

Klasterisasi Wilayah Penjualan Pada Toko Roni Menggunakan *K-Means*

Ermala Handayani, Nirwana Hendrastuty

Program Studi Sistem Informasi, Universitas Teknokrat Indonesia
Jl. Zainal Abidin Pagaralam No.9-11 Labuhan Ratu, Bandar
Lampung

Email : ermalahandayani@gmail.com

ABSTRAK

Penelitian ini bertujuan untuk mengelompokkan wilayah pelanggan toko Roni berdasarkan intensitas pembelian menggunakan algoritma *K-Means Clustering*. Tahapan penelitian meliputi pengumpulan data, pra-pemrosesan, pelabelan data, ekstraksi fitur, pemodelan implementasi, pengujian hingga evaluasi. Data yang digunakan sebanyak 4.999 transaksi penjualan yang terjadi selama Maret hingga April 2022, dan terdiri dari 43 atribut, namun hanya atribut *Kota Penerima* dan *Jumlah Barang* yang dianalisis. Setelah melalui tahap normalisasi dan penghapusan *missing values*, data dibagi menjadi tiga klaster berdasarkan metode Elbow yang menunjukkan titik optimum pada $k=3$, proses klasterisasi dilakukan menggunakan algoritma *K-Means* dengan valuasi berdasarkan nilai SSE (Sum of Squared Errors), yang menunjukkan total 0.1137.

Hasil penelitian menunjukkan bahwa wilayah dapat dikelompokkan menjadi tiga kategori: rendah, sedang, dan tinggi berdasarkan jumlah transaksi. Klaster dengan aktivitas pembelian tertinggi direkomendasikan sebagai target utama untuk strategi promosi toko. Hasil ini membuktikan bahwa algoritma *K-Means* dapat secara efektif digunakan dalam analisis segmentasi wilayah berbasis data transaksi pelanggan.

Kata Kunci : *K-Means, Klasterisasi, Wilayah Penjualan, Data Mining, Elbow Method*

ABSTRACT

This study aims to group the customer areas of Roni's store based on purchasing intensity using the K-Means Clustering algorithm. The research stages include data collection, pre-processing, data labeling, feature extraction, implementation modeling, testing to evaluation. The data used were 4,999 sales transactions that occurred during March to April 2022, and consisted of 43 attributes, but only the Recipient City and Number of Goods attributes were analyzed. After going through the normalization stage and eliminating missing values, the data was divided into three clusters based on the Elbow method which showed an optimum point at $k = 3$, the clustering process was carried out using the K-Means algorithm with a valuation based on the SSE (Sum of Squared Errors) value, which showed a total of 0.1137. The results of the study show that areas can be grouped into three categories: low, medium, and high based on the number of transactions. The cluster with the highest purchasing activity is recommended as the main target for store promotion strategies. These results prove that the K-Means algorithm can be effectively used in regional segmentation analysis based on customer transaction data.

Keywords: *K-Means, Clustering, Sales Region, Data Mining, Elbow Method*

1. PENDAHULUAN

Clustering atau klasterisasi adalah metode pengelompokan data, dapat dikatakan bahwa klasterisasi merupakan proses pembagian data yang menghasilkan kelompok data yang disebut objek berdasarkan dengan tingkat kemiripan karakteristik yang dimiliki. *Clustering* adalah proses mengelompokkan data ke dalam beberapa *cluster* atau kelompok, sehingga data dalam satu *cluster* memiliki tingkat kemiripan yang maksimum dan data antar *cluster* memiliki kemiripan yang minimum (Gantara & Ali, 2023; Indriyani & Irfiani, 2023).

Oleh sebab itu, klasterisasi dapat digunakan dalam menganalisis perencanaan bisnis guna menentukan langkah kegiatan bisnis seperti strategi pemasaran dan segmentasi pasar, misalnya pengelompokan wilayah konsumen berdasarkan tingkat pembelian produk (Widad & Fatah, 2025; Septianto, Faqih, & Rinaldi, 2025). Dalam konteks penjualan daring, strategi promosi yang efektif dapat ditentukan dari wilayah dengan intensitas transaksi tinggi (Putra, Azzahra, & Erlanda, 2023). Data penjualan sangat berguna untuk perencanaan promosi seperti pengadaan iklan pada wilayah yang minat pembelinya tinggi. Bagi para penjual atau *seller online*, pemasaran sangat dibutuhkan dalam memperkenalkan, menawarkan, dan meningkatkan minat calon konsumen untuk membeli produk yang dijual. Maka dari itu data penjualan seharusnya sangat berguna untuk merencanakan promosi penjualan seperti pengadaan iklan pada wilayah yang minat pembelinya paling tinggi sesuai dengan data yang ada. *Online shop* adalah suatu proses pembelian barang atau jasa dari mereka yang menjual melalui internet, atau layanan jual beli secara online tanpa harus bertatap muka dengan penjual atau pihak pembeli secara langsung (Sari, 2015). Toko online biasanya hanya fokus berjualan di media sosial seperti Instagram, Facebook, dan Twitter dengan cara pemesanan melalui WhatsApp.

