

# PENERAPAN DATA MINING UNTUK SEGMENTASI MENU KOPI BERDASARKAN KARAKTERISTIK PEMINAT MENGGUNAKAN ALGORITMA K-MEANS

Rizki Habibah\*, Hikmah Aldinar Siregar, Ade Lestari Hasibuan, Yolanda Listia, Aldi Rahmansyah, M. Idris Sagala

<sup>1</sup>Program Studi Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Labuhanbatu  
e-mail: rizkihabibah151@gmail.com\*, hikmahaldinar13@gmail.com, adelestari.id@gmail.com,  
yolaalistia@gmail.com, ramansyahaldi35@gmail.com, IdrisSagala.id@gmail.com

Received: 2025-12-30 | Accepted: 2026-01-08 | Published: 2026-03-19

## ABSTRAK

Pertumbuhan variasi menu kopi menuntut pelaku usaha memahami karakteristik peminat secara lebih terstruktur, sementara data terkait menu dan preferensi konsumen sering belum dimanfaatkan secara optimal dalam proses pengambilan keputusan. Penelitian ini bertujuan melakukan segmentasi menu kopi berdasarkan karakteristik peminat dengan pendekatan data mining. Metode yang digunakan adalah clustering menggunakan algoritma K-means yang diimplementasikan melalui perangkat lunak Orange Data Mining dengan atribut utama berupa harga dan kategori peminat. Proses analisis meliputi tahap prapemrosesan data, penentuan jumlah cluster, serta evaluasi kualitas cluster menggunakan silhouette coefficient. Hasil penelitian menunjukkan bahwa algoritma K-means mampu mengelompokkan menu kopi ke dalam tiga cluster dengan karakteristik harga dan tingkat minat konsumen yang berbeda. Evaluasi menghasilkan nilai silhouette coefficient sebesar 0,725 yang menunjukkan struktur cluster yang kuat dengan pemisahan yang jelas antar kelompok. Visualisasi hasil klasterisasi memperlihatkan tiga segmen utama, yaitu cluster ekonomis dengan harga rendah dan minat tinggi, cluster menengah dengan harga sedang dan minat yang bervariasi, serta cluster premium dengan harga tinggi dan minat konsumen yang tetap kuat. Segmentasi tersebut memberikan gambaran pola preferensi konsumen yang lebih terstruktur sehingga dapat digunakan sebagai dasar dalam perencanaan produk dan penetapan strategi harga pada usaha kopi.

**Kata Kunci:** data mining, segmentasi, menu kopi, K-Means, clustering

## ABSTRACT

The growth of coffee menu variations requires business owners to understand consumer interest characteristics in a structured manner, while menu data and consumer preference information are often not fully utilized in decision making. This study aims to segment coffee menus based on consumer-interest characteristics using a data-mining approach. The method applied is clustering using the K-Means algorithm implemented in the Orange Data Mining software, with two main attributes: price and interest category. The analysis process includes data preprocessing, determining the optimal number of clusters, and evaluating cluster quality using the silhouette coefficient.

The results show that the K-Means algorithm successfully groups coffee menus into three clusters with distinct price ranges and consumer-interest characteristics. The evaluation yields a silhouette coefficient of 0.725, indicating a strong cluster structure with clear separation between groups. Visualization of the clustering results reveals three main segments: an economical cluster characterized by low prices and high consumer interest, a middle cluster with moderate prices and varying levels of interest, and a premium cluster with high prices and consistently strong consumer interest. These segmentation results provide a clear representation of consumer preference patterns and support decision making in product planning and pricing strategies for coffee businesses.

**Keywords:** data mining, segmentation, coffee menu, K-Means, clustering

## 1. Pendahuluan

Perkembangan industri kopi di Indonesia mengalami peningkatan yang signifikan dalam beberapa tahun terakhir, termasuk di wilayah Rantauprapat sebagai pusat pertumbuhan ekonomi Kabupaten Labuhanbatu. Peningkatan ini ditandai dengan munculnya berbagai jenis usaha kopi, mulai dari kopi

keliling dengan harga ekonomis hingga kafe modern yang menawarkan menu premium. Keberagaman tersebut menciptakan pasar kopi yang dinamis, namun sekaligus menimbulkan tantangan bagi pelaku usaha dalam memahami karakteristik dan preferensi peminat secara akurat [4].

Perbedaan daya beli, frekuensi konsumsi, serta preferensi menu menyebabkan perilaku konsumen kopi menjadi semakin kompleks. Strategi pemasaran dan pengembangan produk yang bersifat umum tidak lagi efektif untuk diterapkan pada seluruh segmen pasar. Oleh karena itu, diperlukan pendekatan berbasis data untuk mengidentifikasi kelompok peminat kopi yang memiliki karakteristik serupa sehingga strategi bisnis dapat disusun secara lebih tepat sasaran [1].

