



Analysis of Best Selling Product Using The K-Means Clustering Method at Distro Engkus Store

Analisis Produk Terlaris Menggunakan Metode K-Means Clustering Pada Toko Distro Engkus

Sofia Dewi¹, Miki Wijana², Asep Saepuloh³
^{1, 2, 3}Sistem Informasi, Universitas Ma'soem, Indonesia

E-Mail: sofiadewi.job@gmail.com, mikiwijana@gmail.com, epul7825@gmail.com

Makalah: Diterima 21 November 2025; Diperbaiki 20 Desember 2025; Disetujui 14 Januari 2026
Corresponding Author: Sofia Dewi

Abstrak

Penelitian ini bertujuan mengatasi masalah Distro Engkus dalam menentukan produk terlaris dan kurang laris secara manual, yang rawan kesalahan dalam pengambilan keputusan pengadaan barang dan pengelolaan stok. Oleh karena itu, diterapkan algoritma K-Means Clustering untuk menganalisis data produk dengan variabel Stok Awal, Stok Terjual, dan Stok Akhir. K-Means digunakan untuk mengelompokkan produk ke dalam tiga kategori: sangat laris, laris, dan kurang laris. Kebaruan penelitian ini terletak pada pemilihan variabel clustering yang secara langsung bertujuan menyelesaikan masalah manajemen inventaris dan bukan segmentasi pelanggan. Temuan menunjukkan bahwa penerapan K-Means efektif dalam mengelompokkan produk secara objektif dan efisien. Hasil validasi internal menggunakan Davies-Bouldin Index (DBI) menunjukkan nilai 0,392 mengindikasikan kualitas clustering yang baik. Penggunaan RapidMiner mendukung manajemen dalam mengambil keputusan strategis yang lebih cepat dan akurat, terutama dalam pengadaan dan pengelolaan persediaan.

Keyword: Data Mining, K-Means Clustering, Penjualan, Distro Engkus, RapidMiner.

Abstract

This study aims to address the problem of Distro Engkus in manually determining best-selling and least-selling products, which is prone to errors in procurement and stock management decisions. Therefore, the K-Means Clustering algorithm is applied to analyze product data with the variables Initial Stock, Sold Stock, and Ending Stock. K-Means is used to group products into three categories: best-selling, best-selling, and least-selling. The novelty of this study lies in the selection of clustering variables that directly aim to solve inventory management problems and not customer segmentation. The findings show that the application of K-Means is effective in grouping products objectively and efficiently. The results of internal validation using the Davies-Bouldin Index (DBI) showed a value of 0.392 indicating good clustering quality. The use of RapidMiner supports management in making faster and more accurate strategic decisions, especially in procurement and inventory management.

Keyword: Data Mining, Distro Engkus, K-Means Clustering, RapidMiner Sales.

1. PENDAHULUAN

Perubahan dalam teknologi informasi, terutama internet, telah secara drastis mempengaruhi cara bisnis beroperasi. Sejumlah pelaku usaha seperti Distro Engkus sekarang memanfaatkan *platform e-commerce* untuk memperluas pasar dan meningkatkan penjualan [1]. Peningkatan jumlah transaksi penjualan ini beriringan dengan timbulnya data dalam jumlah besar. Distro Engkus mengalami kesulitan dalam pengelolaan dan analisis data penjualan yang dilakukan secara manual. Di era digital, jumlah data yang masif membutuhkan teknik analisis yang efektif. Data mining menjadi disiplin ilmu yang krusial untuk menemukan pola tersembunyi dan pengetahuan berharga dari data, yang kemudian dapat digunakan untuk mendukung pengambilan keputusan strategis di berbagai sektor. Penerapannya mulai dari analisis transaksi ritel, evaluasi kualitas layanan untuk menentukan kepuasan pelanggan [2].

