

## MODEL PENGELOMPOKAN DATA PENJUALAN RITEL UNTUK EFISIENSI MENGUNAKAN ALGORITMA K-MEANS

Nurhadiansyah<sup>1</sup>, Ahmad Faqih<sup>2</sup>, Fidy Arie Pratama<sup>3</sup>

Institut Prima Bangsa

STMIK IKMI Cirebon

Universitas Agama Islam Bunga Bangsa Cirebon

E-mail: hadijunior03@gmail.com

### Article Info

#### Article history:

Received 10 03, 2024

Revised 11 21, 2024

Accepted 12 07, 2024

#### Keywords:

First K-Means Clustering  
Second Pengelompokan Data  
Third Efisiensi  
Fourth Data Penjualan  
Fifth keyword Segmentasi  
Pelanggan

### ABSTRACT

Dalam era digital, pengelolaan data penjualan yang kompleks menjadi tantangan bagi perusahaan dalam meningkatkan efisiensi operasional dan strategi bisnis. Penelitian ini menerapkan K-Means Clustering untuk mengelompokkan data penjualan guna mengidentifikasi pola pembelian pelanggan. Metode yang digunakan mengikuti framework CRISP-DM, mencakup pemahaman bisnis, eksplorasi data, persiapan, pemodelan, evaluasi, dan implementasi.

Hasil penelitian menunjukkan bahwa data penjualan terbagi ke dalam lima kluster dengan karakteristik berbeda: Kluster 0 (278 item), Kluster 1 (226 item), Kluster 2 (150 item), Kluster 3 (156 item), dan Kluster 4 (190 item). Kluster 4 memiliki homogenitas tertinggi dengan nilai Silhouette Score 29.067, menjadikannya kluster terbaik dalam analisis ini. Evaluasi menggunakan Davies-Bouldin Index (0.994) menunjukkan pemisahan kluster yang optimal. Model ini membantu optimalisasi stok, strategi pemasaran yang lebih tepat sasaran, dan efisiensi distribusi, sehingga meningkatkan daya saing bisnis.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



## 1. PENDAHULUAN (11 PT)

Dalam era digital, jumlah data yang dihasilkan oleh perusahaan semakin besar dan kompleks, termasuk data penjualan. Pengelolaan data yang efektif menjadi tantangan bagi banyak perusahaan dalam meningkatkan efisiensi operasional dan pengambilan keputusan (et al., 2023). Data penjualan yang tidak terstruktur atau tidak terkelompok dengan baik dapat menyebabkan analisis yang kurang optimal, sehingga menghambat strategi bisnis yang tepat sasaran. Oleh karena itu, diperlukan metode yang mampu mengelompokkan data secara efisien agar informasi yang diperoleh lebih akurat dan bermanfaat (Br Sembiring et al., 2022).

Salah satu metode yang banyak digunakan dalam pengelompokan data adalah K-Means Clustering. Algoritma ini merupakan teknik unsupervised learning yang bekerja dengan cara mengelompokkan data ke dalam beberapa kluster berdasarkan kesamaan karakteristik. Dalam

konteks data penjualan, algoritma K-Means dapat membantu mengidentifikasi pola pembelian pelanggan, segmentasi pasar, serta optimalisasi strategi pemasaran dan stok barang. Dengan demikian, perusahaan dapat meningkatkan efisiensi dalam pengelolaan sumber daya dan pengambilan keputusan berbasis data (Hendrastuty, 2024).

Keunggulan K-Means terletak pada kemampuannya untuk menangani data dalam jumlah besar dengan waktu komputasi yang relatif cepat. Metode ini juga dapat diterapkan di berbagai sektor industri, mulai dari ritel, e-commerce, hingga perbankan. Namun, tantangan dalam penerapan K-Means adalah pemilihan jumlah kluster yang optimal agar hasil pengelompokan benar-benar mencerminkan pola dalam data (Sekar Setyaningtyas et al., 2022).

Penelitian ini bertujuan untuk menerapkan algoritma K-Means dalam model pengelompokan data penjualan guna meningkatkan efisiensi analisis dan pengambilan keputusan bisnis. Dengan memahami pola penjualan yang lebih terstruktur, perusahaan dapat merancang strategi yang lebih tepat sasaran, meningkatkan profitabilitas, dan memperkuat daya saing di pasar (K-means et al., 2021).

