



## Clustering Data Penjualan Toko XYZ Menggunakan Metode K-Means

Putri Margaretta<sup>1</sup>, Betha Nurina Sari<sup>2</sup>, Azhari Ali Ridha<sup>3</sup>

Universitas Singaperbangsa Karawang

Received: 14 November 2024  
Revised : 20 November 2024  
Accepted: 27 November 2024

### Abstract

*Sales, A Process Involving Sellers Offering Goods Or Services To Buyers With The Objective Of Profiting From The Transaction, Is A Pivotal Activity In Business. Xyz Store, A Micro, Small, And Medium Enterprise (Msme) Specializing In Children's Clothing Sales, Has Been Operational Since 2009. Transactions Can Be Conducted Online, Enabling Buyers To Shop Without The Need To Visit The Physical Store. However, Xyz Store Faces A Challenge In Stock Management, Specifically The Mismatch Between Demand And Product Availability, Leading To An Accumulation Of Less Popular Items. By Understanding Sales Trends, Xyz Store Can Optimize Their Stock Management, Either By Curtailing The Purchase Of Stocks For Less Sold Items Or By Substituting Less Popular Items With New Ones That May Be More Appealing To Potential Buyers. Through The Evaluation Stage In The Kdd Method, It Was Determined That The Optimal Number Of Clusters In This Study Is Three, With An Evaluation Result Of 0.5063460425226173. These Three Clusters Were Identified As Less Popular Items, Moderately Popular Items, And Very Popular Items. Cluster 1, Deemed Less Popular, Comprises 84 Items. Cluster 2, Which Is Moderately Popular, Includes 12 Items. Meanwhile, Cluster 3, Identified As Very Popular, Contains Only 4 Items. This Study Provides Valuable Insights Into Sales Strategies And Stock Management At Xyz Store For Enhancing Efficiency And Sales.*

**Keywords:** *K-Means, Knowledge Discovery In Database, Data Mining, Silhouette Coefficient*

(\*) Corresponding Author: [putrimargaretta@gmail.com](mailto:putrimargaretta@gmail.com)

**How To Cite:** Margaretta, P., Sari, B. N., & Ridha, A. A. (2024). Clustering Data Penjualan Toko XYZ Menggunakan Metode K-Means. <https://doi.org/10.5281/zenodo.14586668>

## PENDAHULUAN

Penjualan Adalah Proses Yang Melibatkan Pihak Penjual Yang Menawarkan Barang Atau Jasa Kepada Pembeli Dengan Tujuan Mendapatkan Keuntungan Dari Transaksi Yang Terjadi. Penjualan Juga Dapat Dipahami Sebagai Kegiatan Yang Dijalankan Oleh Penjual Untuk Menjual Barang Atau Jasa Yang Diinginkan Oleh Pembeli Dengan Ekspektasi Mendapat Laba Dari Transaksi-Transaksi Tersebut. Penjualan Berarti Adanya Pemindahan Hak Milik Atas Barang Atau Jasa Dari Penjual Ke Pembeli. (Indra Et Al., 2022).

Bisnis *Online*, Yang Dikenal Sebagai *E-Commerce*, Dapat Diartikan Sebagai Transaksi Jual Beli Melalui Media Internet. Dengan *E-Commerce*, Belanja Dapat Dilakukan Kapan Saja Dari Rumah Melalui Saluran Belanja *Online*. Selain Itu, Transaksi *Online* Dapat Memotong Melalui Saluran Distribusi, Menghemat Waktu Dan Biaya, Serta Mengurangi Kendala Transaksi Yang Dialami Saat Berbelanja Secara Tradisional. (Yahya & Sugiyanto, 2020).

Analisis *Cluster* Merupakan Teknik Yang Digunakan Untuk Mengkategorikan Objek Data Ke Dalam Kelompok Atau *Cluster* Berdasarkan Kesamaan Karakteristik. Jika Objek Data Memiliki Karakteristik Yang Berbeda,

Maka Mereka Akan Ditempatkan Dalam *Cluster* Yang Berbeda. Teknik Ini Termasuk Dalam Teknik Multivariat Dan Proses Analisisnya Dilakukan Oleh Algoritma Pengelompokan, Bukan Oleh Manusia. Tujuan Utama Dari Analisis *Cluster* Adalah Untuk Mengklasifikasikan Data Berdasarkan Karakteristiknya, Sehingga Data Yang Memiliki Kemiripan Paling Dekat Dengan Data Lainnya Akan Dikelompokkan Dalam Satu *Cluster*. Dengan Demikian, Analisis Ini Sangat Berguna Dalam Menemukan Kelompok-Kelompok Yang Sebelumnya Tidak Diketahui Dalam Data. (Nagari & Inayati, 2020).