Distro Engkus, yang merupakan usaha ritel di sektor *fashion*, menghadapi tantangan signifikan dalam mengelola volume data transaksi harian yang terus meningkat akibat perluasan penjualan melalui *e-commerce*. Saat ini, dengan puluhan Stock Keeping Unit (SKU) dan ratusan transaksi per bulan, proses analisis kinerja produk penentuan produk terlaris dan kurang laris masih dilakukan secara manual dan didasarkan pada asumsi

subjektif dari pemilik atau staf penjualan, Pendekatan *non sistematis* ini menimbulkan risiko kesalahan yang tinggi, terutama dalam proyeksi permintaan dan evaluasi persediaan. Pengambilan keputusan strategis mengenai pengadaan barang menjadi tidak akurat, sering kali menyebabkan terjadinya penumpukan stok dan kelebihan persediaan untuk produk yang kurang diminati, atau sebaliknya, kekurangan stok dan kehilangan potensi penjualan untuk produk yang sedang populer. Keterbatasan ini secara langsung berdampak pada inefisiensi modal kerja dan penurunan daya saing dalam pasar ritel yang kompetitif.

Meskipun Algoritma *K-Means Clustering* efektif dalam konteks ritel, sebagian besar studi sejenis berfokus pada segmentasi pelanggan menggunakan metrik *Recency, Frequency, Monetary* (RFM) untuk mengelompokkan konsumen [3] [4]. Perbedaan fokus ini menyisakan kesenjangan (GAP) terhadap masalah Distro Engkus. Kebutuhan Distro Engkus adalah pada kinerja dan manajemen persediaan produk. Kebaruan penelitian ini terletak pada pemilihan variabel *clustering* yang secara spesifik relevan dengan masalah pengendalian stok, yaitu Stok Awal, Stok Terjual, dan Stok Akhir, alih-alih metrik RFM. Optimalisasi pengadaan barang dan mitigasi penumpukan stok menjadi target utama. Penelitian ini memperkuat literatur di bidang Data Mining dengan menyediakan bukti empiris mengenai efektivitas *K-Means* untuk segmentasi produk dalam konteks ritel fashion, didukung oleh validasi kuantitatif *Davies-Bouldin Index* (DBI) [5].

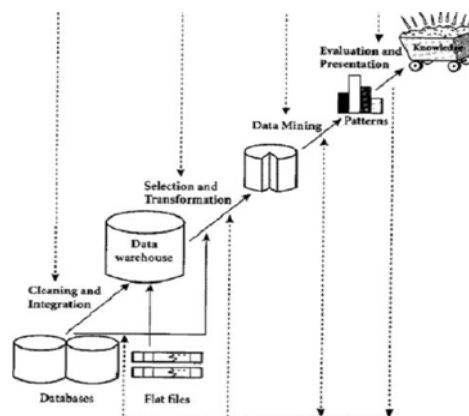
Untuk menyelesaikan masalah tersebut, penelitian ini mengusulkan penggunaan data mining dengan algoritma *K-Means Clustering*. *K-Means* merupakan salah satu teknik pengelompokan yang efektif untuk mengelompokkan data berdasar kesamaan karakteristik [6]. Dalam hal ini, *K-Means* akan diterapkan untuk secara otomatis dan objektif mengelompokkan produk-produk Distro Engkus ke dalam beberapa kategori menurut pola penjualannya, seperti produk yang sangat laris, sedang laris, dan kurang laris. Proses analisis ini akan didukung oleh perangkat lunak *RapidMiner* untuk memvisualisasikan dan menyederhanakan interpretasi hasil pengelompokan.

Penggunaan data mining dengan *K-Means* untuk mengelompokkan produk berdasarkan pola penjualannya, seperti yang telah diterapkan dalam studi kasus analisis produk terlaris pada perusahaan lain, menjadi solusi yang efisien dan objektif [7]. Diharapkan hasil penelitian ini dapat menjadi acuan dan inspirasi bagi pemilik usaha lain untuk menggunakan teknologi data mining dalam mengoptimalkan operasi bisnis mereka. Penerapan algoritma *K-Means* terbukti menjadi strategi yang berhasil diterapkan dalam menganalisis produk terlaris, khususnya untuk mendukung keputusan pada usaha UMKM [8].