Pendekatan data mining telah banyak digunakan untuk menganalisis perilaku konsumen dan menemukan pola tersembunyi dalam data berskala besar. Data mining memungkinkan proses pengolahan data secara sistematis melalui tahapan seleksi data, pra-pemrosesan, pemodelan, hingga evaluasi hasil [1]. Salah satu Teknik utama dalam data mining adalah clustering, yaitu metode pengelompokan data berdasarkan tingkat kemiripan antar objek tanpa memerlukan label kelas sebelumnya [3].

Algoritma K-Means merupakan metode clustering yang paling banyak digunakan karena kesederhanaannya, efisiensi komputasi, serta kemampuannya dalam membentuk kelompok data yang jelas dan mudah diinterpretasikan. Berbagai penelitian sebelumnya menunjukkan bahwa algoritma K-Means efektif digunakan dalam segmentasi pelanggan pada sektor ritel dan industri makanan dan minuman, termasuk pada bisnis kedai kopi [2], [6], [15]. K-Means juga mampu membantu pelaku usaha dalam memahami pola konsumsi pelanggan berdasarkan variabel perilaku seperti frekuensi pembelian dan tingkat pengeluaran [7].

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk menerapkan metode data mining menggunakan algoritma K-Means dalam melakukan segmentasi menu kopi berdasarkan karakteristik peminat di Rantauprapat. Segmentasi dilakukan dengan memanfaatkan atribut frekuensi konsumsi dan rata-rata pengeluaran sebagai dasar pengelompokan. Hasil segmentasi diharapkan dapat memberikan gambaran yang jelas mengenai pola minat konsumen terhadap menu kopi serta menjadi dasar pendukung dalam penentuan strategi produk dan harga beli pelaku usaha kopi lokal.

## 2. Analisa dan pembahasan

Analisis data dilakukan dengan menerapkan algoritma K Means untuk mengelompokkan menu kopi berdasarkan harga dan tingkat minat konsumen. Proses ini bertujuan mengidentifikasi pola segmentasi produk yang dapat digunakan sebagai dasar penyusunan strategi pemasaran pada usaha kopi.

Data penelitian diperoleh melalui observasi terhadap pelaku usaha kopi lokal dengan mencatat atribut nama menu, harga, dan karakteristik peminat. Dataset disusun dalam format tabulasi menggunakan Microsoft Excel kemudian dikonversi ke format CSV agar dapat diproses menggunakan algoritma K Means pada tahap analisis data.

Tahap prapemrosesan meliputi pembersihan data dan transformasi atribut. Atribut kategorikal pada karakteristik peminat diubah ke dalam bentuk numerik agar dapat diproses secara komputasional. Selanjutnya dilakukan normalisasi data untuk menyamakan skala antar fitur sehingga proses pengelompokan tidak dipengaruhi oleh perbedaan rentang nilai antar variabel.

Dalam penelitian ini atribut harga dan karakteristik peminat digunakan sebagai variabel utama analisis. Variabel harga dinyatakan sebagai X1 dalam satuan ribuan rupiah, sedangkan kategori peminat dinyatakan sebagai X2 yang dikonversi ke skala numerik dengan nilai 1 untuk tingkat minat rendah, 2 untuk tingkat minat sedang, dan 3 untuk tingkat minat tinggi. Atribut nama menu digunakan sebagai metadata untuk mengidentifikasi setiap objek data. Dataset penelitian terdiri atas sepuluh menu kopi dengan dua variabel utama, yaitu harga dan kategori peminat. Struktur dataset yang digunakan dalam proses pengelompokan ditunjukkan pada Tabel 1.

Tabel 1 Data hasil konversi atribut harga dan peminat

ID	Nama Menu	(X1)	(X2)
1	Espresso	15	3
2	Americano	18	2
3	Café Latte	25	3
4	Cappuccino	24	3
5	Kopi Susu Aren	12	3
6	Macchiato	22	2
7	Long Black	18	1
8	Mochaccino	26	2
9	Cold Brew	20	2
10	Affogato	28	3

Penentuan cluster menggunakan  $K = 3$ . Berdasarkan iterasi akhir, didapatkan posisi Centroid (C) sebagai berikut :

- C1 (Ekonomis): (13,5 3.0)
- C2 (Medium): (19.5, 1.5)
- C3 (Premium): (25.5, 2.6)

Jarak dihitung menggunakan Euclidean Distance :

$$D = \sqrt{(X_1 - C_1)^2 + (X_2 - C_2)^2}$$

Tabel 2 Hasil Jarak dan Penentuan Cluster

ID	Jarak ke C1	Jarak ke C2	Jarak ke C3	Cluster	$D_{min}^2$
1	1.50	4.74	10.51	C1	2.25
2	4.61	1.58	7.52	C2	2.50
3	11.50	5.70	0.64	C3	0.41
4	10.50	4.74	1.55	C3	2.41
5	1.50	7.65	13.51	C1	2.25
6	8.56	2.55	3.55	C2	6.50
7	4.92	1.58	7.67	C2	2.50
8	12.54	6.52	0.78	C3	0.61
9	6.58	0.71	5.53	C2	0.50
10	14.50	8.63	2.53	C3	6.41

Rata-rata 26.34

Mengacu pada efektivitas klasterisasi diukur dengan membandingkan variansi di dalam dan antar kelompok.