Studi Yang Dilakukan Oleh (Paembonan & Abduh, 2021) Menghasilkan Temuan Bahwa Jumlah *Cluster* Optimal Dan Nilai *Silhouette* Maksimum Untuk Data Obat Diperoleh Saat Jumlah *Cluster* (K) Adalah 2 Dengan Nilai *Silhouette* Sebesar 0,4854. Studi Tersebut Juga Menunjukkan Bahwa Nilai *Silhouette* Cenderung Menurun Seiring Dengan Peningkatan Jumlah *Cluster*. Sementara Itu, Dalam Penelitian Berjudul “Clustering Data Penjualan Pada Toko Perlengkapan Outdoor Menggunakan Metode *K-Means*” Oleh (Indriyani & Irfiani, 2019), Mereka Berhasil Melakukan Pengelompokan Pada 30 Data Master Menggunakan Metode *K-Means* Yang Menghasilkan Pembagian Data Menjadi 3 *Cluster*.

Toko Xyz Adalah Sebuah Usaha Mikro Kecil Dan Menengah (Ukm) Yang Bergerak Di Bidang Penjualan Pakaian Anak. Usaha Ini Telah Berdiri Sejak Tahun 2009. Pembeli Dapat Melakukan Transaksi Jual Beli Produk Toko Xyz Secara *Online* Tanpa Perlu Mengunjungi Toko Fisik Mereka. Namun, Ada Aspek Yang Perlu Diperhatikan Oleh Toko Xyz, Yaitu Manajemen Stok Yang Kurang Detail. Hal Ini Menyebabkan Penumpukan Barang Yang Kurang Laku. Dengan Memahami Tren Penjualan, Toko Xyz Dapat Mengoptimalkan Manajemen Stok Mereka Dengan Mengurangi Pembelian Stok Untuk Barang Yang Jarang Terjual, Atau Bahkan Mengganti Produk Yang Kurang Laku Dengan Produk Baru Yang Mungkin Lebih Disukai Oleh Calon Pembeli.

Mengacu Pada Uraian Dan Perbandingan Metode Dari Berbagai Penelitian Yang Telah Disebutkan Sebelumnya, Penelitian Ini Berfokus Pada '*Clustering* Data Penjualan Toko Xyz Menggunakan Metode *K-Means*'. Penelitian Ini Akan Menerapkan Teknik *Clustering* Pada Produk Penjualan Toko Xyz. Melalui Penelitian Ini, Toko Xyz Dapat Mengetahui Produk Mana Yang Paling Banyak Diminati Dan Produk Mana Yang Kurang Diminati Oleh Pelanggan.

## **KAJIAN TEORI**

### **Data Mining**

Data Mining Merupakan Proses Ekstraksi Dan Penemuan Pola Dalam Kumpulan Data Besar, Yang Merupakan Gabungan Dari Beberapa Disiplin Dalam Ilmu Komputer. Proses Ini Melibatkan Metode *Intersection* Yang Digunakan Dalam *Machine Learning*, *Statistik*, Dan Sistem *Database*. Sebagai Sub-Bidang Dalam Ilmu Komputer Dan *Statistik*, Tujuan Dari *Data Mining* Adalah Untuk Mengekstrak Informasi Menggunakan Metode Cerdas Dari Kumpulan Data Dan Mengubahnya Menjadi Informasi Yang Lebih Terstruktur Dan Mudah Dipahami Untuk Keperluan Lebih Lanjut. (Suyanto & Others, 2017).

### **Clustering**

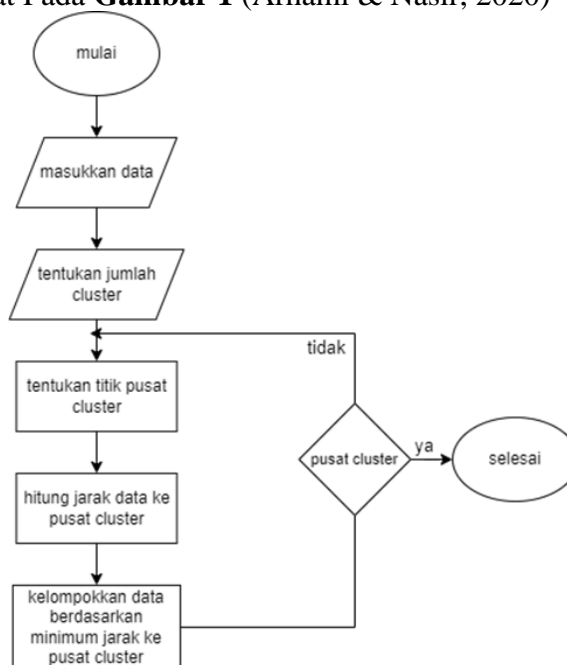
*Clustering* Merupakan Metode *Unsupervised Learning* Yang Bertujuan Untuk Mengelompokkan Data Tanpa Label Kelas Ke Dalam Beberapa Kelompok Atau

*Cluster*. Dalam Proses Ini, Data Dalam Satu *Cluster* Memiliki Tingkat Kemiripan Yang Maksimum, Sementara Antar *Cluster* Memiliki Tingkat Kemiripan Yang Minimum. Oleh Karena Itu, *Clustering* Memungkinkan Prediksi Data Yang Belum Memiliki Label Kelas. (Suyanto & Others, 2017).