2. METODE DAN BAHAN

Penelitian ini menerapkan pendekatan deskriptif kuantitatif yang berfokus pada penggunaan data mining melalui algoritma *K-Means Clustering*. Algoritma *K-Means Clustering* adalah metode analisis data atau teknik data mining yang menjalankan pemodelan tanpa pengawasan (*unsupervised*) dan termasuk salah satu metode yang mengelompokkan data dengan sistem pembagian [9]. *Clustering* adalah proses pengelompokan titik-titik data ke dalam dua kelompok atau lebih, di mana titik-titik dalam kelompok yang sama memiliki kesamaan yang lebih besar dibandingkan dengan titik-titik di kelompok yang berbeda [9]. Sumber data dalam penelitian ini adalah transaksi penjualan produk pakaian dari Distro Engkus. Data yang digunakan terdiri dari atribut 'Nama Produk', 'Stok Awal', 'Stok Terjual', dan 'Stok Akhir'. Jumlah data yang dianalisis mencakup 60 produk. Pemilihan atribut ini didasarkan pada relevansinya dalam menilai kinerja penjualan dan ketersediaan stok barang.

Proses penelitian ini dilaksanakan secara terstruktur mengikuti langkah-langkah *Knowledge Discovery in Databases* (KDD) yang disesuaikan dengan kebutuhan data mining. KDD adalah proses dengan alat bantu komputer untuk menganalisis himpunan data dalam jumlah besar guna menghasilkan pengetahuan baru yang berguna [10]. Hal ini diilustrasikan dalam Gambar 1. Tahapan-tahapan tersebut bersifat interaktif, dengan partisipasi pengguna secara langsung atau melalui basis pengetahuan [11].



Gambar 1. Tahapan Data Mining

Sumber: Meirisah [11]

Tahapan untuk melakukan pengelompokan menggunakan metode *K-Means* adalah sebagai berikut [9]:

1. Tentukan jumlah kluster K yang diinginkan.
2. Lakukan inisialisasi untuk k pusat kluster, dan ini bisa dilakukan dengan berbagai metode. Metode yang paling umum adalah dengan memilih angka secara acak. Pusat-pusat kluster akan diberikan nilai awal menggunakan angka-angka acak.
3. Tempatkan setiap data atau objek pada kluster yang paling dekat. Tingkat kedekatan antara dua objek dihitung berdasarkan jarak di antara mereka. Selain itu, jarak antara data dengan pusat kluster juga menentukan kedekatan data dengan kluster tertentu. Pada tahap ini, perlu dihitung jarak setiap data terhadap semua pusat kluster. Jarak terpendek antara suatu data dan kluster tertentu akan mengindikasikan di kluster mana data tersebut akan ditempatkan. Untuk menghitung jarak semua data ke setiap titik pusat kluster, diterapkan teori jarak Euclidean yang dirumuskan sebagai berikut:

$$D(i, j) = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2}$$

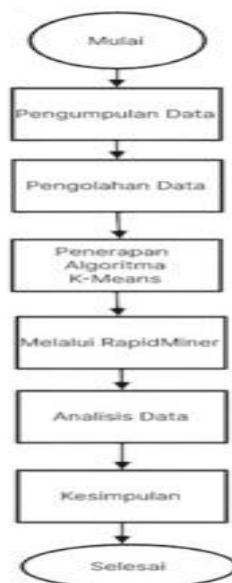
Keterangan:

$D(i, j)$ = Jarak data ke i ke pusat *cluster* j

x_{ki} = Data ke i pada atribut data ke k

x_{kj} = Titik pusat ke j pada atribut ke k

4. Lakukan perhitungan ulang untuk pusat cluster berdasarkan keanggotaan cluster yang ada saat ini. Pusat cluster merupakan nilai rata-rata dari seluruh data atau objek yang tergabung dalam cluster tertentu. Sebagai alternatif, dapat juga dipertimbangkan penggunaan median dari cluster tersebut. Maka, rata-rata atau mean bukanlah satu-satunya cara untuk mengukurnya.
5. Alokasikan kembali setiap objek menggunakan pusat cluster yang telah diperbarui. Jika pusat cluster tidak mengalami perubahan lebih lanjut, maka proses *Clustering* dianggap selesai. Atau, ulangi langkah ketiga hingga pusat cluster tetap sama. Prosedur *K-Means* dapat seperti dijabarkan dalam flowchart berikut:



Gambar 2. Prosedur *K-Means*

Sumber: Rahayu & Dewi, 2025 [10]

Pengujian kualitas *Clustering* dilakukan secara internal menggunakan Davies-Bouldin Index (DBI). DBI mengukur rasio antara jumlah dispersi intra-cluster dan dispersi antar-cluster dengan persamaan:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j})$$

Keterangan:

K = Jumlah *cluster* yang dihitung

Evaluasi kualitas klasterisasi menggunakan DBI adalah metrik umum yang digunakan, bahkan untuk membandingkan kinerja algoritma *clustering* [12]. Semakin kecil nilai DBI, semakin baik kualitas *Clustering* yang dihasilkan, menunjukkan cluster yang padat dan terpisah satu sama lain.