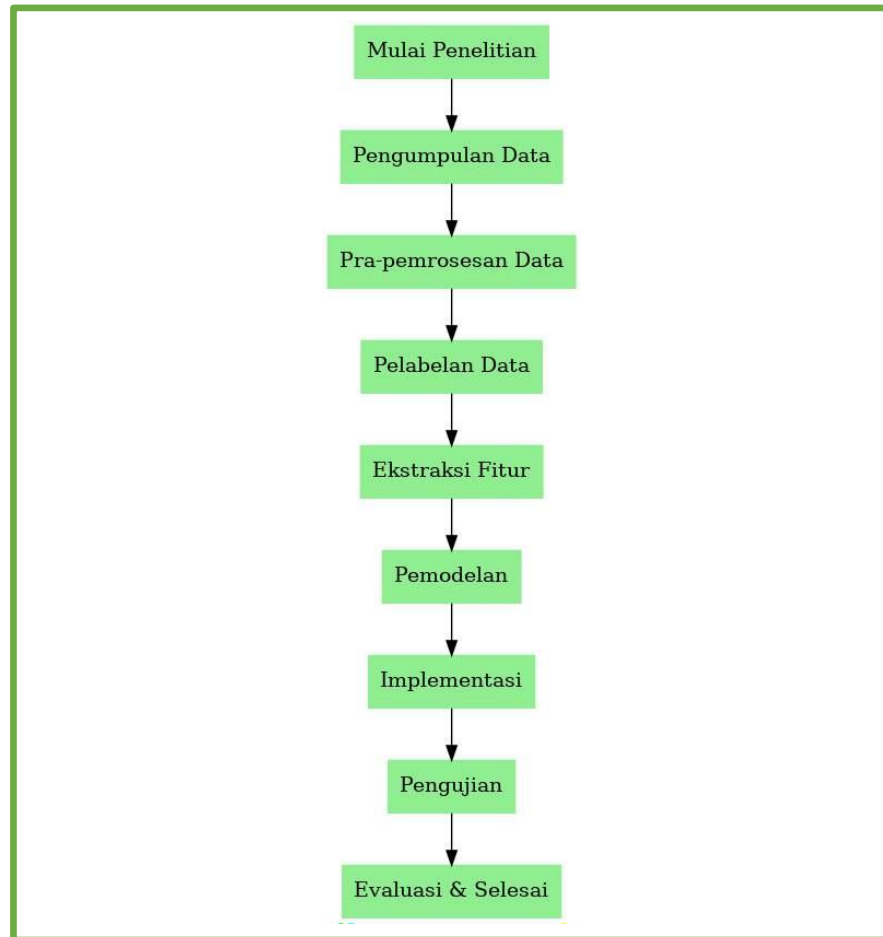
Hal yang sama dilakukan oleh toko *Online Shop* Roni yang beroperasi di Kecamatan Jati Agung, Kabupaten Lampung Selatan. Toko online ini menjual produk-produk seperti pembersih keramik dan serum wajah dengan memanfaatkan berbagai platform media sosial seperti Instagram, Facebook, dan Tiktok Sebagai media promosi dan komunikasi dengan calon konsumen. Proses transaksi dilakukan secara langsung melalui aplikasi WhatsApp, dimana pembeli dapat memesan produk serta mengajukan pertanyaan seputar barang yang ditawarkan. Seiring berjalannya waktu, aktivitas jual

beli yang berlangsung secara berulang menghasilkan kumpulan data transaksi yang cukup besar. Namun data tersebut belum pernah dimanfaatkan secara optimal. Padahal, data ini menyimpan informasi penting untuk mendukung pengambilan keputusan bisnis, seperti pola pembelian pelanggan, segmentasi pasar, dan wilayah dengan permintaan tinggi (Harjono, Utami, & Putri, 2023; Widiyanto, Zakaria, & Irvan, 2025; Prasetyo & Wibowo, 2022). Analisis terhadap data transaksi ini dapat memberikan berbagai informasi penting yang berguna bagi penjual atau seller, seperti pola pembelian pelanggan, segmentasi pasar, dan wilayah dengan permintaan tertinggi. Dengan demikian, data transaksi yang semula hanya tersimpan begitu saja dapat diolah menjadi sumber informasi strategis untuk mendukung pengambilan keputusan bisnis yang lebih efektif dan efisien di masa mendatang.

Dikarenakan banyaknya transaksi yang terjadi, pemilik toko mengalami kesulitan dalam mengetahui daerah dengan tingkat pemesanan tinggi. Maka diperlukan pengelompokan wilayah berdasarkan aktivitas pembelian (Nuryaman et al., 2023; Sandova, Kurniawan, & Supratati, 2024; Damayanti & Widodo, 2021). Penjual hendak melakukan fokus penjualan atau promosi di daerah yang terbilang cukup tinggi minat pembelinya berdasarkan data transaksi yang ada. Maka dari itu, diperlukan sebuah perhitungan pengelompokan wilayah berdasarkan aktivitas pembeli untuk membantu proses promosi penjualan. Penelitian serupa dilakukan oleh Dini Marlina dkk. (2018) yang melakukan penelitian dengan judul “Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak” di mana penelitian ini bertujuan untuk mengetahui wilayah mana yang memiliki tingkat anak cacat yang tinggi sehingga akan dapat dilakukan penanggulangan secara dini, dirasa memiliki kasus yang hampir sama dalam pengelompokan data. Metode K-Means digunakan untuk mengelompokkan wilayah pembelian tertinggi agar dapat membantu penjual merencanakan promosi dan meningkatkan pendapatan (Melisa, Defit, & Sovia, 2025; Almaripat, Faqih, & Rinaldy, 2024; Hendrastuty, 2024).