- WCV: Total kedekatan data di dalam satu kelompok

$$WCV = \sum (D_{min})^2$$

$$WCV = 2.25 + 2.50 + 0.41 + 2.41 + 2.25 + 6.50 + 2.50 + 0.61 + 0.50 + 6.41 = 26.34$$

- BCV: Jarak antar pusat (centroid) cluster.

$$BCV = (C2 - C1)^2 + (C3 - C1)^2 + (C3 - C2)^2$$

$$\text{Dist}(C1, C2)^2 = (19.5 - 13.5)^2 + (1.5 - 3.0)^2 = 6^2 + (-1.5)^2 = 36 + 2.25 = 38.25$$

$$\text{Dist}(C1, C3)^2 = (25.5 - 13.5)^2 + (2.6 - 3.0)^2 = 12^2 + (-0.4)^2 = 144 + 0.16 = 144.16$$

$$\text{Dist}(C2, C3)^2 = (25.5 - 19.5)^2 + (2.6 - 1.5)^2 = 6^2 + 1.1^2 = 36 + 1.21 = 37.21$$

$$= 38.25 + 144.16 + 37.21$$

$$= 219.62$$

- Rasio (R): Semakin tinggi nilai rasio, semakin baik pemisahan datanya.

$$R = \frac{BCV}{WCV}$$

$$\text{Rasio} = \frac{BCV}{WCV} = \frac{219.62}{26.34} = 8.337$$

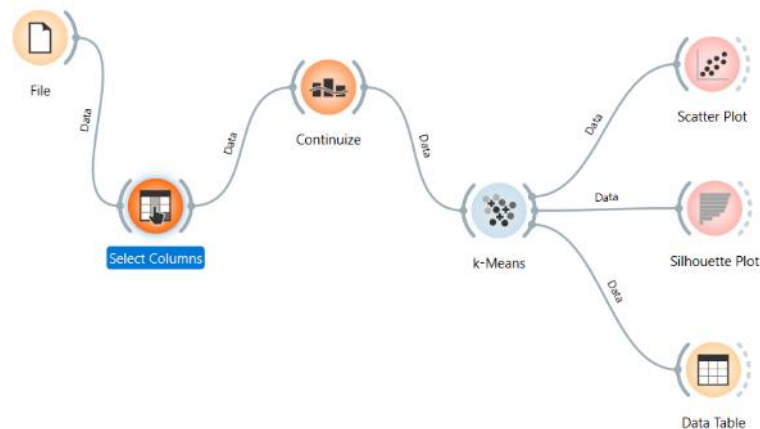
Tabel 3 Perbandingan Evaluasi

Jumlah Cluster (K)	Total WCV	Total BCV	Rasio (BCV/WCV)	Hasil
K = 2	68.45	82.50	1.205	Kurang Optimal
K = 3	26.34	219.62	8.337	Optimal
K = 4	22.10	231.40	10.470	Overfit

Berdasarkan hasil perhitungan di atas, nilai Rasio 8.337 pada  $K = 3$ , menunjukkan bahwa perbedaan antar kelompok (BCV) jauh lebih dominan dibandingkan penyebaran data di dalam kelompok (WCV). Secara fisik, ini berarti menu kopi telah terbagi menjadi tiga zona harga dan minat yang jelas:

1. Cluster 1 (Low Price, High Demand) : Fokus pada volume penjualan tinggi
2. Cluster 2 (Medium Price, Low-Mid Demand) : Perlu promosi tambahan
3. Cluster 3 (High Price, High Demand) : Produk unggulan dengan profit margin besar

Alur kerja penelitian diimplementasikan menggunakan perangkat lunak orange data mining.