3. HASIL DAN PEMBAHASAN

3.1 Analisa Data

Data untuk penelitian ini berasal dari Distro Engkus, di mana informasi yang digunakan mencakup 60 data penjualan. Hasil yang diinginkan adalah menciptakan tiga kelompok yaitu (C1) data penjualan terlaris, (C2) data penjualan yang cukup laris, dan (C3) data penjualan yang kurang diminati. Atribut yang digunakan dalam pengelompokan data penjualan mencakup Stok Awal dan Terjual, sebagaimana ditunjukkan pada Tabel 1.

Tabel 1. Sampel Data

No	Nama Produk	Stok Awal	Stok Terjual	Stok Akhir
1	T-shirt Polos Hitam Basic	150	75	75
2	T-shirt Grafis "Urban Explorer"	120	50	70
3	T-shirt Polos Putih Heavyweight	130	60	70
4	T-shirt Garis-garis Navy	90	35	55
5	T-shirt Band Rock Klasik	80	32	48
6	Celana Jeans Slim Fit Indigo	70	25	45
7	Celana Chino Twill Khaki	85	20	65
8	Hoodie Oversize Hitam	50	18	32
9	Jaket Varsity Retro	40	15	25
10	Kemeja Flanel Merah Kotak	55	20	35
11	T-shirt Oversize Beige	110	45	65
12	T-shirt V-Neck Abu Misty	85	30	55
13	Polo Shirt Pique Navy	75	28	47
14	Celana Pendek Cargo Ripstop	65	22	43
15	Sweatpants Fleece Grey	50	18	32
16	Rompi Denim Vintage	35	10	25
17	Kemeja Denim Washed	40	14	26
18	T-shirt Kerah Tinggi Rib	65	20	45
19	T-shirt Grafis "Old School"	95	38	57
20	Kemeja Hawaiian Print	30	10	20
21	Celana Pendek Chino Navy	50	18	32
22	Kemeja Korduroi Coklat	35	12	23
23	T-shirt Polos Hijau Army	100	40	60
24	Celana Jeans Regular Fit	60	25	35
25	Celana Jogger Cargo	45	16	29
26	Cardigan Rajut Waffle	40	14	26
27	Kemeja Oxford Biru Muda	45	17	28
28	T-shirt Oversize Hijau Botol	70	28	42
29	Jaket Coach Nylon	30	9	21
30	Celana Trackpant Stripes	40	15	25
31	T-shirt Persib Home Jersey Style	200	100	100
32	T-shirt Persib Away Jersey Style	180	90	90
33	T-shirt Persib Training Kit	150	70	80
34	T-shirt Persib Logo Klasik	160	80	80
35	T-shirt Persib "Maung Bandung" Grafis	140	65	75
36	T-shirt Persib Font Retro	130	60	70
37	T-shirt Persib "Pangeran Biru"	120	55	65
38	T-shirt Persib Edisi Juara 2024	250	150	100
39	T-shirt Persib "Bobotoh Sejati"	170	85	85
40	T-shirt Persib Ilustrasi Maskot	110	45	65
41	T-shirt Persib Kode Area Bandung	100	40	60

No	Nama Produk	Stok Awal	Stok Terjual	Stok Akhir
42	T-shirt Persib Angka Kebanggaan	90	35	55
43	T-shirt Persib Motif Batik Biru	80	46	34
44	T-shirt Persib Gelora Siliwangi	70	25	45
45	T-shirt Persib Kolaborasi Artis	60	13	47
46	Hoodie Persib Biru Tua	80	17	63
47	Jaket Persib Tracksuit Emblem	70	25	45
48	Polo Shirt Persib Bordir	90	35	55
49	Celana Jogger Persib	60	17	43
50	Kaos Kaki Motif Persib	100	40	60
51	Kemeja Santai Kotak Coklat	40	15	25
52	Celana Pendek Motif Camo	30	10	20
53	Sweater Crewneck Rajut	35	12	23
54	T-shirt Grafis "Wanderlust"	70	25	45
55	Jaket Windbreaker Navy	25	8	17
56	T-shirt Persib Grafis "Viking"	110	50	60
57	T-shirt Persib Edisi Spesial Anniversary	120	55	65
58	T-shirt Persib Klasik Putih	90	35	55
59	T-shirt Persib Simbol Kebanggaan	80	65	15
60	T-shirt Persib "Jabar Juara"	70	25	45