*Clustering* Memiliki Peran Penting Dalam Prediksi Dan Analisis Berbagai Isu Bisnis, Seperti Segmentasi Pasar Dalam Pemasaran Atau Pemetaan Zonasi Wilayah. Lebih Jauh, *Clustering* Juga Mampu Mengidentifikasi Objek Dalam Berbagai Bidang, Termasuk Computer Vision Dan Pengolahan Gambar. Hasil Optimal Dari Proses *Clustering* Ditandai Dengan Tingkat Kesamaan Yang Tinggi Di Dalam Satu Kelas Dan Tingkat Kesamaan Yang Rendah Antara Kelas-Kelas.

**Algoritma K-Means**

*K-Means* Adalah Algoritma Pengelompokan Yang Masuk Dalam Kategori Pembelajaran Tanpa Pengawasan. Algoritma Ini Sering Digunakan Dalam Proses Pengelompokan Data Berdasarkan Kesamaan Atribut Atau Fitur Yang Relevan, Yang Dapat Dilihat Pada **Gambar 1** (Arhami & Nasir, 2020)



**Gambar 1** Algoritma *K-Means*

Dalam Lingkup *Data Science* Dan *Machine Learning*, Teknik Pengelompokan *K-Means* Dikenal Sebagai Salah Satu Teknik *Unsupervised* Yang Kuat Untuk Mengidentifikasi Struktur Dari Dataset Yang Diberikan. Algoritma *Clustering* Ini Menjadi Pilihan Terbaik Untuk Memisahkan Data Ke Dalam Kelompok-Kelompok Dan Sangat Sering Digunakan Karena Kesederhanaannya (Zubair Et Al., 2022).

Dalam Artikel Penelitian Oleh Zhao (2022) Algoritma *K-Means* Memiliki Sejumlah Kelebihan, Termasuk Pemikiran Yang Jelas, Konvergensi Yang Cepat, Dan Implementasi Yang Mudah. Akan Tetapi, Pemilihan Centroid Awal Sangat Mempengaruhi Efisiensi Operasional Algoritma. Pemilihan Centroid Secara Acak Bisa Menghasilkan Banyak Iterasi Atau Hanya Mencapai Keadaan Optimal Lokal. Selain Itu, Algoritma Ini Memerlukan Pengguna Untuk Menentukan Jumlah

*Cluster* K. Perbedaan Nilai K Akan Berdampak Besar Pada Hasil Pengelompokan, Yang Pada Akhirnya Mempengaruhi Akurasi Hasil Pengelompokan.

### **Silhouette Coefficient**

Penelitian Yang Dilakukan Oleh Yunistya Et Al. (2022) Mengatakan Bahwa Metode *Silhouette Coefficient* (Sc) Adalah Salah Satu Pendekatan Evaluasi Yang Bisa Digunakan Untuk Menilai Kualitas Dan Kekuatan *Cluster*. Metode Ini Merupakan Teknik Validasi *Cluster* Yang Mengintegrasikan Metode *Cohesion* Dan *Separation*. Nilai Sc Yang Dihasilkan Dari Perhitungan Ini Dapat Berkisar Antara -1 Hingga 1. Nilai Sc = 1 Menunjukkan Bahwa Objek Sudah Berada Di *Cluster* Yang Sesuai. Jika Nilai Sc = 0, Berarti Objek Berada Di Antara Dua *Cluster*. Sedangkan, Jika Sc = -1, Ini Menunjukkan Bahwa Objek Seharusnya Berada Di *Cluster* Lain Yang Lebih Tepat.

## **METODOLOGI PENELITIAN**

Dalam Penelitian Ini, Metodologi Yang Digunakan Adalah *Knowledge Discovery In Database* (Kdd). Kdd Adalah Proses Yang Digunakan Untuk Menemukan Pola Yang Berguna Dan Berarti Dari Data Besar. Metodologi Kdd Terdiri Dari Enam Tahap Utama, Yaitu:

### 1) Data Selection

Tahap Ini Melibatkan Proses Seleksi Atau Pemilihan Data Dari Kumpulan Data Yang Diperoleh Dari Toko Xyz. Mengingat Volume Data Yang Tersedia, Tidak Semua Data Akan Diproses. Hanya Data Yang Relevan Dengan Kebutuhan Penelitian, Yaitu Data Penjualan Yang Sama, Yang Akan Diproses. Oleh Karena Itu, Target Data Ditentukan. Dalam Konteks Seleksi Data Ini, Penelitian Ini Fokus Pada Item Yang Sama Yang Dijual Selama Periode Penjualan 2019-2021. Ada Empat Atribut Yang Diperlukan Dalam Target Data, Yaitu Nama Produk, Dan Jumlah Penjualan Pada Tahun 2019, 2020, Dan 2021.

### 2) Data Pre-Processing

Mengikuti Tahap Seleksi Data, Tahap Data Pre-Processing Dijalankan. Dalam Tahap Ini, Atribut Diorganisir Berdasarkan Jenis Datanya. Tujuan Dari Tahap Ini Adalah Untuk Memfasilitasi Proses Transformasi Data Pada Tahap Berikutnya.

### 3) Data Transformation

Pada Tahap Data Transformation, Atribut Yang Sebelumnya Telah Dipisahkan Berdasarkan Jenis Datanya Kemudian Ditransformasi. Atribut Penjualan Pada Tahun 2019, 2020, Dan 2021 Ditransformasi Menjadi *Numeric*, Sementara Atribut Nama Barang Ditransformasi Menjadi *Categorical*.