Gambar 1. Workflow Pemrosesan Data K-Means pada Orange

Penjelasan komponen workflow :

1. Widget File  
Digunakan untuk mengimpor dataset dalam format CSV ke dalam lingkungan analisis pada perangkat lunak Orange Data Mining. Pada tahap ini dilakukan pengaturan peran atribut, yaitu nama menu sebagai meta attribute yang berfungsi sebagai identitas data, sedangkan atribut harga dan tingkat minat konsumen ditetapkan sebagai fitur utama yang digunakan dalam proses analisis kluster.
2. Select Columns  
Tahap ini digunakan untuk menyeleksi variabel yang relevan dengan tujuan analisis. Proses seleksi difokuskan pada dua fitur utama yaitu harga dan minat konsumen agar proses klusterisasi berjalan lebih terarah dan tidak dipengaruhi atribut yang tidak berkontribusi terhadap pembentukan kelompok data.
3. Preprocess Normalize  
Tahap prapemrosesan dilakukan melalui normalisasi data untuk menyamakan skala antar fitur. Proses ini bertujuan mencegah dominasi salah satu variabel dalam perhitungan jarak sehingga kontribusi masing masing fitur tetap proporsional pada proses klusterisasi.
4. K Means  
Algoritma K Means diterapkan untuk melakukan proses pengelompokan data. Parameter jumlah kluster ditetapkan sebesar  $K = 3$  berdasarkan hasil evaluasi sebelumnya. Proses ini mengelompokkan menu kopi berdasarkan tingkat kemiripan karakteristik harga dan minat konsumen.
5. Silhouette Plot  
Widget ini digunakan untuk mengevaluasi kualitas kluster yang terbentuk dengan mengukur tingkat kohesi dan separasi antar data. Nilai silhouette memberikan indikasi seberapa dekat suatu objek dengan cluster tempatnya berada dibandingkan dengan cluster lain.
6. Scatter Plot  
Visualisasi scatter plot digunakan untuk menampilkan distribusi data dalam ruang dua dimensi berdasarkan variabel harga dan minat konsumen. Setiap titik data diberi label warna sesuai dengan cluster hasil proses klusterisasi sehingga pola pengelompokan dapat diamati secara visual.

7. Data Table

Widget data table digunakan untuk menampilkan dan mendokumentasikan hasil pelabelan cluster pada setiap baris data. Informasi ini digunakan sebagai dasar interpretasi hasil analisis serta mendukung proses pelaporan dalam penulisan hasil penelitian.

Analisis data dilakukan menggunakan algoritma K Means untuk mengelompokkan menu kopi ke dalam cluster yang optimal berdasarkan kesamaan karakteristik data. Proses pengelompokan bertujuan mengidentifikasi pola segmentasi menu yang memiliki kemiripan pada atribut harga dan tingkat minat konsumen. Evaluasi kualitas cluster dilakukan melalui dua pendekatan analisis, yaitu sebagai berikut.

1. Analisis Varian

Metode ini digunakan untuk mengukur kekuatan struktur cluster yang terbentuk melalui perbandingan antara variasi data di dalam cluster dan variasi data antar cluster. Parameter yang digunakan meliputi Within Cluster Variance WCV dan Between Cluster Variance BCV. Nilai rasio antara BCV dan WCV digunakan untuk menilai tingkat pemisahan cluster. Semakin tinggi nilai rasio tersebut, semakin jelas pemisahan antar kelompok data.

2. Silhouette Coefficient

Metode ini digunakan untuk mengukur tingkat kesesuaian setiap objek data terhadap cluster tempatnya berada. Nilai silhouette diperoleh dari perbandingan jarak rata-rata objek terhadap anggota cluster yang sama dengan jarak rata-rata objek terhadap cluster terdekat lainnya. Nilai koefisien ini berada pada rentang minus satu hingga satu. Nilai yang mendekati satu menunjukkan bahwa objek berada pada cluster yang tepat, sedangkan nilai yang mendekati nol menunjukkan posisi objek berada pada batas antar cluster. Nilai negatif menunjukkan kemungkinan kesalahan penempatan objek pada cluster.

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

Di mana:

- $a(i)$  : Jarak rata-rata objek ke semua objek lain dalam satu cluster
- $b(i)$  : Jarak rata-rata objek ke objek di cluster terdekat lainnya.

Tabel 4 Data hasil konversi atribut harga dan peminat

ID	Nama Menu	(X1)	(X2)
1	Espresso	15	3
2	Americano	18	2
3	Café Latte	25	3
4	Cappuccino	24	3
5	Kopi Susu Aren	12	3
6	Macchiato	22	2
7	Long Black	18	1
8	Mochaccino	26	2
9	Cold Brew	20	2
10	Affogato	28	3

Tabel 5 Hasil Pengelompokan Dan Nilai Silhouette

ID	Nama Menu	Cluster	Silhouette Score
1	Espresso	C1	0.82
2	Americano	C2	0.65
3	Café Latte	C3	0.75
4	Cappuccino	C3	0.70
5	Kopi Susu Aren	C1	0.85
6	Macchiato	C2	0.58
7	Long Black	C2	0.60
8	Mochaccino	C3	0.78
9	Cold Brew	C2	0.82
10	Affogato	C3	0.81

Nilai rata-rata silhouette sebesar 0,725 menunjukkan bahwa struktur cluster yang terbentuk memiliki kualitas pengelompokan yang kuat. Nilai tersebut mendekati batas atas koefisien silhouette yaitu 1, sehingga menunjukkan tingkat kohesi yang baik di dalam cluster serta pemisahan yang jelas antar cluster. Untuk memverifikasi hasil analisis yang diperoleh dari perangkat lunak Orange Data Mining, dilakukan perhitungan tambahan menggunakan metode analisis varian secara manual. Perhitungan ini melibatkan

evaluasi nilai Within Cluster Variance dan Between Cluster Variance untuk menilai kekuatan pemisahan antar cluster serta konsistensi objek data dalam kelompoknya.