3.2 Penerapan K-Means Clustering

1. Tahap pertama adalah menetapkan jumlah kelompok yang ingin dihasilkan. Dalam studi ini, jumlah kelompok yang akan dibuat adalah sebanyak 3 kelompok. Selanjutnya, tentukan titik pusat awal untuk masing-masing kelompok, Centroid awal (pusat *cluster*) dipilih secara acak (Random Seed: 42) dari sampel data. Nilai yang terpilih adalah sebagai berikut [13]..

Tabel 2. Pusat Awal *Cluster*

<i>Cluster</i>	Nama	SA	ST	SK
C0	Kurang Terjual	50	17	33
C1	Sedang Terjual	99	42	57
C2	Laris terjual	175	89	86

2. Arahkan seluruh data atau objek menuju kluster yang paling dekat dan hitung jarak antara data pertama dan pusat kluster. Jarak antara setiap produk dan centroid dihitung menggunakan rumus Euclidean Distance.

$$d(x, c_i) = \sqrt{\sum_{j=1}^p (x_j - c_{ij})^2}$$

3. Dari penghitungan jarak antara data pertama dengan pusat *cluster* yang telah disebutkan, dapat disimpulkan bahwa data pertama memiliki jarak terdekat dengan *cluster* 3. Detail lengkap dari perhitungan untuk 60 data produk bisa dilihat pada tabel 3.

Tabel 3. Hasil Perhitungan Jarak *Clustering*

No	Nama Produk	C0	C1	C2	Jarak	<i>Cluster</i>
1	T-shirt Polos Hitam Basic	123.00	63.36	30.69	30.69	2
2	T-shirt Grafis "Urban Explorer"	85.78	25.96	69.30	25.96	1
3	T-shirt Polos Putih Heavyweight	98.07	38.13	55.87	38.13	1
4	T-shirt Garis-garis Navy	49.07	11.58	105.37	11.58	1