### 4) Data Mining

Setelah Tahap Transformasi Data, Metode Elbow Diaplikasikan Untuk Mencari Jumlah *Cluster* Atau 'K' Yang Optimal. Metode Elbow Ini Beroperasi Dengan Cara Memvisualisasikan Hasil Dari Berbagai Nilai 'K' Dan Mencari 'Elbow' Atau Titik Belok Dalam Visualisasi Tersebut. Nilai 'K' Pada Titik 'Elbow' Ini Dianggap Sebagai Jumlah *Cluster* Yang Optimal. Setelah Menemukan 'K' Yang Optimal, Data Kemudian Diproses Menggunakan Algoritma *K-Means* Untuk Melakukan Proses *Clustering*. Algoritma *K-Means* Ini Berfungsi Dengan Membagi Data Menjadi 'K' *Cluster* Yang Berbeda, Di Mana Setiap Data Dalam Suatu *Cluster* Memiliki Jarak Terdekat Dengan Centroid Dari *Cluster* Tersebut.

### 5) Evaluation

Setelah Proses *Data Mining*, Tahap Evaluasi Dilakukan. Di Tahap Ini, *Silhouette Coefficient* Digunakan Untuk Menghasilkan Skor *Silhouette*. Skor Ini Akan Digunakan Untuk Mengevaluasi Kualitas Dari Hasil *Clustering*.

### 6) Knowledge Representation

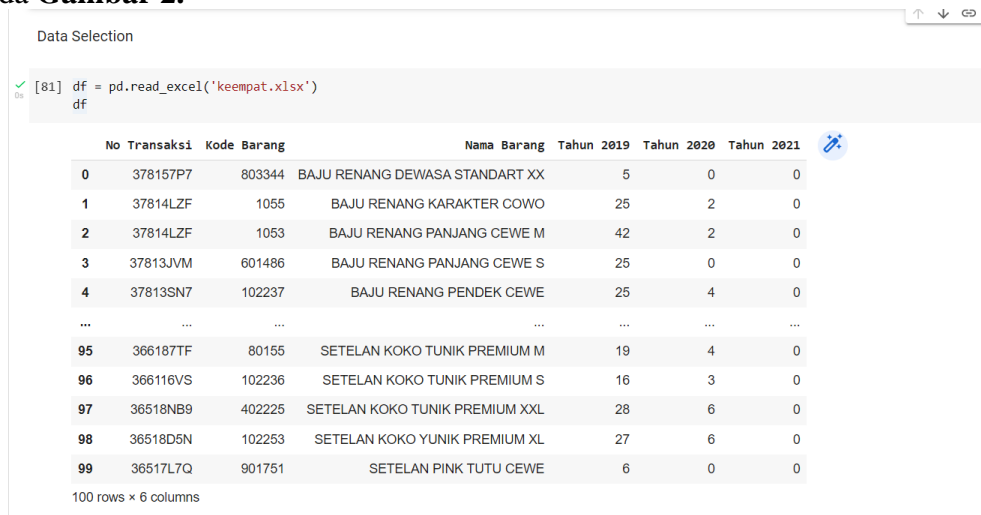
Knowledge Representation Merupakan Tahap Akhir, Dimana Setelah Semua Langkah Atau Tahap Telah Diselesaikan Dan Hasil Akhir Dari Proses *Data Mining* Telah Dievaluasi, Maka Dapat Dilanjutkan Dengan Penyusunan Laporan Atau Presentasi Dari Hasil Tersebut. Hasil Akhir Ini Kemudian Diserahkan Kepada Pemilik Toko Xyz Untuk Dapat Diimplementasikan Selanjutnya.

## HASIL & PEMBAHASAN

Penelitian Ini Menghasilkan Implementasi Metode Pengelompokan Dengan Memanfaatkan Algoritma *K-Means*. Tujuan Implementasi Ini Adalah Untuk Memperbaiki Proses Pengelompokan Penjualan Di Toko Xyz, Yang Dapat Berkontribusi Dalam Pengawasan Stok Produk Dan Penerapan Strategi Penjualan Yang Lebih Baik Dengan Menggunakan *Knowledge Discovery In Database* (Kdd) Sebagai Metodologi Penelitian. Dengan Mengadopsi Metodologi Kdd, Penelitian Ini Berusaha Untuk Mengekstrak Pengetahuan Baru Dan Relevan Dari Kumpulan Data Penjualan, Yang Selanjutnya Dapat Digunakan Untuk Mendukung Proses Pengambilan Keputusan Dalam Konteks Penjualan Di Toko Xyz.

### 1. Data Selection

Pada Tahap Ini, Penentuan Dataset Atau Subset Data Yang Akan Dianalisis Dilakukan. Dataset Penjualan Dari Toko Xyz Menjadi Titik Sentral Analisis. Dataset Ini Mencakup Berbagai Variabel Yang Relevan Untuk Penjualan, Seperti 'Nama Barang', 'Tahun 2019', 'Tahun 2020', Dan 'Tahun 2021' Yang Dapat Dilihat Pada **Gambar 2**.