2. METODE

Berikut merupakan metode yang digunakan pada penelitian ini:



Gambar 1. Metode penelitian

Pengumpulan Data

Dalam penelitian ini, penulis mengumpulkan data dengan melakukan tanya jawab secara langsung dengan pemilik toko online, kemudian pemilik toko memberikan file data transaksi penjualan kepada penulis. Dataset yang diberikan merupakan data transaksi penjualan dari tahun 2021 hingga 2022 sebanyak 19.766 transaksi, dimuat dari file excel. Namun dalam penelitian ini hanya akan mengolah data sebanyak 4.999 saja yaitu data transaksi penjualan dalam kurun waktu bulan Maret 2022 – April 2022. Dataset awal merupakan file Excel yang diubah dalam format CSV. Dataset mencakup atribut seperti nama barang, jumlah barang, kota penerima dan informasi lainnya. Namun, sebelum proses analisis dilakukan, data dibersihkan dengan menghapus kolom yang tidak relevan terhadap tujuan penelitian.

Pra-pemrosesan Data

Data yang digunakan dalam penelitian ini terdiri dari 43 kolom dan 4.999 data transaksi penjualan. Kolom yang akan digunakan adalah Kota Penerima dan Jumlah Barang. Langkah yang dilakukan pada tahap ini sebagai berikut:

1. Penghapusan kolom yang tidak relevan.
2. Penanganan missing value.
3. Label Encoding jika diperlukan.
4. Normalisasi menggunakan StandardScaler.
5. Pemeriksaan nilai unik pada atribut utama.

Pelabelan Data

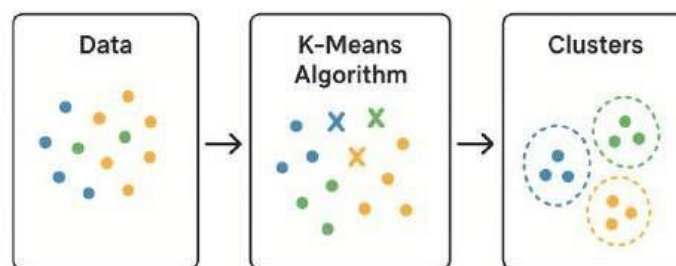
Pelabelan data menganotasi data mentah dengan label yang bermakna, sehingga memberikan konteks dan kategorisasi untuk dipahami oleh model Machine Learning (ML). Label ini berfungsi sebagai panduan penting untuk model ML, sehingga dapat menafsirkan data secara efektif.

Ekstraksi Fitur

Ekstraksi fitur adalah proses mengubah data mentah menjadi representasi numerik atau fitur yang dapat diolah oleh algoritma pembelajaran mesin (machine learning). Proses ini penting karena data mentah seringkali tidak dapat langsung diolah oleh algoritma ML, sehingga perlu diubah menjadi bentuk yang lebih terstruktur dan relevan dengan tujuan analisis.

Pemodelan

Tahap pemodelan dilakukan untuk mengelompokkan data penjualan berdasarkan karakteristik tertentu menggunakan Algoritma K-Means Clustering. Adapun variabel yang digunakan dalam pemodelan ini meliputi Kota Penerima dan Jumlah Barang.



Gambar 2. Tahap pemodelan

Implementasi

Implementasi dalam data mining meliputi berbagai metode dan teknik untuk memperoleh pola dan pengetahuan dari data yang besar. Implementasi sistem yang digunakan pada tahap ini ialah Klastering (*Clustering*) menggunakan Algoritma K-Means, yaitu mengelompokkan data yang serupa, misalnya pada kasus ini untuk mengetahui wilayah *customer* dengan pembelian atau transaksi tertinggi. K-Means adalah algoritma yang bertujuan mengelompokkan data ke dalam sejumlah klaster berdasarkan tingkat kemiripan. Proses K-Means dimulai dengan menentukan jumlah klaster k . Algoritma *K-Means* digunakan untuk membagi data menjadi 3 klaster. Penentuan jumlah klaster didukung oleh visualisasi metode Elbow.

Langkah-langkah:

1. Data numerik dari jumlah pembelian dan lokasi dikumpulkan.
2. Dilakukan normalisasi.
3. Model K-Means dijalankan dengan $n_clusters=3$.
4. Hasil klaster ditampilkan menggunakan visualisasi scatter.

Pengujian

Tahap pengujian pada penelitian ini dilakukan untuk mengevaluasi kualitas hasil pengelompokan data menggunakan algoritma K-Means, yang bertujuan untuk menguji apakah model yang sudah dibuat bisa bekerja dengan baik pada data baru yang belum pernah dilihat sebelumnya.

Evaluasi

Evaluasi dilakukan untuk mengukur kualitas dan akurasi model dalam mengenali pola dan pengelompokan data baru (Rahmawati, Prihartono, & Fathurrohman, 2025; Agustiar, Sari, & Maulana, 2025). Proses ini penting untuk memastikan bahwa model tidak hanya bekerja baik pada data latih, tetapi juga mampu memberikan hasil yang akurat dan dapat diandalkan saat diterapkan pada data baru.

3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan Python dengan pustaka Pandas, Scikit-Learn, dan Matplotlib. Klusterisasi dilakukan untuk 3 klaster sesuai dengan hasil Elbow Method. Proses klusterisasi diawali dengan pembersihan dan transformasi data, seperti penghapusan kolom tidak relevan, penanganan missing value, label encoding, dan normalisasi (Bui & Bahtiar, 2024; Alvianatinova et al., 2024). Evaluasi dilakukan menggunakan SSE, dan hasil akhir visualisasi memperlihatkan wilayah dengan jumlah pembelian tertinggi, sedang, dan rendah.