5	T-shirt Band Rock Klasik	36.74	23.28	117.12	23.28	1
6	Celana Jeans Slim Fit Indigo	24.66	35.69	129.62	24.66	0
7	Celana Chino Twill Khaki	47.52	27.28	115.33	27.28	1
8	Hoodie Oversize Hitam	1.41	60.02	153.56	1.41	0
9	Jaket Varsity Retro	12.96	72.35	165.60	12.96	0
10	Kemeja Flanel Merah Kotak	6.16	53.89	147.52	6.16	0
11	T-shirt Oversize Beige	73.54	13.93	81.25	13.93	1
12	T-shirt V-Neck Abu Misty	43.34	18.55	111.99	18.55	1
13	Polo Shirt Pique Navy	30.69	29.53	123.46	29.53	1
14	Celana Pendek Cargo Ripstop	18.71	41.86	135.79	18.71	0
15	Sweatpants Fleece Grey	1.41	60.02	153.56	1.41	0
16	Rompi Denim Vintage	18.38	78.38	171.94	18.38	0
17	Kemeja Denim Washed	12.57	72.29	165.68	12.57	0
18	T-shirt Kerah Tinggi Rib	19.44	42.24	136.17	19.44	0
19	T-shirt Grafis "Old School"	55.15	5.66	99.21	5.66	1
20	Kemeja Hawaiian Print	24.86	84.58	177.83	24.86	0
21	Celana Pendek Chino Navy	1.41	60.02	153.56	1.41	0
22	Kemeja Korduroi Coklat	18.71	78.43	171.75	18.71	0
23	T-shirt Polos Hijau Army	61.30	3.74	93.28	3.74	1
24	Celana Jeans Regular Fit	12.96	47.90	141.15	12.96	0
25	Celana Jogger Cargo	6.48	66.15	159.62	6.48	0
26	Cardigan Rajut Waffle	12.57	72.29	165.68	12.57	0
27	Kemeja Oxford Biru Muda	7.07	66.20	159.52	7.07	0
28	T-shirt Oversize Hijau Botol	24.54	35.52	129.16	24.54	0
29	Jaket Coach Nylon	24.66	84.53	177.90	24.66	0
30	Celana Trackpant Stripes	12.96	72.35	165.60	12.96	0
31	T-shirt Persib Home Jersey Style	184.06	124.15	30.69	30.69	2
32	T-shirt Persib Away Jersey Style	159.62	99.77	6.48	6.48	2
33	T-shirt Persib Training Kit	122.55	62.56	31.97	31.97	2
34	T-shirt Persib Logo Klasik	135.20	75.46	18.49	18.49	2
35	T-shirt Persib "Maung Bandung" Grafis	110.31	50.34	43.84	43.84	2
36	T-shirt Persib Font Retro	98.07	38.13	55.87	38.13	1
37	T-shirt Persib "Pangeran Biru"	85.84	25.96	67.99	25.96	1
38	T-shirt Persib Edisi Juara 2024	249.36	190.56	97.68	97.68	2
39	T-shirt Persib "Bobotoh Sejati"	147.40	87.60	6.48	6.48	2
40	T-shirt Persib Ilustrasi Maskot	73.54	13.93	81.25	13.93	1
41	T-shirt Persib Kode Area Bandung	61.30	3.74	93.28	3.74	1
42	T-shirt Persib Angka Kebanggaan	49.07	11.58	105.37	11.58	1
43	T-shirt Persib Motif Batik Biru	41.74	30.10	116.52	30.10	1
44	T-shirt Persib Gelora Siliwangi	24.66	35.69	129.62	24.66	0
45	T-shirt Persib Kolaborasi Artis	17.66	49.62	143.26	17.66	0
46	Hoodie Persib Biru Tua	42.43	31.97	121.40	31.97	1
47	Jaket Persib Tracksuit Emblem	24.66	35.69	129.62	24.66	0
48	Polo Shirt Persib Bordir	49.07	11.58	105.37	11.58	1
49	Celana Jogger Persib	14.14	48.39	142.33	14.14	0
50	Kaos Kaki Motif Persib	61.30	3.74	93.28	3.74	1
51	Kemeja Santai Kotak Coklat	12.96	72.35	165.60	12.96	0
52	Celana Pendek Motif Camo	24.86	84.58	177.83	24.86	0

53	Sweater Crewneck Rajut	18.71	78.43	171.75	18.71	0
54	T-shirt Grafis "Wanderlust"	24.66	35.69	129.62	24.66	0
55	Jaket Windbreaker Navy	31.02	90.73	183.91	31.02	0
56	T-shirt Persib Grafis "Viking"	73.61	13.93	80.14	13.93	1
57	T-shirt Persib Edisi Spesial Anniversary	85.84	25.96	67.99	25.96	1
58	T-shirt Persib Klasik Putih	49.07	11.58	105.37	11.58	1
59	T-shirt Persib Simbol Kebanggaan	59.40	51.52	121.00	51.52	1
60	T-shirt Persib "Jabar Juara"	24.66	35.69	129.62	24.66	0

4. Setelah semua informasi dikelompokkan ke dalam *cluster* yang paling mendekati, selanjutnya hitung kembali pusat *cluster* yang baru berdasarkan nilai rata-rata anggota yang terdapat dalam *cluster* itu. Pembuatan centroid yang baru dihitung menggunakan rumus berikut. [5].

$$C = \frac{\sum m}{n}$$

Keterangan:

- C = centroid data
m = anggota data yang termasuk kedalam centroid tertentu
n = jumlah data yang menjadi anggota centroid tertentu.

Ulangi perhitungan *cluster* baru untuk parameter Terjual, sehingga hasil perhitungan diperoleh titik pusat baru dari setiap *cluster* dibawah ini.

Tabel 4. Titik Pusat Baru *Cluster*

<i>Cluster</i>	Nama	SA	ST	SK
C0	Kurang Terjual	49,83	17,28	32,55
C1	Sedang Terjual	98,7	41,57	57,13
C2	Laris terjual	175	89,38	85,63

5. Setelah memperoleh titik pusat baru untuk setiap kelompok, ulangi dari langkah ketiga sampai titik pusat untuk setiap kelompok tetap konstan dan tidak ada lagi data yang berpindah antara satu kelompok dengan kelompok yang lain.