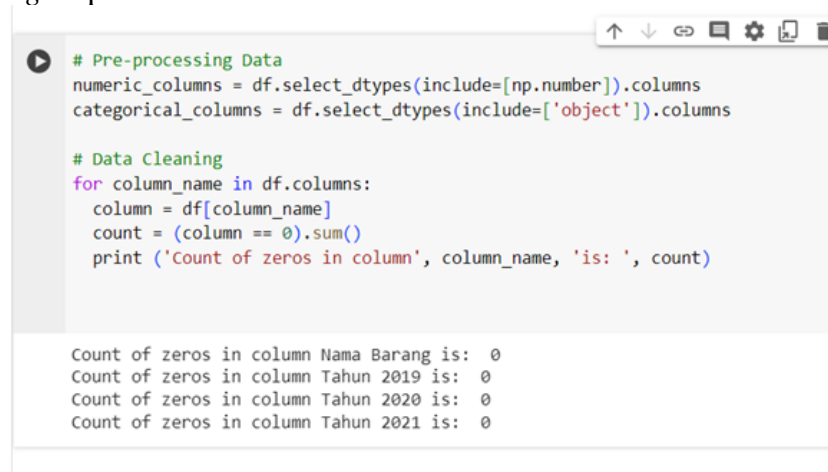
| No Transaksi | Kode Barang | Nama Barang                           | Tahun 2019 | Tahun 2020 | Tahun 2021 |
|--------------|-------------|---------------------------------------|------------|------------|------------|
| 0            | 378157P7    | 803344 BAJU RENANG DEWASA STANDART XX | 5          | 0          | 0          |
| 1            | 37814LZF    | 1055 BAJU RENANG KARAKTER COWO        | 25         | 2          | 0          |
| 2            | 37814LZF    | 1053 BAJU RENANG PANJANG CEWE M       | 42         | 2          | 0          |
| 3            | 37813JVM    | 601486 BAJU RENANG PANJANG CEWE S     | 25         | 0          | 0          |
| 4            | 37813SN7    | 102237 BAJU RENANG PENDEK CEWE        | 25         | 4          | 0          |
| ...          | ...         | ...                                   | ...        | ...        | ...        |
| 95           | 366187TF    | 80155 SETELAN KOKO TUNIK PREMIUM M    | 19         | 4          | 0          |
| 96           | 366116VS    | 102236 SETELAN KOKO TUNIK PREMIUM S   | 16         | 3          | 0          |
| 97           | 36518NB9    | 402225 SETELAN KOKO TUNIK PREMIUM XXL | 28         | 6          | 0          |
| 98           | 36518D5N    | 102253 SETELAN KOKO YUNIK PREMIUM XL  | 27         | 6          | 0          |
| 99           | 36517L7Q    | 901751 SETELAN PINK TUTU CEWE         | 6          | 0          | 0          |

**Gambar 2** Dataset Penjualan

### 2. Data Preprocessing

Setelah Seleksi Data, Tahap Kedua Dalam *Knowledge Discovery In Database* Adalah Melakukan *Preprocessing* Data, Yang Melibatkan Pembersihan Data, Transformasi Data, Dan Pemilihan Fitur. Dalam Penelitian Ini, Tahap *Preprocessing* Melibatkan Pemisahan Variabel Berdasarkan Tipe Data Mereka

Untuk Mengoptimalkan Proses *Clustering*. Variabel Dengan Kolom Tipe Data ‘Object’ Dipisahkan Menjadi Kolom Kategorikal, Sementara Variabel Dengan Kolom Tipe Data ‘Numerical’ Dipisahkan Menjadi Kolom Numerik. Proses Pembersihan Data Kemudian Dilakukan Pada Data Missing Value. Hasil Data *Preprocessing* Dapat Dilihat Pada **Gambar 3**.



```

# Pre-processing Data
numeric_columns = df.select_dtypes(include=[np.number]).columns
categorical_columns = df.select_dtypes(include=['object']).columns

# Data Cleaning
for column_name in df.columns:
    column = df[column_name]
    count = (column == 0).sum()
    print ('Count of zeros in column', column_name, 'is: ', count)

```

Count of zeros in column Nama Barang is: 0  
 Count of zeros in column Tahun 2019 is: 0  
 Count of zeros in column Tahun 2020 is: 0  
 Count of zeros in column Tahun 2021 is: 0

**Gambar 3** Hasil *Data Preprocessing*

### 3. Data Transformation

*Data Transformation* Bertujuan Untuk Mengubah Dataset Menjadi Format Yang Lebih Mudah Diolah Untuk Tahapan Berikutnya. Data Yang Telah Diproses Pada Tahap *Preprocessing* Diubah Menjadi Format Yang Paling Sesuai Untuk Algoritma *K-Means*. Fungsi Yang Digunakan Dalam Tahap Ini Dapat Dilihat Pada **Gambar 4**.



```

from sklearn.preprocessing import StandardScaler, OneHotEncoder
from sklearn.compose import make_column_transformer

# Create a preprocessor
preprocessor = make_column_transformer(
    (StandardScaler(), numeric_columns),
    (OneHotEncoder(), categorical_columns)
)