Data yang digunakan dalam penelitian ini terdiri dari 43 kolom dan 4.999 diambil dari data transaksi penjualan pada bulan Maret 2022 – April 2022. Kolom yang akan digunakan adalah Kota Penerima dan Jumlah Barang. Langkah yang dilakukan pada tahap ini sebagai berikut:

1. Penghapusan kolom yang tidak digunakan seperti informasi pengirim, status kirim, dan rincian e-commerce.

	No Waybill	Sudah Terkirim	No Order	Tanggal Pengiriman	Status	Klien VIP	Nama Pengirim	Telepon Pengirim	Provinsi Pengirim	Kota Pengirim	...	Total Biaya	Biaya discount	Total Biaya Setelah diskon	Diskon	100	Diterima Oleh	Hubungan
0	1131438805	Blum Terkirim	2.200000e+13	4/25/2022	Menunggu Diproses	BERLIAN SHOPS	JH NATURE SERUM	6.290000e+11	JAWA BARAT	BANDUNG	...	31.0	12.4	18.6	Ada diskon	157.0	NaN	NaN
1	1131438697	Blum Terkirim	2.200000e+13	4/25/2022	Menunggu Diproses	BERLIAN SHOPS	JH NATURE SERUM	6.290000e+11	JAWA BARAT	BANDUNG	...	83.0	21.2	31.8	Ada diskon	175.0	NaN	NaN
2	1131430616	Blum Terkirim	2.200000e+13	4/25/2022	Menunggu Diproses	BERLIAN SHOPS	JH NATURE SERUM	6.290000e+11	JAWA BARAT	BANDUNG	...	31.0	12.4	18.6	Ada diskon	150.0	NaN	NaN

3 rows x 43 columns

Gambar 3. Data sebelum penghapusan kolom

	Kota Penerima	Jumlah Barang
0	275	1
1	437	1
2	206	1

Gambar 4. Data setelah penghapusan kolom

2. Mengatasi nilai Nan (missing value) pada data.

Kota Penerima	0
Jumlah Barang	5

Gambar 5. Data kolom yang terdapat nilai Nan

Kota Penerima	0
Jumlah Barang	0

Gambar 6. Data kolom setelah dibersihkan

Setelah data dibersihkan dari nilai Nan (kosong), otomatis baris data juga menjadi berkurang dari 4999 baris menjadi 4994 baris.

3. Label Encoding untuk mengubah data menjadi suatu kategorikal, ditahap ini data diubah menjadi bilangan integer.

	Kota Penerima	Jumlah Barang
0	OGAN KOMERING ULU TIMUR	2.0
1	WAJO	2.0
2	LUBUK LINGGAU	2.0

Gambar 7. Data sebelum proses kategorikal ke bentuk integer

	Kota Penerima	Jumlah Barang
0	275	1
1	437	1
2	206	1

Gambar 8. Data sudah dalam bentuk interger

4. Normalisasi menggunakan StandardScaler untuk menyetarakan skala nilai numerik. Tahap normalisasi terhadap data untuk menormalkan setiap fitur dengan menskalakan data ke suatu rentang, sehingga nilainya memiliki nilai minimum dan maksimum yang ditentukan dengan default 0 hingga 1. Dalam arti lain normalisasi untuk mengatur database agar lebih rapi dan bebas dari

redundansi data (menyimpan data secara berulang), sehingga analis dapat menggunakan secara efisien.

	Kota Penerima	Jumlah Barang
0	275	1
1	437	1
2	206	1

Gambar 9. Data sebelum dinormalisasi

	Kota Penerima	Jumlah Barang
0	0.619369	0.166667
1	0.984234	0.166667
2	0.463964	0.166667

Gambar 10. Data setelah dinormalisasi (skala)

5. Pemeriksaan nilai unik pada atribut utama seperti Kota Penerima, Jumlah Barang, dan Nama Barang. Pada tahap ini digunakan untuk membantu melihat berapa banyak kategori atau variasi data yang ada dalam suatu atribut. Kemudian didapatkan hasil bahwa :
 - a. Kolom “Kota Penerima” memiliki 445 nilai unik, berarti data berasal dari 445 kota yang berbeda.
 - b. Kolom “Jumlah Barang” memiliki 6 nilai unik, berarti data berasal dari 6 jenis barang yang berbeda.
6. Menggunakan *Elbow Method* untuk menentukan jumlah cluster yang terbaik berdasarkan nilai SSE (Sum of Squared Errors) atau WCSS (Within-Cluster Sum of Squares). Berdasarkan metode elbow, dilakukan perhitungan SSE terhadap nilai k dari 1 hingga 10. Berikut merupakan tahapan yang dikerjakan:
 - a. Dimulai dari menghitung jarak centeroid, inisialisasi centeroid awal (dipilih dari data).

