

# PENERAPAN DATA MINING UNTUK PREDIKSI PENJUALAN BAJU DISTRO MENGGUNAKAN METODE K-MEANS CLUSTERING

Oleh :

Shafa Arrizqa Az Zahroh

211080200126

Program Studi Informatika

Universitas Muhammadiyah Sidoarjo

Mei, 2025

# Pendahuluan

- Strategi untuk meningkatkan penjualan produknya pada Toko Aldi, khususnya penjualan baju distro
- Industri ini menjadi semakin kompetitif, dan penjualan baju distro perlu mengelola inventaris dan memenuhi permintaan pelanggan dengan lebih efektif dan efisien
- Metode ini membantu toko untuk mengembangkan strategi persediaan stok baju
- Oleh karena itu penelitian ini muncul dengan adanya kelemahan yg ada pada refrensi sebelumnya, sehingga kami membuat penelitian ini dilaksanakan dengan memanfaatkan metode *K-Means Clustering* dapat dimanfaatkan dalam mengklasifikasikan berbagai produk menurut pola penjualan, musim penjualan, kategori produk, dan variabel-variabel lainnya



# Rumusan Masalah

- Bagaimana menggunakan metode *K-Means* dalam prediksi penjualan?
- Bagaimana mengklasifikasi stok barang pada prediksi penjualan dalam bentuk *clustering*?

# Metode Penelitian



# Metode Penelitian

1.

## Studi Literatur

Tahapan studi literatur dilakukan untuk mengumpulkan referensi yang relevan dan mendukung penelitian. Proses ini bertujuan memperkaya landasan teori untuk merancang program yang efektif dan inovatif, serta memperdalam pemahaman konsep data mining dan penerapannya dalam pengelolaan data di Toko Baju Distro.

# Metode Penelitian

2.

## Pengumpulan Data

Dalam tahapan ini ini, proses pengumpulan data mulai dilaksanakan di Toko Aldi sebagai langkah strategis untuk mendukung jalannya penelitian ini. Data yang berhasil dihimpun berjumlah sebanyak 393 entri, yang dikumpulkan selama periode waktu bulan April hingga Juli tahun 2024. Informasi tersebut diperoleh melalui analisis hasil rekapan penjualan yang dilakukan secara cermat dan sistematis selama dua bulan tersebut, sehingga diharapkan dapat memberikan gambaran yang komprehensif terkait aktivitas penjualan di toko tersebut.



# Metode Penelitian

3.

## *Pre-Processing*

Pre processing data merupakan tahap penting sebelum data dianalisis proses ini meliputi :

- Pembersihan Data (Data *Cleaning*)
- Normalisasi Data
- Pemilihan Fitur (*Feature Selection*)



# Metode Penelitian

4.

## Implementasi Algoritma

Pada tahap implementasi, dilakukan pemrosesan data dengan memanfaatkan platform Google Colaboratory sebagai alat utama untuk menjalankan algoritma k-means. Dalam proses ini, algoritma k-means tidak hanya diterapkan secara mandiri, tetapi juga dikombinasikan dengan dua teknik evaluasi yang relevan, yaitu Silhouette Coefficient dan Davies-Bouldin Index. Kombinasi kedua teknik tersebut bertujuan untuk mengevaluasi kinerja model clustering secara lebih komprehensif, sehingga memungkinkan penentuan jumlah kluster optimal serta memastikan bahwa pengelompokan yang dihasilkan terdapat kualitas secara tinggi juga sejalan terhadap karakteristik data yang dianalisis.



# Metode Penelitian

5.

## Evaluasi

Setelah clustering selesai, kita harus mengevaluasi kualitas pengelompokan yaitu untuk menentukan apakah data tersebut sudah optimal menggunakan perhitungan :

- Silhoutte Score
- Davies-Bouldin Index (DBI)



# Hasil dan Pembahasan

1.