```

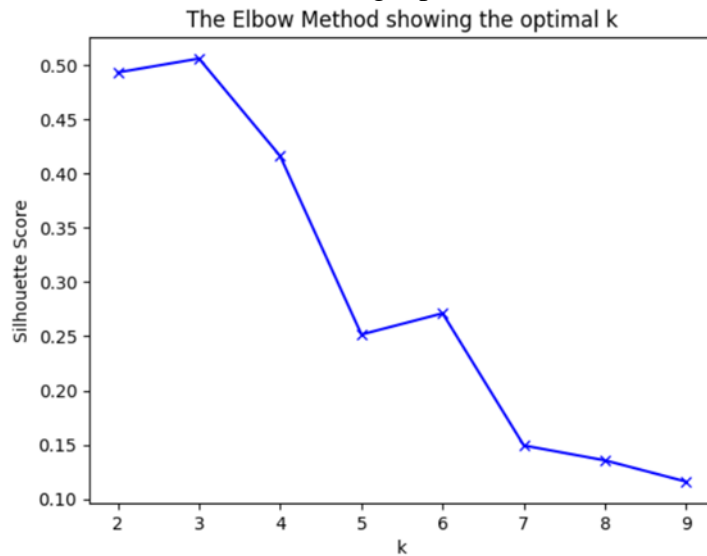
**Gambar 4** *Data Transformation*

Dalam Konteks Ini, Fungsi ‘*Make\_Column\_Transformer*’ Digunakan, Yang Menciptakan Estimator Yang Menerapkan Transformer Yang Berbeda Ke Kolom Yang Berbeda Dari Array Atau Pandas Dataframe. Dalam Hal Ini, *StandardScaler()* Diterapkan Ke Kolom Numerik Dan *Onehotencoder()* Diterapkan Ke Kolom Kategorikal.

### 4. Data Mining

Tahap Selanjutnya Dalam *Knowledge Discovery In Database* Adalah *Data Mining*, Di Mana Pada Tahap Ini Proses Algoritma Yang Telah Ditentukan Sebelumnya Dijalankan. Pada Tahap Ini, Proses *Data Mining* Menggunakan Algoritma *K-Means* Dilakukan Untuk Mendapatkan Hasil *Clustering*. Untuk Menentukan Jumlah *Cluster* Yang Optimal, Pencarian Dilakukan Menggunakan

Metode Elbow, Dan Hasil Grafik Elbow Yang Dapat Dilihat Pada **Gambar** Menunjukkan Bahwa Jumlah *Cluster* Yang Optimal Adalah 3 *Cluster*.



**Gambar 5** Hasil Grafik Elbow

Setelah Menemukan Jumlah *Cluster* Yang Optimal, Proses *Data Mining* Dilanjutkan Dengan Menggunakan *Pipeline* Dari Library Sklearn. *Pipeline* Adalah Metode Untuk Menggabungkan Beberapa Langkah Pemrosesan Data Dan Pemodelan Menjadi Satu Proses Yang Dapat Dilihat Pada **Gambar 6**.



**Gambar 6** Pipeline

Berikut Adalah Detail Penggunaan *Pipeline*:

- i. Columntransformer: Transformer Ini Melakukan Operasi Yang Berbeda-Beda Pada Kolom Yang Berbeda Dari Dataframe. Dalam Konteks Ini, Ada Dua Transformer:
  - StandardScaler: Ini Digunakan Untuk Melakukan Standarisasi Fitur Dengan Menghilangkan Mean Dan Menskalakan Ke Varians Unit. Dalam Kasus Ini, Diterapkan Pada Kolom 'Tahun 2019', 'Tahun 2020', Dan 'Tahun 2021'.
  - Onehotencoder: Ini Digunakan Untuk Mengubah Variabel Kategori Menjadi Format Yang Dapat Dipahami Oleh Algoritma Machine Learning. Dalam Kasus Ini, Diterapkan Pada Kolom 'Nama Barang'.
- ii. Kmeans: Ini Adalah Algoritma *K-Means* Yang Digunakan Untuk Melakukan *Clustering*. Parameter Yang Digunakan Adalah:

- $N\_Clusters=3$ : Ini Menunjukkan Bahwa Data Akan Dikelompokkan Menjadi 3 *Cluster* Sesuai Dengan Hasil Metode Elbow.
- $N\_Init=10$ : Ini Menunjukkan Bahwa Algoritma *K-Means* Akan Dijalankan 10 Kali Dengan Centroid Awal Yang Berbeda.
- $Random\_State=0$ : Ini Digunakan Untuk Menjaga Konsistensi Hasil Ketika Menjalankan Kode Ini Beberapa Kali. Dengan Menetapkan  $Random\_State$  Ke Suatu Nilai Tetap (Misalnya 0), Maka Akan Mendapatkan Hasil Yang Sama Setiap Kali Kode Ini Dijalankan.

Jadi, Secara Keseluruhan, *Pipeline* Ini Pertama-Tama Melakukan Pra-Pemrosesan Pada Data (Standarisasi Dan Encoding One-Hot), Dan Kemudian Menerapkan Algoritma *K-Means* Pada Data Yang Telah Diproses.