Index	Kota Penerima	Jumlah Barang
0	0.619369	0.2
1	0.984234	0.2
2	0.463964	0.2
3	0.416667	0.2
4	0.903153	0.2
5	0.272523	0.8
6	0.209459	0.2
7	0.479730	0.8

Gambar 11. Contoh data training

Misalnya kita ambil centeroid awal dari tiga data:

$$C_1 = \text{Data ke-0} = (0.619369, 0.2)$$

$$C_2 = \text{Data ke-5} = (0.272523, 0.8)$$

$$C_3 = \text{Data ke-1} = (0.984234, 0.2)$$

b. Hitung jarak Euclidean ke setiap centroid

Rumus jarak Euclidean

$$d = \sqrt{(x_1 - c_1)^2 + (x_2 - c_2)^2}$$

Keterangan:

d = Jarak Euclidean antara sebuah titik data dan centroid
(pusat kluster)

x_1, x_2 = Nilai fitur/atribut dari titik data (misalnya: jumlah barang)

c_1, c_2 = Nilai fitur dari centroid klaster yang sedang dihitung jaraknya terhadap titik data

Hitung jarak masing-masing data ke tiga centroid, contoh untuk data ke-0:

Data ke-0: (0.619369, 0.2)

$$\text{- Jarak ke } C_1 = \sqrt{((0.619369 - 0.619369)^2 + (0.2 - 0.2)^2)} = 0$$

$$\begin{aligned} \text{- Jarak ke } C_2 &= \sqrt{((0.619369 - 0.272523)^2 + (0.2 - 0.8)^2)} \\ &= \sqrt{(0.1204 + 0.36)} = 0.717 \end{aligned}$$

$$\begin{aligned} \text{- Jarak ke } C_3 &= \sqrt{((0.619369 - 0.984234)^2 + (0.2 - 0.2)^2)} \\ &= \sqrt{(0.1328)} = 0.364 \end{aligned}$$

Lakukan hal yang sama untuk semua data, kemudian setiap data dimasukkan ke kluster dengan jarak terpendek.

Index	Data Point (x, y)	Jarak ke C1	Jarak ke C2	Jarak ke C3	Cluster
0	(0.619, 0.2)	0.000	0.717	0.364	C1
1	(0.984, 0.2)	0.364	0.857	0.000	C3
2	(0.464, 0.2)	0.155	0.625	0.520	C1
3	(0.417, 0.2)	0.202	0.568	0.567	C1
4	(0.903, 0.2)	0.284	0.788	0.082	C3
5	(0.273, 0.8)	0.717	0.000	0.857	C2
6	(0.209, 0.2)	0.410	0.618	0.775	C1
7	(0.480, 0.8)	0.609	0.215	0.726	C2

Gambar 12. Hasil perhitungan jarak Euclidean

- c. Memperbarui centroid baru dengan menghitung rata-rata koordinat X dan Y dari masing-masing cluster.

C1 (Cluster 1):

- Anggota: Index 0, 2, 3, 6
- Rata-rata X = $(0.619 + 0.464 + 0.417 + 0.209)/4 = 0.427$
- Rata-rata Y = $(0.2 + 0.2 + 0.2 + 0.2)/4 = 0.2$
- Centroid baru C1 = (0.427, 0.2)

C2 (Cluster 2):

- Anggota: Index 5, 7
- Rata-rata X = $(0.273 + 0.480)/2 = 0.377$
- Rata-rata Y = $(0.8 + 0.8)/2 = 0.8$
- Centroid baru C2 = (0.377, 0.8)

C3 (Cluster 3):

- Anggota: Index 1, 4, 8, 9
- Rata-rata X = $(0.984 + 0.903 + 0.878 + 0.939)/4 = 0.926$
- Rata-rata Y = 0.2
- Centroid baru C3 = (0.926, 0.2)

Ulangi perhitungan jarak dengan centroid baru, lalu lihat apakah cluster anggota berubah. Jika tidak ada perubahan, maka proses selesai.

d. SSE dihitung menggunakan rumus:

$$SSE = \sum_{i=1}^n [(x_i - c_x)^2 + (y_i - c_y)^2]$$

Keterangan:

SSE = Total kesalahan kuadrat dari semua titik data terhadap centroid clusternya

$\sum_{i=1}^n$ = Jumlahkan dari data ke-1 hingga data ke-n

n = Jumlah data dalam satu cluster

x_i = Nilai fitur pertama (misal: kota penerima yang sudah dinormalisasi)

y_i = Nilai fitur kedua (misal: jumlah barang yang sudah dinormalisasi)

Cluster 1 (C1 = (0.427, 0.2))

Anggota: 0, 2, 3, 6

Index	Data (x, y)	SSE per data
0	(0.619, 0.2)	$(0.619 - 0.427)^2 + (0.2 - 0.2)^2 = 0.0369$
2	(0.464, 0.2)	$(0.464 - 0.427)^2 = 0.0014$
3	(0.417, 0.2)	$(0.417 - 0.427)^2 = 0.0001$
6	(0.209, 0.2)	$(0.209 - 0.427)^2 = 0.0475$

Gambar 13. SSE cluster 1

$$SSE \text{ C1} = 0.0369 + 0.0014 + 0.0001 + 0.0475 = 0.0859$$

Cluster 2 (C2 = (0.337, 0.8))

Anggota: 5, 7

Index	Data (x, y)	SSE per data
5	(0.273, 0.8)	$(0.273 - 0.377)^2 = 0.0108$
7	(0.480, 0.8)	$(0.480 - 0.377)^2 = 0.0106$

Gambar 14. SSE cluster 2

$$SSE \text{ C2} = 0.0108 + 0.0106 = 0.0214$$

Cluster 3 (C3 = (0.926, 0.2))