1. Perhitungan WCV dihitung dari total kuadrat jarak ( $D^2$ ) setiap data ke centroid clusternya :  
 $WCV = \sum(D_{min})^2 = 26.34$  (hasil dari tabulasi jarak iterasi terakhir)
2. Perhitungan BCV mengukur seberapa jauh jarak antar pusat cluster:  
 $Dist(C1, C2)^2 = 38.25$   
 $Dist(C1, C3)^2 = 144.16$   
 $Dist(C2, C3)^2 = 37.21$   
 $Total\ BCV = 38.25 + 144.16 + 37.21 = 219.62$
3. Perhitungan Rasio

$$Rasio = \frac{BCV}{WVC} = \frac{219.62}{26.34} = 8.337$$

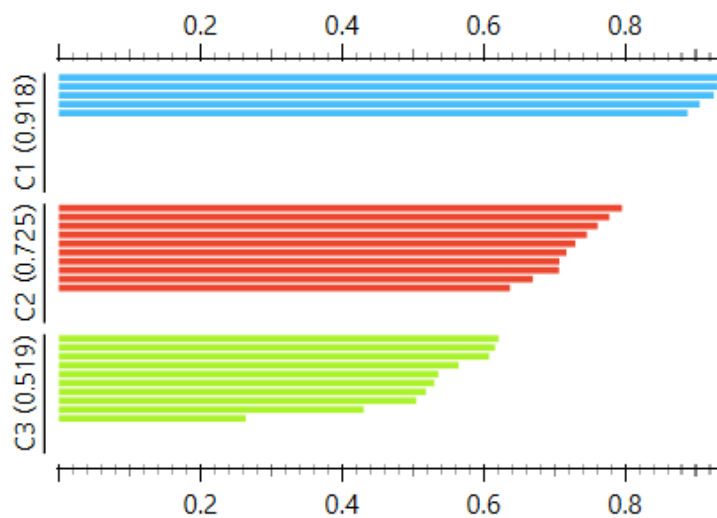
Tabel 6 Tabel Perbandingan Evaluasi

K (Cluster)	Silhouette Score (Orange)	Total WCV	Total BCV	Rasio (BCV/WCV)
K = 2	0.612	68.45	82.50	1.205
K = 3	0.725	26.34	219.62	8.337
K = 4	0.685	22.10	231.40	10.470

Berdasarkan hasil pada tabel evaluasi, meskipun nilai  $K = 4$  menunjukkan rasio pemisahan yang relatif tinggi, nilai silhouette score menurun menjadi 0,685. Penurunan ini menunjukkan berkurangnya kualitas pengelompokan karena sebagian objek data mulai berada pada batas antar cluster sehingga tingkat kohesi dalam kelompok menjadi lebih rendah.

Jumlah cluster yang paling optimal diperoleh pada  $K = 3$  dengan nilai silhouette score sebesar 0,725 serta rasio pemisahan sebesar 8,337. Hasil tersebut menunjukkan bahwa struktur cluster memiliki pemisahan yang jelas serta konsistensi internal yang lebih baik dibandingkan konfigurasi jumlah cluster lainnya.

Selain analisis varian antar cluster, evaluasi kualitas pengelompokan juga dilakukan menggunakan visualisasi silhouette plot melalui widget Silhouette Plot pada perangkat lunak Orange Data Mining. Visualisasi ini digunakan untuk mengamati tingkat kesesuaian setiap objek data terhadap cluster tempatnya berada. Nilai silhouette yang tinggi menunjukkan kedekatan objek dengan cluster asalnya, sedangkan nilai yang mendekati nol mengindikasikan posisi objek berada pada batas antar cluster.



Gambar 2. Evaluasi Clustering dengan Silhouette Plot

Tabel 7 Nilai Silhoutte Score/Cluster

Cluster	Rata-rata Nilai Silhoutte	Keterangan
C1	0.83	Kuat (Strong)
C2	0.68	Sedang (Medium)
C3	0.76	Kuat (Strong)