## Pengumpulan Data

Data yang dikumpulkan berupa data penjualan baju distro di Toko Aldi Sidoarjo dengan format .csv (microsoft excel) sebanyak 393 baris data dan memiliki 6 atribut utama yang dijadikan variabel untuk mendukung proses prediksi yaitu Tanggal, Status Pembayaran, Type Barang, Quantity, Harga, Jumlah. Berikut merupakan sampel data yang disajikan pada tabel 1 :

**Tabel 1** Sampel Data

No	Tanggal	Status Pembayaran	Type Barang	Quantity	Harga	Jumlah
1.	01-Apr-24	Transfer	PU	1	Rp.45.000	Rp.45.000
2.	01-Apr-24	Cash	KU	2	Rp.30.000	Rp.60.000
3.	01-Apr-24	Cash	PH-WL	1	Rp.65.000	Rp.65.000
4.	01-Apr-24	Cash	KU	1	Rp.30.000	Rp.30.000
...	...	...	...	...	...	...
390.	26-Jul-24	Transfer	PH-WP	2	Rp.65.000	Rp.130.000
391.	27-Jul-24	Transfer	PH-WL	1	Rp.65.000	Rp.65.000
392.	28-Jul-24	Transfer	PJU	1	Rp.50.000	Rp.50.000
393.	29-Jul-24	Transfer	PU	1	Rp.45.000	Rp.45.000

# Hasil dan Pembahasan

2.

## Pre-Processing Data

Proses pengolahan data awal dilaksanakan dengan tujuan dalam memperbaiki kualitas data serta membuat data siap untuk diproses lebih lanjut. Tahapan yang dilakukan meliputi pengambilan atribut yang dibutuhkan yaitu atribut “Type Barang”, “Quantity”, dan “Jumlah” dan menghilangkan atribut yang tidak digunakan. Selanjutnya menghilangkan simbol “Rp” dan tanda baca “titik (.)” pada isi atribut biaya sewa per hari dari yang awalnya “Rp. 35.000” menjadi “35000”. Sampel data yang sudah melalui proses Pre-Processing data sebelumnya akan ditunjukkan oleh tabel 2 :

**Tabel 2. Data Setelah *Pre-Processing***

	Type Barang	Quantity	Jumlah
1	1	1	45000
2	2	2	60000
3	3	1	65000
4	2	1	30000
...	...	...	...
390.	6	2	130000
391.	3	1	65000
392.	4	1	50000
393.	1	1	45000

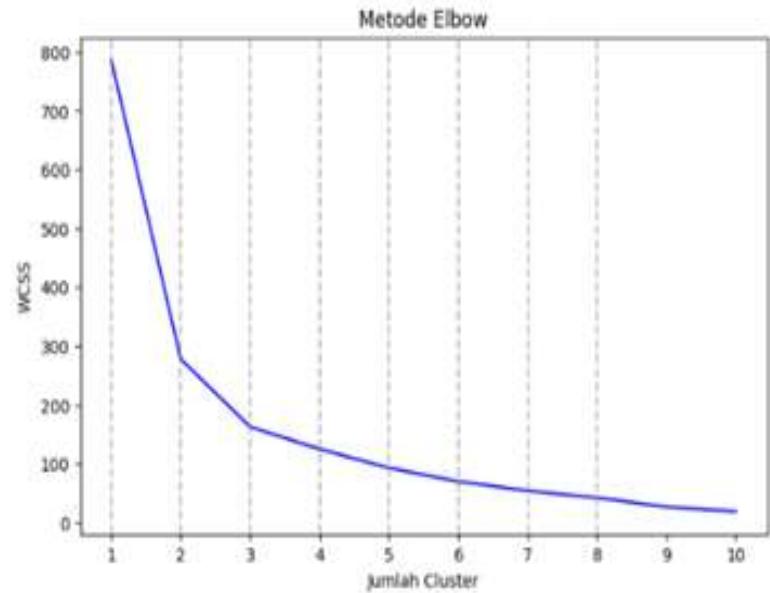


# Hasil dan Pembahasan

3.