Anggota: 1, 4, 8, 9

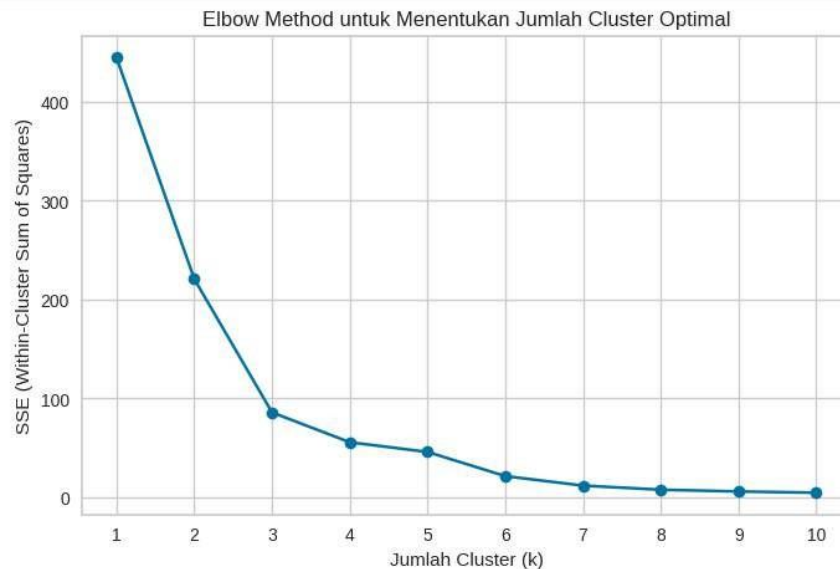
Index	Data (x, y)	SSE per data
1	(0.984, 0.2)	$(0.984 - 0.926)^2 = 0.0034$
4	(0.903, 0.2)	$(0.903 - 0.926)^2 = 0.0005$
8	(0.878, 0.2)	$(0.878 - 0.926)^2 = 0.0023$
9	(0.939, 0.2)	$(0.939 - 0.926)^2 = 0.0002$

Gambar 15. SSE cluster 3

$$\text{SSE C3} = 0.0034 + 0.0005 + 0.0002 = 0.0064$$

$$\begin{aligned} \text{SSE Total (K=3)} &= \text{SSE C1} + \text{SSE C2} + \text{SSE C3} \\ &= 0.0859 + 0.0214 + 0.0064 = 0.1137 \end{aligned}$$

- e. Tahap Elbow Method, untuk metode ini menghitung SSE untuk beberapa nilai K, pada tahap ini menggunakan K range = 1 hingga 11 artinya mencoba jumlah cluster dari 1 sampai 10, kemudian ditampilkan dalam bentuk grafik seperti berikut:

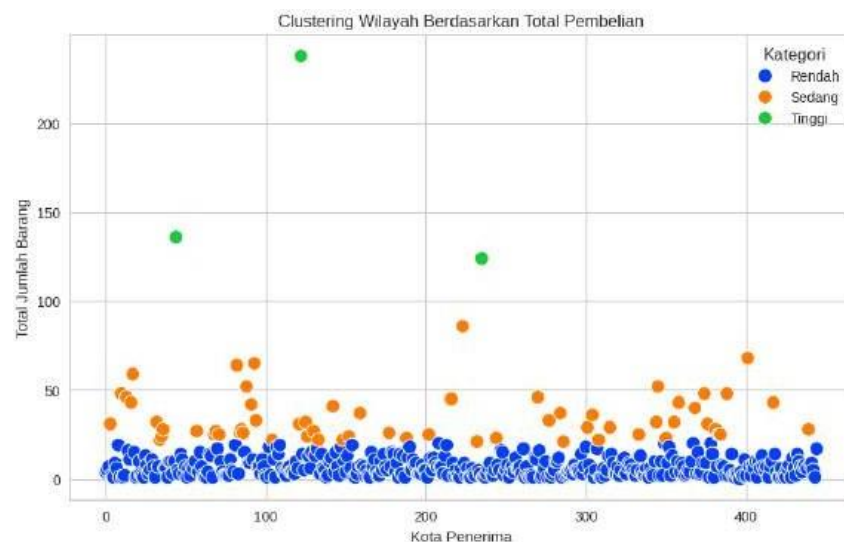
**Grafik 1.** Tampilan grafik *Elbow*

Nilai SSE kemudian diplot terhadap jumlah kluster K. Dapat dilihat grafik menunjukkan penurunan yang tajam dari $k = 1$ ke $k = 3$, titik tekuk (elbow) pada grafik juga menunjukkan bahwa jumlah kluster optimal adalah $k = 3$, di mana penurunan SSE mulai melambat.

Dari hasil implementasi K-Means yang dilakukan, wilayah penjualan berhasil dikelompokkan ke dalam tiga klaster:

- Klaster Rendah: Wilayah dengan jumlah transaksi rendah
- Klaster Sedang: Wilayah dengan jumlah transaksi sedang
- Klaster Tinggi: Wilayah dengan jumlah transaksi tinggi

Visualisasi data menunjukkan pemetaan terkait wilayah mana saja yang perlu difokuskan dalam kegiatan promosi. Grafik 2 menggambarkan hubungan antara kota penerima dengan total jumlah barang, setiap titik pada grafik mewakili satu pasang data.



