimensi utama dalam membedakan pola pembelian pelanggan. Temuan ini mengindikasikan bahwa algoritma K-Modes efektif digunakan sebagai pendekatan segmentasi pelanggan berbasis data kategorikal dalam konteks ritel. Berdasarkan hasil segmentasi tersebut, pengelola Toko XYZ disarankan untuk menerapkan strategi pemasaran yang lebih tersegmentasi, seperti pengembangan program loyalitas dan layanan bernilai tambah bagi pelanggan bernilai tinggi, optimalisasi promosi dan bundling produk untuk pelanggan dengan nilai transaksi menengah, serta penawaran harga kompetitif dan kemudahan pembayaran bagi segmen pelanggan yang sensitif terhadap harga. Untuk penelitian selanjutnya, disarankan agar analisis diperluas dengan menambahkan variabel perilaku lain, seperti frekuensi pembelian dan durasi hubungan pelanggan, serta melakukan perbandingan dengan algoritma segmentasi lain yang mendukung data kategorikal guna meningkatkan ketahanan dan generalisasi hasil segmentasi.

DAFTAR PUSTAKA

- [1] S. Nurjanah and B. Santoso, "Analisis implementasi teknologi digital dalam manajemen bisnis ritel: Studi literatur," *Musytari: Jurnal Manajemen dan Akuntansi*, vol. 7, no. 3, pp. 510–522, 2024, doi: 10.35897/musytari.v7i3.7752.
- [2] Iskandar, Muhsidi, Q. E. Hasan, D. D. Muttaqien, and A. Sumiah, "Analisis perilaku pembelian dan prediksi konsumen pada usaha retail dengan metode data mining," *Jurnal Tekno Kompak*, vol. 19, no. 2, pp. 78–87, 2025, doi: 10.33365/jtk.v19i2.728.
- [3] E. K. Nduru, E. Buulolo, and P. Pristiwanto, "Implementasi algoritma K-Modes untuk menentukan strategi marketing STMIK Budi Darma," in *Konferensi Nasional Teknologi Informasi dan Komputer (KOMIK)*, 2018.
- [4] N. Awaliyah, E. Prasetyo, E. Muzayanah, and A. Lestari, "Optimizing customer segmentation in online retail using RFM and clustering methods," *Jurnal Innovative*, vol. 14, no. 2, pp. 123–131, 2024.
- [5] P. M. K. Mado and 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, doi: 10.63447/jimik.v6i3.1563.
- [6] B. T. Kristanti, A. Junaidi, and 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, pp. 1–10, 2024, doi: 10.23960/jitet.v12i3.4677.
- [7] T. D. Dinh *et al.*, "Categorical Data Clustering: 25 Years Beyond K-Modes," *ArXiv*, 2024, [Online]. Available: <https://arxiv.org/abs/2408.17244>
- [8] B. Thapaliya and Y. Zhuang, "Fast Clustering of Categorical Big Data," *ArXiv*, 2025, [Online]. Available: <https://arxiv.org/abs/2502.07081>
- [9] S. R. Niloy, T. H. Muktaashid, and A. Saiduzzaman, "Customer Segmentation and Classification Using K-Modes Clustering with Ensemble Learning," in *Communications in Computer and Information Science*, Springer, 2025, pp. 3–18.
- [10] J. Herina, P. 1✉, and J. Heikal, "Customer Segmentation Using Python K-Modes Clustering Method (Case Study on Netflix Userbase)".
- [11] M. A. Yaqin, N. Ayunda, S. Sujarwo, and M. A. Murtadho, "Segmentasi pelanggan agen beauty local brand menggunakan algoritma K-Modes berbasis Python," *Nusantara Journal of Computers and Its Applications*, vol. 7, no. 1, pp. 1–7, 2022.
- [12] A. Huang, Z. Huang, and M. K. Ng, "Title A fuzzy k-modes algorithm for clustering categorical data A Fuzzy k-Modes Algorithm for Clustering Categorical Data," 1999. [Online]. Available: <http://hdl.handle.net/10722/42992>
- [13] R. J. Kuo, M. Cendana, T. P. Q. Nguyen, and F. E. Zulvia, "A Multivariate Fuzzy Weighted K-Modes Algorithm with Probabilistic Distance for Categorical Data," *Journal of ICT Research and Applications*, vol. 18, no. 2, pp. 93–107, Sep. 2024, doi: 10.5614/ITBJ.ICT.RES.APPL.2023.18.2.2.
- [14] A. Y. N. Rumapea, F. A. Ryandi, D. Pratiwi, and S. Sari, "Analisis segmentasi pelanggan ritel online menggunakan K-Means clustering berdasarkan model Recency, Frequency, Monetary (RFM)," *Jurnal Sains dan Teknologi*, vol. 7, no. 1, pp. 45–57, 2024, [Online]. Available: <https://ejournal.sisfokomtek.org/index.php/saintek/article/view/4607>
- [15] A. G. Ramadhan, S. Yuliana, and F. Aziz, "Data mining untuk segmentasi pelanggan dengan algoritma K-Means: Studi kasus pada data pelanggan di toko retail," *Jurnal Syntax Literate*, vol. 8, no. 7, pp. 4345–4358, 2023, doi: 10.36418/syntax-literate.v8i7.13746.
- [16] S. P. L. Geni, H. Safitri, F. Merry, M. Wati, and H. Haviluddin, "Penerapan K-Means clustering untuk segmentasi konsumen e-commerce berdasarkan pola pembelian," *JUKI: Jurnal Komputer dan Informatika*, vol. 7, no. 2, pp. 450–459, 2025, doi: 10.31328/juki.v7i2.1049.
- [17] J. H. Putri, R. Kurniawan, and A. Sari, "Customer segmentation using Python K-Modes clustering method for Netflix persona analysis," *Jurnal Innovative*, vol. 14, no. 3, pp. 210–222, 2024, [Online]. Available: <https://j-innovative.org/index.php/Innovative/article/view/14883>
- [18] W. N. Fadhilah, "Penerapan algoritma binning pada preprocessing data klasifikasi SDGs," *INFOMEDIA: Jurnal Ilmiah Program Studi Teknik Informatika*, vol. 10, no. 1, pp. 1–8, 2025.
- [19] D. Rifaldi, "Teknik preprocessing pada text mining menggunakan machine learning," *Jurnal Teknologi Informasi*, vol. 6, no. 1, pp. 45–52, 2023