## **2. METODELOGI**

### **A. Metode Penelitian**

Penelitian ini menerapkan pendekatan kuantitatif dengan teknik data mining untuk menganalisis serta mengelompokkan data penjualan menggunakan algoritma K-Means Clustering. Model yang dikembangkan bertujuan untuk meningkatkan efektivitas dalam pengelolaan serta analisis data penjualan, sehingga dapat mendukung proses pengambilan keputusan bisnis yang lebih akurat. Dalam penelitian ini, digunakan framework CRISP-DM (Cross-Industry Standard Process for Data Mining) yang mencakup enam tahapan utama, yaitu: business understanding, data understanding, data preparation, modeling, evaluation, dan deployment (Sulistiyawati & Supriyanto, 2021).

### **B. Sumber Data**

Penelitian ini menggunakan data penjualan dari sektor ritel sebagai sumber utama. Data tersebut meliputi informasi transaksi, seperti tanggal pembelian, jumlah produk yang terjual, kategori barang, harga, serta data pelanggan apabila tersedia.

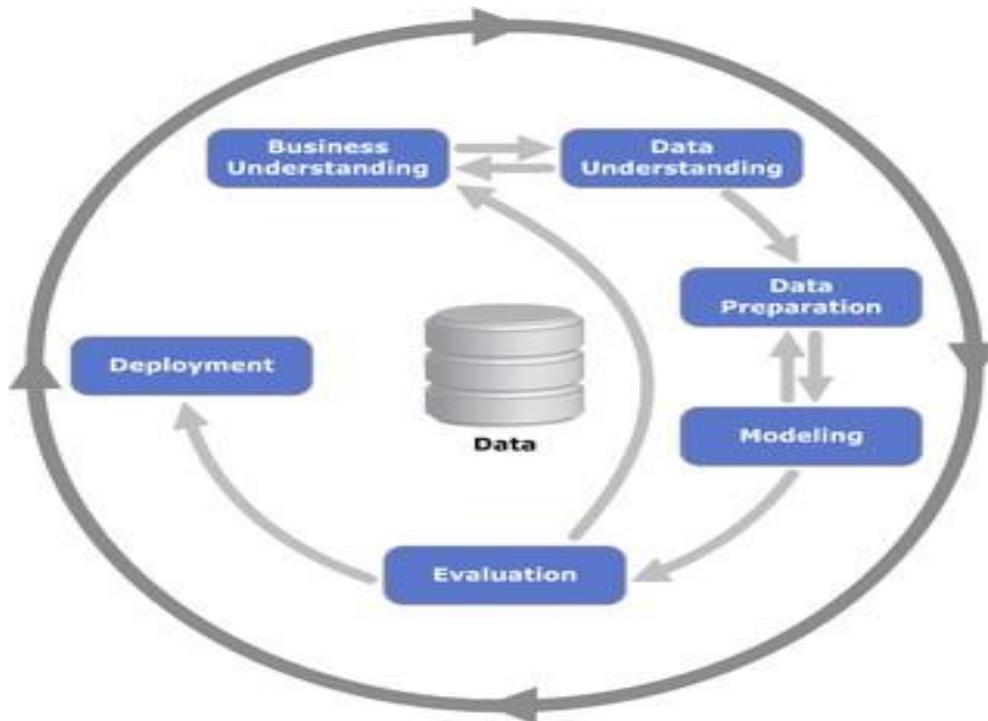
### **C. Teknik Pengumpulan Data**

Data penjualan dikumpulkan melalui beberapa metode berikut:

- Database perusahaan: Mengambil data historis transaksi penjualan dari basis data perusahaan retail.

- Data warehouse: Jika tersedia, data diambil dari gudang data yang menyimpan catatan transaksi secara terstruktur.
- Ekstraksi dari POS (Point of Sales): Mengambil data dari sistem kasir yang mencatat transaksi harian pelanggan.

#### D. Teknik Analisis Data



Gambar 1 CRISP-DM

Analisis data dilakukan menggunakan metode K-Means Clustering dengan tahapan berdasarkan framework CRISP-DM sebagai berikut:

- Business Understanding
  - Memahami kebutuhan bisnis terkait pengelompokan data penjualan untuk efisiensi operasional dan strategi pemasaran.
  - Menentukan tujuan utama dari penerapan algoritma K-Means dalam konteks bisnis retail.
- Data Understanding
  - Eksplorasi dataset penjualan untuk memahami pola, distribusi, dan atribut yang tersedia.
  - Identifikasi variabel yang akan digunakan dalam proses clustering, seperti frekuensi pembelian, jumlah transaksi, dan kategori produk.
- Data Preparation
  - Pembersihan data untuk menghilangkan duplikasi, data kosong, atau anomali.