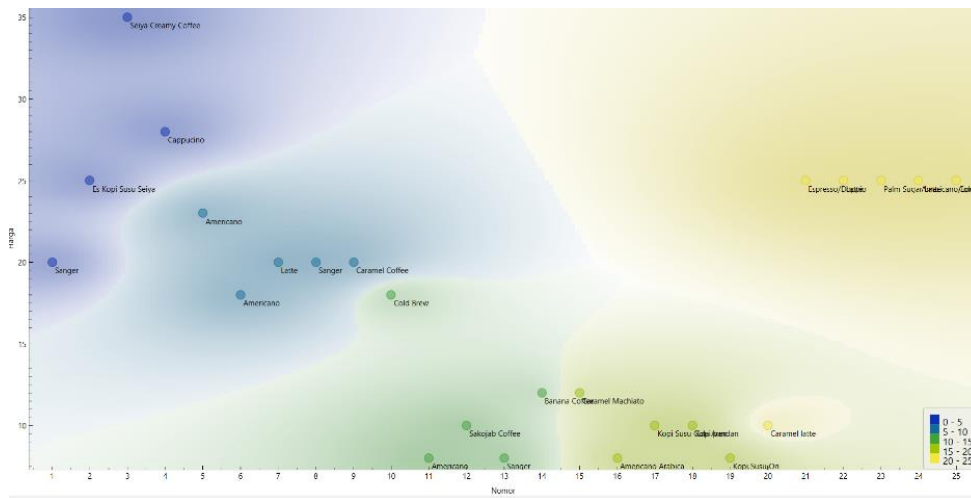
Hasil perhitungan menunjukkan bahwa nilai Rasio dan Silhoutte Score berada pada kategori tinggi. Hal ini membuktikan bahwa :

- Cluster 1 (Ekonomis - Minat Tinggi): Terdiri dari menu harga rendah dengan perputaran cepat
- Cluster 2 (Medium - Minat Sedang): Terdiri dari menu harga menengah yang memerlukan strategi promosi lebih agresif
- Cluster 3 (Premium – Minat Tinggi): Terdiri dari menu eksklusif dengan margin keuntungan besar.

Setelah proses pelabelan cluster menggunakan algoritma K Means selesai dilakukan, distribusi data kemudian divisualisasikan menggunakan widget Scatter Plot pada perangkat lunak Orange Data Mining. Visualisasi ini digunakan untuk menampilkan posisi setiap menu kopi dalam ruang dua dimensi berdasarkan dua variabel utama penelitian.

Sumbu horizontal merepresentasikan variabel harga, sedangkan sumbu vertikal merepresentasikan karakteristik peminat yang telah dikonversi ke dalam bentuk numerik. Setiap titik pada grafik menunjukkan satu menu kopi, dan warna yang berbeda digunakan untuk menandai cluster hasil proses klasterisasi.

Melalui visualisasi ini, pola pengelompokan data dapat diamati secara lebih jelas sehingga memudahkan interpretasi terhadap segmentasi menu berdasarkan tingkat harga dan minat konsumen. Scatter plot juga membantu mengidentifikasi jarak antar cluster serta distribusi objek data dalam setiap kelompok.



Gambar 3. Visualisasi Hasil Clustering menggunakan Scatter Plot

- **Cluster 1**  
 Cluster ini terkonsentrasi pada area kiri atas pada grafik sebaran. Posisi tersebut menunjukkan kelompok menu dengan harga relatif rendah namun memiliki tingkat minat konsumen yang tinggi. Karakteristik ini mengindikasikan produk yang memiliki daya tarik kuat dengan harga yang terjangkau sehingga berpotensi menjadi menu populer.
- **Cluster 2**  
 Cluster ini tersebar pada area tengah hingga bagian bawah grafik. Kelompok ini merepresentasikan menu dengan harga pada kategori menengah dengan tingkat minat konsumen yang bervariasi dari rendah hingga sedang. Pola distribusi yang lebih menyebar menunjukkan variasi preferensi konsumen terhadap menu pada kategori harga ini.
- **Cluster 3**  
 Cluster ini berada pada area kanan atas grafik. Posisi tersebut menunjukkan kelompok menu kopi dengan harga tinggi yang termasuk dalam kategori premium. Meskipun memiliki harga lebih tinggi,

tingkat minat konsumen pada kelompok ini tetap berada pada kategori tinggi sehingga menunjukkan adanya segmen pasar yang memiliki preferensi terhadap produk kopi premium.

### 3. Hasil

Berdasarkan proses iterasi algoritma k-means dengan jumlah  $K = 3$ , berikut adalah hasil akhir pelabelan setiap menu kopi :

ID	Nama Menu	X 1	X 2	Cluster	$D_{min}^2$
1	Espresso	15.000	3	C1	2.25
2	Americano	18.000	2	C2	2.50
3	Café Latte	25.000	3	C3	0.63
4	Cappuccino	24.000	3	C3	3.13
5	Kopi Susu Aren	12.000	3	C1	2.25
6	Macchiato	22.000	2	C2	6.50
7	Long Black	18.000	1	C2	2.50
8	Mochaccino	26.000	2	C3	0.63
9	Cold Brew	20.000	2	C2	0.50
10	Affogato	28.000	3	C3	5.13

Visualisasi menunjukkan bahwa distribusi data terbagi ke dalam tiga zona utama tanpa adanya tumpang tindih antar kelompok. Pola ini menunjukkan pemisahan cluster yang jelas berdasarkan dua variabel utama yaitu harga dan karakteristik peminat konsumen.