### 5. Evaluasi

Tahap Evaluasi Menggunakan *Silhouette Coefficient*. *Silhouette Score* Yang Diperoleh Adalah Sekitar 0.506, Yang Menandakan Bahwa Pengelompokan Telah Dilakukan Dengan Baik. Nilai Yang Hampir Mencapai 1 Menunjukkan Pengelompokan Yang Sangat Baik, Nilai Yang Hampir Mencapai 0 Menunjukkan Bahwa Titik Data Berada Di Batas Keputusan Antara Dua *Cluster* Yang Berdekatan Atau Sangat Dekat, Dan Nilai Yang Hampir Mencapai -1 Menunjukkan Bahwa Titik Data Telah Ditempatkan Di *Cluster* Yang Tidak Tepat. Hasil Evaluasi Dapat Dilihat Pada **Gambar 7**.



```

from sklearn.metrics import silhouette_score

# Get the cluster labels
labels = pipeline.named_steps['kmeans'].labels_

# Calculate silhouette score
score = silhouette_score(preprocessor.transform(df), labels)

# Print silhouette score
print(f'Silhouette score: {score}')

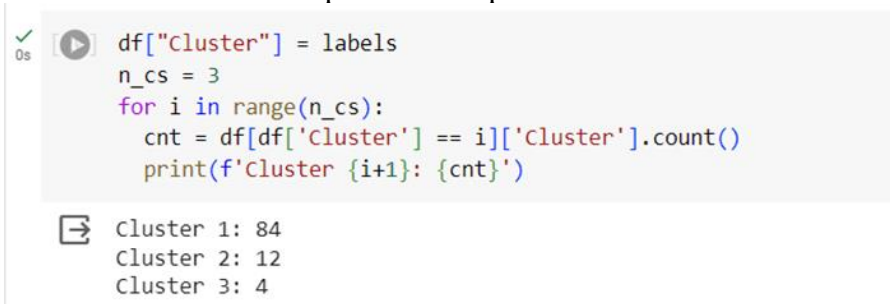
```

Silhouette score: 0.5063460425226173

**Gambar 7** Hasil *Silhouette Coefficient*

### 6. Knowledge Representation

Setelah Mendapatkan Hasil Pengelompokan Berdasarkan Tahap-Tahap Sebelumnya, Kemudian Mengelompokkan Setiap Item Sesuai Dengan *Clusternya*. Label *Cluster* Ditambahkan Ke Dataframe Df Dengan Menggunakan Variabel Labels. Jumlah Item Dalam Tiap *Cluster* Dapat Dilihat Pada **Gambar 8**.



```

df["Cluster"] = labels
n_cs = 3
for i in range(n_cs):
    cnt = df[df['Cluster'] == i]['Cluster'].count()
    print(f'Cluster {i+1}: {cnt}')

```

Cluster 1: 84  
Cluster 2: 12  
Cluster 3: 4

**Gambar 8** Jumlah Item Dalam Tiap *Cluster*

Selanjutnya, Jumlah *Cluster* Ditentukan Dan Disimpan Dalam Variabel *N\_Cs*, Yang Dalam Hal Ini Adalah Tiga. Untuk Setiap *Cluster*, Dihitunglah Jumlah Sampel Dalam Dataframe *Df* Di Mana Nilai Kolom 'Cluster' Sama Dengan Indeks *Cluster*. Hasil Perhitungan Tersebut Kemudian Dicitak Untuk Memberikan Gambaran Tentang Distribusi Sampel Dalam Setiap *Cluster*. Dalam Proses Ini,  $I+1$  Digunakan Untuk Mencetak Nomor *Cluster* Yang Dimulai Dari 1, Bukan 0, Untuk Memudahkan Pemahaman Dan Sesuai Dengan Konvensi Penomoran Yang Biasa Digunakan Dalam Penelitian Dan Laporan.

Melalui Tahapan Evaluasi Dalam Metode *Kdd*, Dapat Diketahui Bahwa Jumlah *Cluster* Yang Optimal Dalam Penelitian Ini Adalah Tiga *Cluster* Dengan Hasil Evaluasi 0.5063460425226173 Yang Didapatkan Dengan Mencoba Jumlah *Cluster* Dari Dua Hingga Sembilan *Cluster*. Tiga *Cluster* Tersebut Mendapatkan Identitas Antara Lain, Produk Tidak Laris, Produk Cukup Laris Dan Produk Sangat Laris Yang Dapat Dilihat Pada **Tabel 1**. Setiap *Cluster* Yang Telah Terbentuk Memiliki Keterkaitan Antar Satu Sama Lain Di Setiap Anggotanya.