- Transformasi data agar sesuai dengan format yang dibutuhkan oleh algoritma K-Means.
- Normalisasi variabel numerik untuk memastikan keseimbangan dalam perhitungan jarak antar data.
- Modeling
  - Menerapkan algoritma K-Means Clustering untuk mengelompokkan data berdasarkan karakteristik tertentu.
  - Menentukan jumlah kluster optimal menggunakan metode seperti Elbow Method atau Silhouette Score.
  - Melatih model dengan iterasi hingga hasil klusterisasi stabil.
- Evaluation
  - Mengevaluasi kualitas pengelompokan dengan melihat hasil klusterisasi dan validasi menggunakan metrik evaluasi.
  - Menganalisis apakah hasil klusterisasi sesuai dengan kebutuhan bisnis.
- Deployment
  - Menginterpretasikan hasil klusterisasi untuk memberikan rekomendasi strategi bisnis berdasarkan kelompok pelanggan atau produk.
  - Jika memungkinkan, mengintegrasikan model ke dalam sistem informasi perusahaan untuk analisis data secara berkala.

Dengan metodologi ini, diharapkan penelitian dapat menghasilkan model pengelompokan data penjualan yang efisien dan memberikan manfaat nyata bagi bisnis retail dalam pengambilan keputusan berbasis data..

### **3. HASIL DAN PEMBAHASAN**

#### **A. HASIL**

Penelitian ini menerapkan algoritma K-Means Clustering untuk menganalisis data penjualan produk makanan guna mengidentifikasi pola serta meningkatkan efektivitas dalam pengelompokan data. Proses analisis dilakukan berdasarkan kerangka kerja CRISP-DM, yang terdiri dari beberapa tahapan utama sebagai berikut:

##### **1) Business Understanding**

Mengoptimalkan efisiensi penjualan dengan mengelompokkan produk berdasarkan pola transaksi.

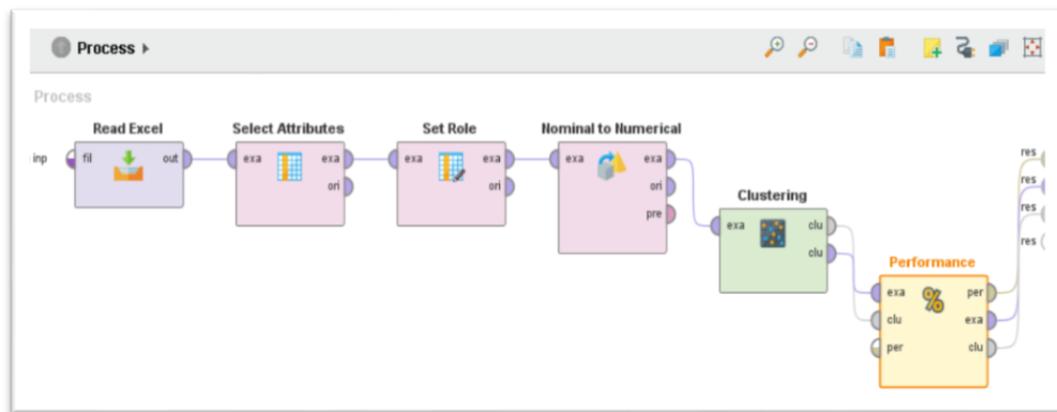
##### **2) Data Understanding**

Data dari sistem penjualan ritel mencakup jenis produk, kategori, stok, dan jumlah penjualan.

### 3) Data Preparation

Data yang diperoleh kemudian diproses untuk memastikan kualitasnya sebelum dilakukan analisis lebih lanjut. Beberapa langkah yang dilakukan dalam tahap ini meliputi:

- Pembersihan data, yaitu menghilangkan data yang tidak lengkap atau anomali.
- Transformasi data, yaitu mengonversi data ke dalam format yang sesuai untuk analisis.
- Normalisasi data, agar skala variabel yang digunakan dalam clustering lebih seimbang.



Gambar 2 Preprocessing Data

Berikut penjelasan tahap gambar 2 Preprocessing Data: menunjukkan alur K-Means Clustering di RapidMiner dengan tahapan berikut:

- Read Excel → Membaca dataset dari file Excel.