#### - Cluster 1 Ekonomis

Cluster ini terdiri dari menu dengan harga relatif rendah namun memiliki tingkat minat konsumen yang tinggi. Karakteristik tersebut menunjukkan potensi produk dengan volume penjualan tinggi karena daya tarik harga yang terjangkau bagi konsumen.

#### - Cluster 2 Menengah

Cluster ini merepresentasikan menu dengan harga pada kategori menengah dengan tingkat minat yang bervariasi dari rendah hingga sedang. Variasi ini menunjukkan dinamika preferensi konsumen pada kategori harga menengah sehingga strategi pemasaran yang relevan adalah peningkatan promosi untuk memperkuat daya tarik produk.

#### - Cluster 3 Premium

Cluster ini berisi menu dengan harga tinggi namun tetap memiliki tingkat minat konsumen yang tinggi. Karakteristik ini menunjukkan adanya segmen pasar yang memiliki preferensi terhadap produk kopi premium. Strategi yang sesuai pada kelompok ini berfokus pada margin keuntungan serta penekanan pada kualitas produk.

Interpretasi visual menunjukkan bahwa titik data yang berada dekat dengan pusat cluster atau centroid memiliki kesamaan karakteristik yang tinggi sehingga menunjukkan tingkat kohesi yang baik di dalam kelompoknya. Jarak antar cluster yang relatif jauh menunjukkan tingkat separasi yang jelas antar kelompok data. Temuan ini konsisten dengan nilai silhouette coefficient yang tinggi serta rasio antara between cluster variance dan within cluster variance yang telah dihitung sebelumnya. Kedua indikator tersebut menunjukkan bahwa konfigurasi tiga cluster merupakan struktur pengelompokan yang optimal untuk dataset menu kopi pada penelitian ini.

### 4. Kesimpulan

Penelitian ini mengimplementasikan algoritma K Means untuk melakukan segmentasi menu kopi berdasarkan harga dan karakteristik peminat dengan pendekatan data mining. Hasil analisis menunjukkan bahwa dataset menu kopi dapat dikelompokkan ke dalam tiga cluster dengan karakteristik yang berbeda. Pembagian tersebut menunjukkan adanya pola segmentasi yang jelas pada preferensi konsumen terhadap produk kopi.

Evaluasi menggunakan metode silhouette coefficient menunjukkan bahwa struktur cluster yang terbentuk memiliki tingkat kohesi dan separasi yang baik. Hasil ini menunjukkan bahwa konfigurasi cluster yang dihasilkan memiliki stabilitas yang memadai sehingga layak digunakan sebagai dasar interpretasi data. Segmentasi yang diperoleh memberikan informasi yang relevan bagi pelaku usaha dalam mendukung pengambilan keputusan terkait perencanaan produk, pengelompokan menu, serta strategi penetapan harga.

Keterbatasan penelitian ini terletak pada jumlah variabel yang digunakan yang masih terbatas serta sifat dataset yang bersifat statis. Kondisi tersebut membatasi kemampuan model dalam menangkap

dinamika preferensi konsumen yang berubah dari waktu ke waktu. Penelitian selanjutnya disarankan menambahkan variabel perilaku konsumen yang lebih beragam serta menggunakan dataset yang bersifat longitudinal atau dinamis sehingga proses segmentasi dapat menghasilkan gambaran pasar yang lebih komprehensif dan representatif.