**Tabel 1** Hasil *Clustering*

| <i>Cluster</i> | Kategori <i>Cluster</i> | Jumlah Produk |
|----------------|-------------------------|---------------|
| 1              | Kurang Laris            | 84            |
| 2              | Cukup Laris             | 12            |
| 3              | Sangat Laris            | 4             |

### **Kesimpulan**

Berdasarkan Penelitian Yang Telah Dilakukan Dengan Menggunakan Teknik *Data Mining* Dengan Menggunakan Metodologi *Kdd*, Dapat Disimpulkan Bahwa:

- 1) Penerapan Algoritma *K-Means* Dalam Mengelompokkan Data Stok Produk Toko Xyz Menghasilkan Tiga *Cluster*. *Cluster* 1, Yang Dianggap Kurang Laris, Terdiri Dari 84 Produk. *Cluster* 2, Yang Cukup Laris, Mencakup 12 Produk. Sementara Itu, *Cluster* 3, Yang Sangat Laris, Hanya Berisi 4 Produk. Produk Dalam *Cluster* 2 Memiliki Penjualan Yang Stabil Dan Dapat Dipertahankan. Namun, Untuk Produk Dalam *Cluster* 1 Yang Kurang Laris, Strategi Peningkatan Penjualan Perlu Diterapkan. Salah Satu Pendekatan Yang Bisa Dilakukan Adalah Dengan Memberikan Promosi Atau Menggabungkan Produk Ini Dengan Produk Dalam *Cluster* 3 Yang Sangat Laris.
- 2) Evaluasi Dilakukan Dengan Menggunakan *Silhouette Coefficient*, Dimana Nilai Yang Hampir Mencapai Satu Dianggap Sebagai Hasil Optimal Untuk Proses Pengelompokan. Spesifik Untuk Kasus Ini, Ketika Jumlah *Cluster*(*K*) Ditetapkan Menjadi Tiga, Hasil *Silhouette Coefficient* Mencapai 0,506, Menunjukkan Hasil Yang Cukup Baik Dalam Proses Pengelompokan.

### **Saran**

Berdasarkan Penelitian Yang Telah Dilakukan, Beberapa Saran Yang Dapat Diberikan Adalah:

- 1) Peninjauan Algoritma Lain: Meskipun Algoritma *K-Means* Telah Memberikan Hasil Yang Memuaskan, Mungkin Perlu Dipertimbangkan Untuk Menggunakan Algoritma Pengelompokan Lain Seperti *Dbscan* Atau *Hierarchical Clustering* Untuk Melihat Apakah Hasilnya Bisa Lebih Optimal.

Penerapan Teknik Prediksi: Selain Teknik Pengelompokan, Teknik Prediksi Seperti Arima Atau Lstm Dapat Digunakan Untuk Memprediksi Tren Penjualan Di Masa Depan Berdasarkan Data Historis.

#### DAFTAR PUSTAKA

- Arhami, M., & Nasir, M. (2020). Data Mining: Algoritma Dan Implementasi. In R. I. Utami (Ed.), *Data Mining Algoritma Dan Implementasi* (1st Ed.). Penerbit Andi.
- Indra, I., Pratiwi, W. A. A., & Putra, Y. D. (2022). Pengaruh Biaya Promosi Terhadap Penjualan. *Jurnal Ekonomi, Manajemen Dan Akuntansi*, 24(4), Hal 711-716. <https://doi.org/10.30872/jfor.v24i4.11704>
- Indriyani, F., & Irfiani, E. (2019). Clustering Data Penjualan Pada Toko Perlengkapan Outdoor Menggunakan Metode *K-Means*. *Juita : Jurnal Informatika*, 7(2). <https://doi.org/10.30595/juita.v7i2.5529>
- Nagari, S. S., & Inayati, L. (2020). Implementation Of Clustering Using K-Means Method To Determine Nutritional Status. *Jurnal Biometrika Dan Kependudukan*, 9(1). <https://doi.org/10.20473/jbk.v9i1.2020.62-68>
- Paembonan, S., & Abduh, H. (2021). Penerapan Metode Silhouette Coefficient Untuk Evaluasi Clustering Obat. *Pena Teknik: Jurnal Ilmiah Ilmu-Ilmu Teknik*, 6(2). [https://doi.org/10.51557/pt\\_jit.v6i2.659](https://doi.org/10.51557/pt_jit.v6i2.659)
- Suyanto, D., & Others. (2017). Data Mining Untuk Klasifikasi Dan Clusterisasi Data. In *Bandung: Informatika Bandung* (Pertama). Informatika Bandung.
- Yahya, S., & Sugiyanto, C. (2020). Indonesian Demand For Online Shopping: Revisited. *Journal Of Indonesian Economy And Business*, 35(3), 188–203. <https://doi.org/10.22146/jieb.55358>
- Yunistya, D. I., Goejantoro, R., Deny, F., & Amijaya, T. (2022). *The Application Of K-Harmonic Means Method In District/City Grouping (Case Study: Poverty In Kalimantan Island In 2020) Penerapan Metode K-Harmonic Means Dalam Pengelompokan Kabupaten/Kota (Studi Kasus: Kemiskinan Di Pulau Kalimantan Tahun 2020)*. 19(1), 51–64. <https://doi.org/10.20956/j.v19i1.21116>
- Zhao, H. (2022). Design And Implementation Of An Improved *K-Means* Clustering Algorithm. *Mobile Information Systems*, 2022, 6041484. <https://doi.org/10.1155/2022/6041484>
- Zubair, M., Iqbal, M. A., Shil, A., Chowdhury, M. J. M., Moni, M. A., & Sarker, I. H. (2022). An Improved K-Means Clustering Algorithm Towards An Efficient Data-Driven Modeling. *Annals Of Data Science*. <https://doi.org/10.1007/S40745-022-00428-2>