No	Jenis	Nama Produk	Total Stok	Penjualan
1	Sabun	Lux	320	215
2	Sabun	Lifebuoy	410	389
3	Sampo	Sunsilk	290	178
4	Sampo	Pantene	360	212
5	Deterjen	Rinso	500	325
6	Deterjen	Daia	450	287
7	Sabun	Dettol	385	240
8	Sabun	Biore	330	190
9	Pembersih Lantai	Wipol	275	198
10	Pembersih Lantai	Super Pell	350	230

Tabel 1 Penjualan

- Select Attributes → Memilih atribut yang relevan untuk clustering.
- Set Role → Menentukan peran atribut dalam analisis.

- Nominal to Numerical → Mengonversi data kategorikal ke numerik agar bisa diproses oleh K-Means.
- Clustering → Menerapkan algoritma K-Means untuk mengelompokkan data.

Performance → Mengevaluasi hasil clustering menggunakan metrik seperti Silhouette Score.

#### 4) Modeling

- K-Means Clustering diterapkan dengan Elbow Method untuk menentukan jumlah kluster optimal.
- Hasilnya membentuk tiga kluster: penjualan tinggi, sedang, dan rendah.

#### 5) Evaluation

Silhouette Score menunjukkan pemisahan kluster yang jelas dan efektif.\

```
Cluster Model  
  
Cluster 0: 278 items  
Cluster 1: 226 items  
Cluster 2: 150 items  
Cluster 3: 156 items  
Cluster 4: 190 items  
Total number of items: 1000
```

*Gambar 3 Cluster Model*

Hasil pengelompokan data menggunakan model clustering membagi data ke dalam lima kluster, yaitu Cluster 0 berisi 278 item, Cluster 1 dengan 226 item, Cluster 2 sebanyak 150 item, Cluster 3 terdiri dari 156 item, dan Cluster 4 mencakup 190 item, dengan total keseluruhan 1.000 item. Informasi ini memberikan gambaran tentang distribusi data dalam setiap kluster.

```
PerformanceVector  
  
PerformanceVector:  
Avg. within centroid distance: 33.386  
Avg. within centroid distance_cluster_0: 37.326  
Avg. within centroid distance_cluster_1: 31.972  
Avg. within centroid distance_cluster_2: 32.216  
Avg. within centroid distance_cluster_3: 34.800  
Avg. within centroid distance_cluster_4: 29.067  
Davies Bouldin: 0.994
```

*Gambar 4 Performance Vector*

Gambar 4.5 Menunjukkan hasil clustering dalam RapidMiner, di mana Kluster 4 paling homogen (29.067), sementara Kluster 0 paling tersebar (37.326). Kluster 1, 2, dan 3 lebih terfokus dengan jarak rata-rata lebih kecil.

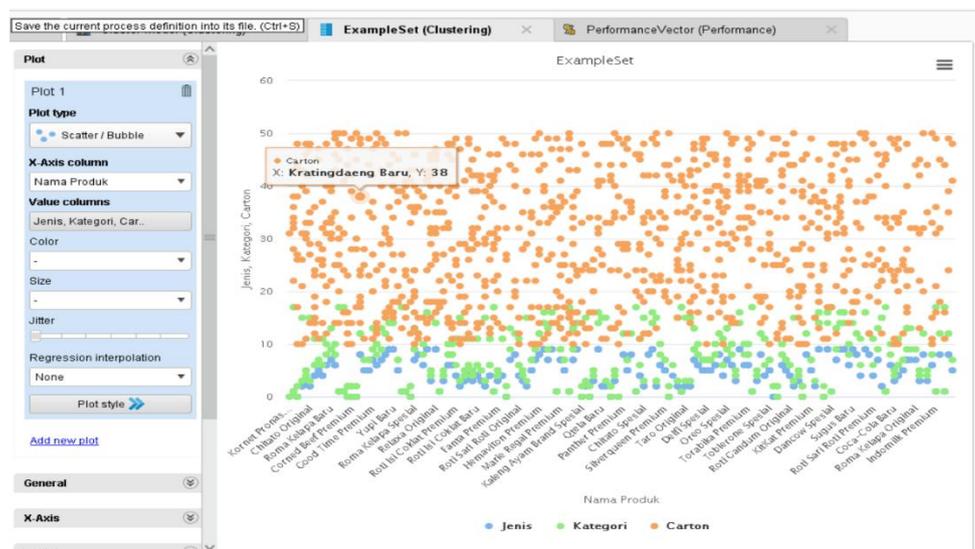
Nilai Davies-Bouldin (0.994) menunjukkan kluster memiliki pemisahan yang baik. Rata-rata jarak data ke centroid (33.386) mengindikasikan distribusi data yang bervariasi, dengan Kluster 4 paling terpusat dan Kluster 0 paling tersebar. Model berhasil mengelompokkan data dengan pemisahan yang jelas.