Grafik 2. Plot scatter clustering wilayah berdasarkan total pembelian

Berikut ini adalah tampilan hasil wilayah yang memiliki transaksi 10 teratas setelah dilakukan proses klasterisasi K-Means.

	Kota Penerima	Jumlah Barang	Cluster	Kategori	Cluster
122	122	238	1		Tinggi
44	44	136	1		Tinggi
235	235	124	1		Tinggi
223	223	80	0		Sedang
93	93	65	0		Sedang
82	82	64	0		Sedang
401	401	62	0		Sedang
17	17	59	0		Sedang
88	88	52	0		Sedang
345	345	52	0		Sedang

Gambar 16. 10 Wilayah pembelian teratas

Keterangan kode kota penerima:

122 = Jakarta	82 = Cibinong
44 = Bekasi	401 = Tangerang
235 = Medan	17 = Bandung
223 = Manado	88 = Ciputat
93 = Denpasar	345 = Semarang

Proses manual perhitungan centroid, jarak Euclidean, serta visualisasi scatter plot dan grafik Elbow mendukung hasil bahwa $k = 3$ adalah jumlah klaster optimal (Valentina et al., 2025). Dari hasil klasterisasi ditemukan 10 wilayah teratas berdasarkan transaksi terbanyak seperti Jakarta, Bekasi, Medan, dan Bandung. Visualisasi ini dapat menjadi dasar bagi strategi promosi toko berdasarkan data aktual pelanggan (Ardana, Khoyum, & Faisal, 2024; Hendrastuty, 2024).

4. SIMPULAN

Penelitian ini berhasil menerapkan metode *K-Means Clustering* untuk mengelompokkan wilayah penjualan berdasarkan intensitas transaksi pada Toko Roni selama periode Maret hingga April 2022. Dataset yang dianalisis berjumlah 4.999 data transaksi, setelah melalaui tahap pembersihan dari total awal 4.999 data. Proses penelitian meliputi tahapan pengumpulan data, pra-pemrosesan (penghapusan kolom tidak relevan, penanganan missing value, normalisasi, dan encoding), pelabelan data, ekstraksi fitur, pemodelan menggunakan algoritma K-Means, hingga pengujian dan evaluasi model.

Penentuan jumlah kluster dilakukan melalui Metode Elbow, dengan hasil optimal pada jumlah 3 kluster ($k = 3$). Model uji menggunakan perhitungan jarak Euclidean dan Sum of Squared Errors (SSE), dengan total nilai SSE sebesar 0.1137, yang menunjukkan bahwa model membentuk kluster dengan tingkat kesalahan rendah dan pengelompokan yang baik. Dari hasil klusterisasi, dapat disimpulkan bahwa wilayah dalam **kluster Sedang** merupakan target potensial promosi karena menunjukkan aktivitas transaksi tertinggi. Sementara itu, wilayah dalam **kluster Rendah** menunjukkan potensi untuk dikembangkan atau dianalisis lebih lanjut terkait faktor yang memengaruhi rendahnya pembelian.

Secara keseluruhan, metode *K-Means Clustering* mampu dalam mengelompokkan wilayah berdasarkan jumlah transaksi pembelian. Informasi yang dihasilkan dapat dimanfaatkan oleh Toko Roni untuk menentukan wilayah prioritas dalam promosi produk serta mengoptimalkan strategi pemasaran online berbasis data pelanggan yang aktual. Penelitian ini dapat dikembangkan lebih lanjut dengan mempertimbangkan atribut lain seperti waktu transaksi atau jenis produk untuk hasil klusterisasi yang lebih tajam (Alvianatinova et al., 2024; Gantara & Ali, 2023; Hendrastuty, 2024; Prasetyo & Wibowo, 2022; Damayanti & Widodo, 2021).

5. DAFTAR PUSTAKA

- Agustiar, F. D., Sari, B. N., & Maulana, I. (2025). Penerapan data mining untuk pengelompokan produk penjualan menggunakan algoritma K-Means (studi kasus Toko Agung Makmur Jaya). *JATI: Jurnal Mahasiswa Teknik Informatika*, 9(1), 1–12. <https://doi.org/10.36040/jati.v9i1.12178>
- Almaripat, M., Faqih, A., & Rinaldy, A. R. (2024). Sales data clustering analysis using K-Means method for marketing strategy development. *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, 4(2), 1–11. <https://doi.org/10.59934/jaiea.v4i2.792>
- Alvianatinova, V., Ali, I., Rahaningsih, N., & Bahtiar, A. (2024). Penerapan algoritma K-Means clustering dalam pengelompokan data penjualan supermarket berdasarkan cabang. *JATI: Jurnal Mahasiswa Teknik Informatika*, 8(2), 1–12. <https://doi.org/10.36040/jati.v8i2.8993>
- Ardana, C. H., Khoyum, A. A. A., & Faisal, M. (2024). Segmentasi pelanggan penjualan online menggunakan metode K-Means clustering. *JISKA: Jurnal Informatika Sunan Kalijaga*, 9(1), 1–9. <https://doi.org/10.14421/jiska.2024.9.1.1-9>
- Bui, M. A., & Bahtiar, A. (2024). Implementasi metode algoritma K-Means clustering untuk mengelompokkan transaksi penjualan barang di Toko Arino. *JATI: Jurnal Mahasiswa Teknik Informatika*, 8(2), 1–10. <https://doi.org/10.36040/jati.v8i2.8975>
- Damayanti, N. A., & Widodo, A. (2021). Analisis segmentasi pelanggan berdasarkan data transaksi menggunakan metode K-Means clustering. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 5(10), 3892–3898. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/10929>