## Implementasi Algoritma

Metode Elbow digunakan untuk menentukan jumlah cluster optimal dengan mencari titik siku pada grafik WCSS. Pengelompokan cluster diuji dari 1 hingga 10 cluster, dengan sumbu X mewakili jumlah cluster dan sumbu Y adalah nilai WCSS.



Gambar 2. Grafik Metode Elbow

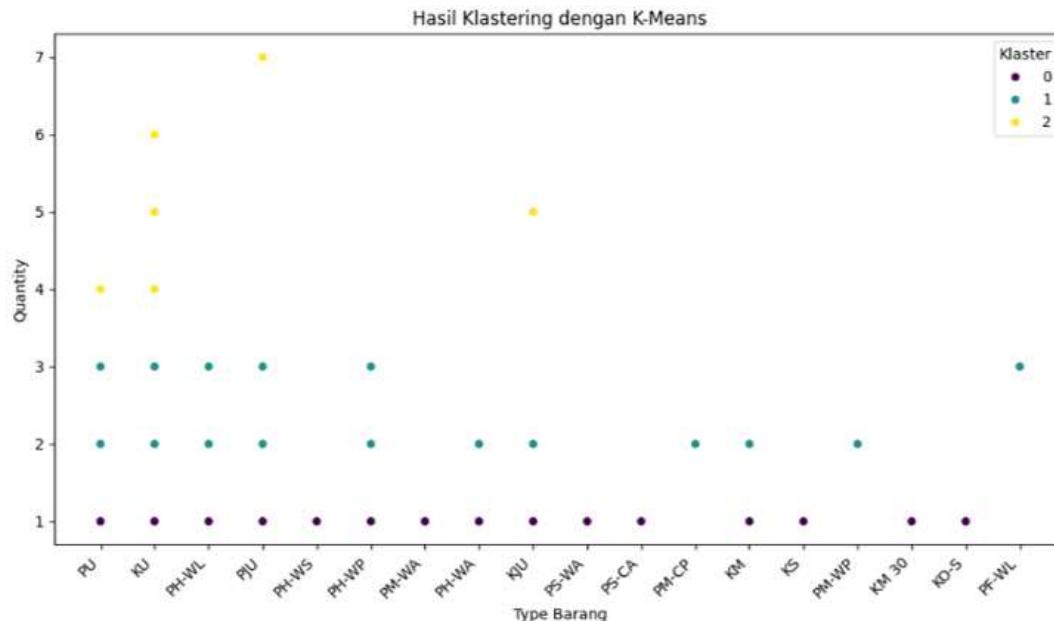
Pada gambar 2 terjadi sudut lengkung pada cluster 2 dan 3. Hal ini menunjukkan bahwa cluster optimal ada pada cluster 2 dan 3. Namun, untuk menguji keakuratan cluster harus menggunakan matriks untuk menghitungnya.

# Hasil dan Pembahasan

4.

## Evaluasi Hasil

Hasil clustering yang divisualisasikan dalam diagram scatter mengindikasikan bahwa data tersegmentasi ke dalam sejumlah kelompok yang dibedakan berdasarkan warna, di mana setiap kelompok merepresentasikan satu klaster yang dibentuk oleh algoritma K-Means..



Gambar 3. Grafik Hasil Clustering

# Hasil dan Pembahasan

4.

## Evaluasi Hasil

Pada gambar 3, dapat diidentifikasi bahwa grafik menunjukkan pola bahwa produk dengan penjualan rendah cenderung masuk klaster 0 (quantity 1), penjualan sedang di klaster 1 (quantity 2–3), dan penjualan tinggi di klaster 2 (quantity 4–7).