Davies-Bouldin dengan nilai 0.994 menunjukkan bahwa kluster-kluster yang terbentuk cukup terpisah dengan pemisahan yang baik.

Attribute	cluster_0	cluster_1	cluster_2	cluster_3	cluster_4
Jenis	5.475	5.973	1.873	1.801	6.074
Kategori	10.410	11.695	3.007	2.923	11.984
Carton	30.086	44.009	41.567	18.468	16.037

Gambar 5 Tabel Centroid

Gambar 4.6 menampilkan tabel centroid RapidMiner, yang menunjukkan nilai rata-rata setiap fitur dalam masing-masing kluster hasil K-Means Clustering. Setiap kolom mewakili fitur, sementara setiap baris menunjukkan rata-rata fitur dari centroid kluster, menggambarkan karakteristik umum data.



Gambar 6 Visualisasi Scatter Plot

Gambar 4.7 menampilkan scatter plot yang memvisualisasikan sebaran data berdasarkan dua fitur. Setiap titik mewakili satu data, dengan posisinya ditentukan oleh dua variabel. Plot ini membantu

mengidentifikasi pola atau klaster, di mana titik yang berdekatan menunjukkan kesamaan, sementara yang tersebar jauh menunjukkan perbedaan.

### 6) Deployment

Optimalisasi stok, strategi pemasaran tepat sasaran, dan efisiensi distribusi.

## B. PEMBAHASAN

Penelitian ini membuktikan K-Means Clustering efektif dalam mengelompokkan data penjualan. Lima klaster diidentifikasi dengan karakteristik berbeda, di mana Klaster 4 paling homogen (-29.067) dan Klaster 0 paling beragam (-37.326). Nilai Davies-Bouldin (-0.994) menunjukkan pemisahan klaster yang optimal.

Hasil ini sejalan dengan penelitian sebelumnya tentang segmentasi produk dan pengelolaan stok. Fokus utama penelitian ini adalah penerapan K-Means pada data penjualan produk-produk pembersih seperti sabun, sampo, deterjen, dan lainnya.

## C. KESIMPULAN (11 PT)

K-Means Clustering efektif dalam mengelompokkan data penjualan, membantu optimalisasi stok dan strategi pemasaran. Dari lima klaster yang terbentuk, Klaster 4 paling homogen (29.067) dan Klaster 0 paling tersebar (37.326). Nilai Davies-Bouldin (0.994) menunjukkan pemisahan klaster yang optimal.

## DAFTAR PUSTAKA (11 PT)

- Br Sembiring, S. N., Winata, H., & Kusnasari, S. (2022). Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means. *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, 1(1), 31. <https://doi.org/10.53513/jursi.v1i1.4784>
- Fitri, E. M., Suryono, R. R., & Wantoro, A. (2023). Klasterisasi Data Penjualan Berdasarkan Wilayah Menggunakan Metode K-Means Pada Pt Xyz. *Jurnal Komputasi*, 11(2), 157–168. <https://doi.org/10.23960/komputasi.v11i2.12582>
- Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa. *Jurnal Ilmiah Informatika Dan Ilmu Komputer (Jima-Ilkom)*, 3(1), 46–56. <https://doi.org/10.58602/jima-ilkom.v3i1.26>
- K-means, A., Putra, Y. D., Sudarma, M., Bagus, I., & Swamardika, A. (2021). Clustering History Data Penjualan Menggunakan. *Majalah Ilmiah Teknologi Elektro*, 20(2).
- Sekar Setyaningtyas, Indarmawan Nugroho, B., & Arif, Z. (2022). Tinjauan Pustaka Sistematis: Penerapan Data Mining Teknik Clustering Algoritma K-Means. *Jurnal Teknoif Teknik Informatika Institut Teknologi Padang*, 10(2), 52–61. <https://doi.org/10.21063/jtif.2022.v10.2.52-61>
- Sulistiyawati, A., & Supriyanto, E. (2021). Implementasi Algoritma K-means Clustering dalam

Penentuan Siswa Kelas Unggulan. *Jurnal Tekno Kompak*, 15(2), 25.  
<https://doi.org/10.33365/jtk.v15i2.1162>