- Marlina, D., Putri, N. F., Fernando, A., & Ramadhan, A. (2018). Implementasi algoritma K-Medoids dan K-Means untuk pengelompokkan wilayah sebaran cacat pada anak. *CoreIT: Jurnal Hasil Penelitian Ilmu Komputer dan Teknologi Informasi*, 4(2), 64–71. <https://doi.org/10.24014/coreit.v4i2.4498>. moraref.kemenag.go.id/jurnal/seminar-id.com+3informatika.universitasdumai.ac.id+3journal.irpi.or.id+3ejournal.uin-suska.ac.id
- Gantara, N. P., & Ali, I. (2023). Penerapan metode K-Means clustering pada penjualan barang di Sports Station. *E-Link: Jurnal Teknik Elektro dan Informatika*, 18(1), 28–39. <https://doi.org/10.30587/e-link.v18i1.5339>
- Harjono, S. W., Utami, N. W., & Putri, I. G. A. P. D. (2023). Klasterisasi tingkat penjualan pada startup Panak.id dengan algoritma K-Means. *Jurnal Ilmiah Teknologi Informasi Asia*, 17(1), 55–66. <https://doi.org/10.32815/jitika.v17i1.888>
- Hendrastuty, N. (2024). Penerapan data mining menggunakan algoritma K-Means clustering dalam evaluasi hasil pembelajaran siswa. *JIMA-ILKOM: Jurnal Ilmiah Informatika dan Ilmu Komputer*, 3(1), 1–7. <https://doi.org/10.58602/jima-ilkom.v3i1.26>
- Indriyani, F., & Irfiani, E. (2023). Clustering data penjualan pada toko perlengkapan outdoor menggunakan metode K-Means. *JUITA: Jurnal Informatika*, 7(2), 112–122. <https://doi.org/10.30595/juita.v7i2.5529>
- Melisa, S., Defit, S., & Sovia, R. (2025). Penerapan algoritma K-Means clustering untuk optimalisasi persediaan liquid vape berdasarkan data penjualan. *Jurnal KomtekInfo*, 12(1), 38–45. <https://doi.org/10.35134/komtekinfo.v12i1.620>
- Nuryaman, Y., Sudarsono, B., Faddillah, U., & Asistiyasari, A. (2023). Klasifikasi data penjualan menggunakan algoritma K-Means dan Analytic Hierarchy Process. *Jurnal Ladang Artikel Ilmu Komputer (LARIK)*, 2(1), 1–10. <https://doi.org/10.31294/larik.v2i1.1371>
- Prasetyo, E., & Wibowo, F. W. (2022). Klasterisasi data pelanggan menggunakan algoritma K-Means untuk strategi pemasaran produk. *Jurnal Ilmu Komputer dan Informatika (JIKI)*, 6(1), 55–62. <https://doi.org/10.30865/jiki.v6i1.3461>
- Putra, B. Y., Azzahra, F. Y., & Erlanda, I. A. (2023). Klasterisasi pengunjung mall menggunakan algoritma K-Means berdasarkan pendapatan dan pengeluaran. *Jurnal Informatika dan Teknik Elektro Terapan*, 11(3 s1), 1–12. <https://doi.org/10.23960/jitet.v11i3s1.3392>
- Rahmawati, R., Prihartono, W., & Fathurrohman. (2025). Optimasi stok dengan clustering data transaksi penjualan menggunakan algoritma K-Means di Konter Agung Cell. *Jurnal Informatika dan Teknik Elektro Terapan*, 13(2), 622–630. <https://doi.org/10.23960/jitet.v13i2.6302>
- Sandova, F., Kurniawan, R., & Supratati, T. (2024). Penerapan data mining menggunakan metode K-Means clustering pada penjualan tas di Asia Toserba Cirebon. *JATI: Jurnal Mahasiswa Teknik Informatika*, 8(1), 1–12. <https://doi.org/10.36040/jati.v8i1.8330>
- Septianto, M. A., Faqih, A., & Rinaldi, A. R. (2025). Klasterisasi data produksi pertanian di Kabupaten Cirebon dengan algoritma K-Means. *Jurnal Informatika dan Teknik Elektro Terapan*, 13(2), 260–269. <https://doi.org/10.23960/jitet.v13i2.6174>

- Valentina, M. F., Guntari, P. S., Jofilla, S., Gani, A., Ramadan, E. T., Albarra, M. A., & Pujiyanto. (2025). Penerapan algoritma K-Means clustering untuk analisis pola penjualan toko kue dengan RapidMiner. *Jurnal Mahasiswa Ilmu Komputer*, 6(1), 1–10. <https://doi.org/10.24127/ilmukomputer.v6i1.7768>
- Widad, R., & Fatah, Z. (2025). Penerapan algoritma K-Means clustering pada penjualan produk skincare. *Jamastika: Jurnal Mahasiswa Teknik Informatika*, 4(1), 1–11. <https://doi.org/10.35473/jamastika.v4i1.3611>
- Widianto, W., Zakaria, M. R., & Irvan. (2025). Penerapan data mining untuk klasifikasi data anggaran pendapatan dan belanja daerah menggunakan algoritma K-Means. *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, 9(2), 622–630. <https://doi.org/10.35870/jtik.v9i2.3425>