Hasil clustering menunjukkan :

- Klaster 0: 214 unit terjual, pendapatan Rp10.267.000
- Klaster 1: 317 unit, pendapatan Rp13.555.000
- Klaster 2: 206 unit, pendapatan Rp13.160.000 (harga rendah, volume tinggi)

Produk dengan penjualan dan pendapatan tinggi berkontribusi besar dan layak menjadi fokus pemasaran. Jumlah anggota tiap klaster tercantum pada Tabel 4.

# Hasil dan Pembahasan

4.

## Evaluasi Hasil

Tabel 4. Data Anggota Cluster

c	n	x	nx	hx
0	216	PU : Polo Urgan	77	3.465.000
		KU : Kaos Urgan	43	1.290.000
		PH-WL : Polo Hold/W/List	26	1.690.000
		PJU : Polo Jumbo Urgan	13	650.000
		KJU : Kaos Jumbo Urgan	10	350.000
		PH-WS : Polo Hold/W/Salur	9	657.000
		PH-WA : Polo Hold/W/Apl	8	520.000
		PH-WP : Polo Hold/W/Polos	6	390.000
		PM-WA : Polo Monk1/W/Apl	6	390.000
		PS-WA : Polo Six/W/Apl	4	280.000
		PS-CA : Polo Six/C/Apl	3	195.000
		KM : Kaos Mirror	3	150.000
		KO-S : Kaos Oastok/Slur	3	90.000
		KM 30 : Kaos Mirror 30'S	2	100.000
		KS : Kaos Salur	1	50.000
1	317	KU : Kaos Urgan	138	4.140.000
		PU : Polo Urgan	97	4.365.000
		PJU : Polo Jumbo Urgan	18	900.000
		KJU : Kaos Jumbo Urgan	4	140.000
		PH-WP : Polo Hold/W/Polos	18	1.170.000
		PF-WL : Polo Feold/W/List	9	675.000
		PH-WL : Polo Hold/W/List	13	845.000
		PH-WA : Polo Hold/W/Apl	6	390.000
		PM-WP : Polo Monk1/W/Print	6	438.000
		PM-CP : Polo Monk1/C/Polos	4	292.000
		KM : Kaos Mirror	4	200.000
2	206	KU : Kaos Urgan	152	4.560.000
		PU : Polo Urgan	20	900.000
		KJU : Kaos Jumbo Urgan	20	700.000
		PJU : Polo Jumbo Urgan	14	700.000

Keterangan :

- c : Cluster
- n : Jumlah total terjual keseluruhan type barang
- x : Jenis type barang
- nx : Jumlah total tejual pada masing-masing type barang
- hx : Total harga keseluruhan pada masing-masing type barang

Pada tabel 4 data anggota cluster menunjukkan pembagian data pada masing-masing cluster. Cluster 0 terdapat 216 jumlah total terjual type barang dengan 15 type barang, cluster 1 terdapat 317 jumlah total terjual type barang dengan 11 type barang, dan cluster 2 terdapat 206 jumlah total terjual type barang dengan 4 type barang.