#### Daftar Pustaka

- [1] L. Abednego, C. E. Nugraheni, and A. Salsabina, "Customer Segmentation: Transformation from Data to Marketing Strategy," *Conf. Ser.*, vol. 4, no. 1, pp. 139–152, Dec. 2023, doi: 10.34306/conferenceseries.v4i1.645.
- [2] K. Affandi, H. O. J. Pramono, S. V. Handjaja, R. Febrianto, K. P. Choiss, and H. Widjojo, "THE FACTORS INFLUENCING CONTEMPORARY COFFEE CONSUMPTION IN THE PANDEMIC ERA," *J. Manaj.*, vol. 19, no. 1, pp. 1–17, Jul. 2022, doi: 10.25170/jm.v19i1.2951.
- [3] E. W. Agustin, K. Uthami, A. I. Ulfa, L. Efrizoni, and R. Rahmaddeni, "Optimization of Customer Segmentation in the Retail Industry Using the K-Medoid Algorithm," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 5, no. 3, pp. 766–775, Jun. 2025, doi: 10.57152/malcom.v5i3.1977.
- [4] O. Akande, E. O. Asani, and B. Dautare, "Customer Segmentation Through RFM Analysis and K-Means Clustering: Leveraging Data-Driven Insights for Effective Marketing Strategy," *Ceddi J. Inf. Syst. Technol.*, vol. 3, no. 1, pp. 14–25, Apr. 2024, doi: 10.56134/jst.v3i1.81.
- [5] F. Arifianto, J. Hasudungan, A. Muzaky, and H. T. Y. Achsan, "Segmentasi Pelanggan Berdasarkan Recency, Frequency, dan Monetary dengan K-Means Clustering: Studi Kasus Toko Pakaian Almost Famous," *J. Teknol. Inform. dan Komput.*, vol. 10, no. 1, pp. 122–135, Mar. 2024, doi: 10.37012/jtik.v10i1.2096.
- [6] F. Arifianto, J. Hasudungan, A. Muzaky, and H. T. Y. Achsan, "Segmentasi Pelanggan Berdasarkan Recency, Frequency, dan Monetary dengan K-Means Clustering: Studi Kasus Toko Pakaian Almost Famous," *J. Teknol. Inform. dan Komput.*, vol. 10, no. 1, pp. 122–135, 2024, doi: 10.37012/jtik.v10i1.2096.
- [7] D. A. Awaliyah, Budi Prasetyo, R. Muzayanah, and A. D. Lestari, "Optimizing Customer Segmentation in Online Retail Transactions through the Implementation of the K-Means Clustering Algorithm," *Sci. J. Informatics*, vol. 11, no. 2, pp. 539–548, Jun. 2024, doi: 10.15294/sji.v11i2.6137.
- [8] R. Y. Daulay, R. A. Passalaras, and J. Heikal, "Customer Segmentation Using K-Means Clustering with SPSS Program in a Case Study of Consumer Interest in Current Coffee Shop," *Budg. J. Business, Manag. Account.*, vol. 5, no. 2, pp. 721–740, Apr. 2024, doi: 10.31539/budgeting.v5i2.9288.
- [9] S. F. Djun, I. G. A. Gunadi, and S. Sariyasa, "Analisis Segmentasi Pelanggan pada Bisnis dengan Menggunakan Metode K-Means Clustering pada Model Data RFM," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 5, no. 4, pp. 354–364, Feb. 2024, doi: 10.35746/jtim.v5i4.434.
- [10] C. Hafidz Ardana *et al.*, "Segmentasi Pelanggan Penjualan Online Menggunakan Metode K-means Clustering," 2024.
- [11] N. Hanum, C. Prianto, W. I. Rahayu, and H. D. Kishendrian, "SINTECH Journal | 137 Penerapan Metode Clustering Dalam Segmentasi Pelanggan Perusahaan Logistik", [Online]. Available: <https://doi.org/10.31598>
- [12] A. M. Husein, F. K. Waruwu, Y. M. T. Batu Bara, M. Donpril, and M. Harahap, "Clustering Algorithm For Determining Marketing Targets Based Customer Purchase Patterns And Behaviors," *Sinkron*, vol. 6, no. 1, pp. 137–143, Oct. 2021, doi: 10.33395/sinkron.v6i1.11191.
- [13] I. Iqbal, N. Hidayat, D. P. Gevano, and A. P. R. Ilahi, "Segmentasi Pelanggan Menggunakan K-Means Clustering Berdasarkan Data Kepribadian dan Pola Konsumsi," *J. Tek. Inform.*, vol. 6, no. 5, pp. 3914–3924, Oct. 2025, doi: 10.52436/1.jutif.2025.6.5.5140.
- [14] Y. Liu, "Customer Segmentation in User Behavior Analysis: A Comparative Study of Clustering Algorithms," 2023.
- [15] A. G. Ramadhan, "How to cite: Ade Guntur Ramadhan (2023) Data Mining Untuk Segmentasi Pelanggan dengan Algoritma K-Means: Studi Kasus pada Data Pelanggan di Toko Retail DATA MINING UNTUK SEGMENTASI PELANGGAN DENGAN ALGORITMA K-MEANS: STUDI KASUS PADA DATA PELANGGAN DI TOKO RETAIL," vol. 8, no. 10, 2023, doi: 10.36418/syntax-literate.v6i6.
- [16] E. R. Sitorus and I. Nugraha, "Customer Segmentation Analysis with RFM Model (Recency, Frequency, Monetary) and K-Means Clustering: Case Study of Bottled Water Sales at PT XYZ,"

- vol. X, no. 2, 2025.
- [17] T. Soni Madhulatha, "AN OVERVIEW ON CLUSTERING METHODS," vol. 2, no. 4, pp. 719–725, 2012, [Online]. Available: [www.iosrjen.org](http://www.iosrjen.org)
- [18] K. Tabianan, S. Velu, and V. Ravi, "K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data," *Sustain.*, vol. 14, no. 12, Jun. 2022, doi: 10.3390/su14127243.