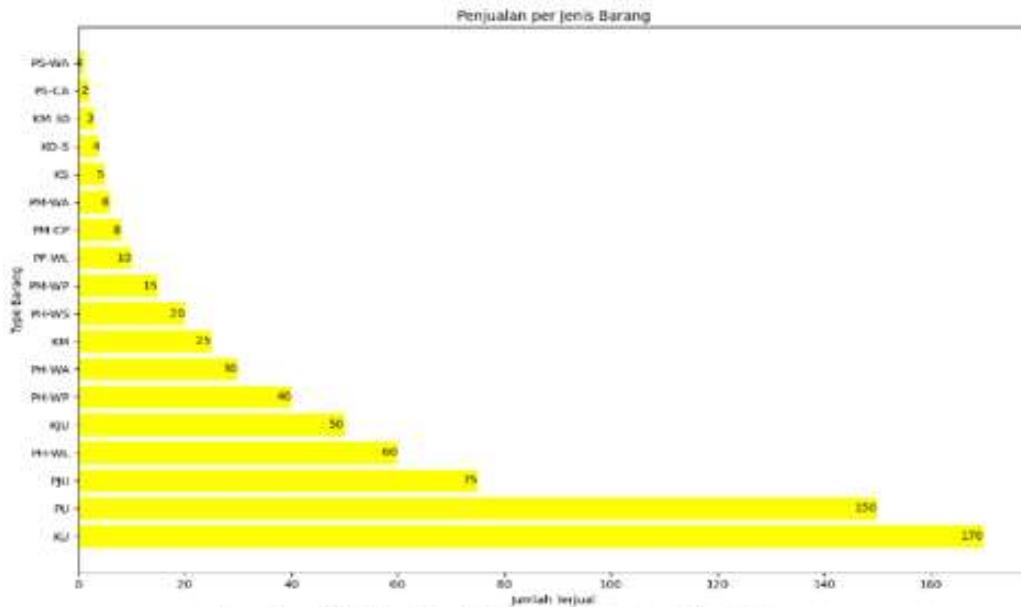


# Hasil dan Pembahasan

4.

## Evaluasi Hasil

### GRAFIK JUMLAH PENJUALAN PER JENIS BARANG



Gambar 4. Tabel Jumlah Penjualan Per Jenis Barang

Analisis data penjualan penting untuk mengidentifikasi pola dan tren pasar. Berdasarkan diagram batang horizontal, terlihat perbedaan signifikan antar kategori produk. Baju tipe KU, PU, dan PJU paling laku (170, 150, dan 75 unit), disusul tipe PH-WL, KJU, PH-WP, dan PH-WA dengan penjualan sedang (60–30 unit). Sementara itu, tipe KS, KO-S, KM 30, PS-CA, dan PS-WA kurang diminati (1–5 unit). Hal ini menunjukkan adanya preferensi konsumen terhadap produk tertentu.

# Hasil dan Pembahasan

5.

## Hasil Klasterisasi

Hasil evaluasi menunjukkan kualitas klaster yang sangat baik. Nilai Silhouette Coefficient sebesar 0.576 menunjukkan kemiripan tinggi dalam klaster dan perbedaan antar klaster. Calinski-Harabasz Index sebesar 19.125 mengindikasikan struktur klaster yang jelas. Davies-Bouldin Index sebesar 0.308 menunjukkan batas klaster yang tegas dan minim tumpang tindih.

➡ Estimated number of clusters: 3  
Silhouette Coefficient: 0.576  
Calinski\_harabasz Index: 19.125  
Davies-Bouldin Index: 0.308

**Gambar 5.** Hasil Silhouette dan DBI



# Kesimpulan

Berdasarkan penelitian yang dilakukan, metode K-Means Clustering terbukti mampu mengelompokkan pola penjualan baju distro di Toko Aldi secara efektif berdasarkan tingkat penjualannya. Dari 393 transaksi yang dianalisis, hasil klasterisasi mengidentifikasi tiga kategori utama, yakni produk dengan tingkat penjualan tinggi (laku) sebanyak 170, 150 dan 75 barang, tingkat penjualan menengah (cukup laku) sebanyak 60, 50, 40 dan 30 barang, dan tingkat penjualan rendah (kurang laku) sebanyak 1 hingga 5 barang. Evaluasi menggunakan Silhouette Coefficient (0.576), Calinski-Harabasz Index (19.125), dan Davies-Bouldin Index (0.308) membuktikan mengenai klaster yang terbentuk terdapat tingkat kohesi secara baik serta pemisahan yang jelas. Temuan ini memberikan wawasan bagi pemilik usaha dalam merancang strategi pemasaran yang lebih optimal serta meningkatkan efisiensi pengelolaan stok, sehingga dapat meningkatkan kinerja operasional dan daya saing bisnis di industri distro..



“SELESAI”