Tabel 5. Hasil Akhir *Cluster* (sampel)

No	Nama Produk	Stok Terjual (ST)	<i>Cluster</i> Final	Kategori
1	T-shirt Polos Hitam Basic	75	C2	Sangat Laris
8	Hoodie Oversize Hitam	18	C0	Kurang Laris
31	T-shirt Persib Home Jersey Style	100	C2	Sangat Laris
38	T-shirt Persib Edisi Juara 2024	150	C2	Sangat Laris
45	T-shirt Persib Kolaborasi Artis	13	C0	Kurang Laris

Iterasi akhir penelitian ini dari proses *clustering*, maka di dapatkan hasil dengan total *cluster*: C0 (Kurang Laris) 29 produk, C1 (Laris) 23 produk, dan C2 (Sangat Laris) 8 Produk

3.3 Perhitungan DBI

1. Perhitungan DBI dimulai dengan mencari nilai melalui perhitungan nilai *sum of square within cluster* Perhitungan DBI dimulai dengan mencari nilai melalui perhitungan nilai *sum of square within cluster* (SSW), untuk mengetahui matrik kohesi dalam sebuah *cluster* ke-i [12]. Perhitungan SSW menggunakan rumus berikut.

$$SSW = \sum_{i=1}^k \sum_{x_j \in C_i} \|x_j - c_i\|^2$$

2. langkah selanjutnya hitung *sum of square between cluster* (SSB), untuk mengetahui separasi antar *cluster* menggunakan rumus berikut.

$$SSB = \sum_{i=1}^k n_i \|c_i - \bar{x}\|^2$$

3. Mencari Rasio dari hasil perhitungan SSW dan SSB.

$$\text{Rasio} = \frac{SSB}{SSW}$$

Setelah semua nilai Rasio Separasi R(i,j) dihitung, nilai DBI akhir diperoleh dengan mencari nilai maksimum dari R(i,j) untuk setiap *cluster* dan merata-ratakannya, seperti ditunjukkan pada Tabel 6.

4. proses perhitungan tahap akhir menghitung persamaan berikut untuk memperoleh nilai DBI.

$$s_i = \frac{1}{|C_i|} \sum_{x \in C_i} d(x, c_i)$$

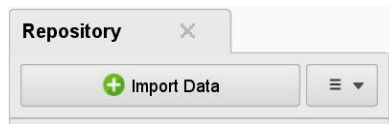
Tabel 6. Perhitungan DBI

i	j	SSWi	SSWj	SSB(i,j)	R(i,j)=(SSWi+SSWj)/SSB(i,j)	max _{j≠i} {R(i,j)}
0	1	16.549	20.493	60.017	0.617	0.617
0	2	16.549	33.551	153.681	0.326	
1	0	20.493	16.549	60.017	0.617	0.617
1	2	20.493	33.551	93.94	0.575	
2	0	33.551	16.549	153.681	0.326	0.575
2	1	33.551	20.493	93.94	0.575	
Hasil Akhir						∑maxR(i,j)=1,809
DBI						0,603

Semakin kecil nilai DBI, semakin baik kualitas *Clustering* yang dihasilkan, menunjukkan *cluster* yang padat dan terpisah satu sama lain [14].

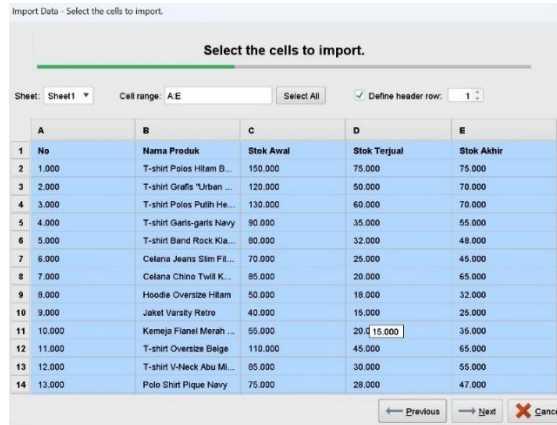
3.4 Penerapan K-Means Clustering Menggunakan Rapidminer

1. Pada gambar 3 data dari file Excel di masukan ke dalam (*workflow*) Rapid Miner untuk diproses [8].



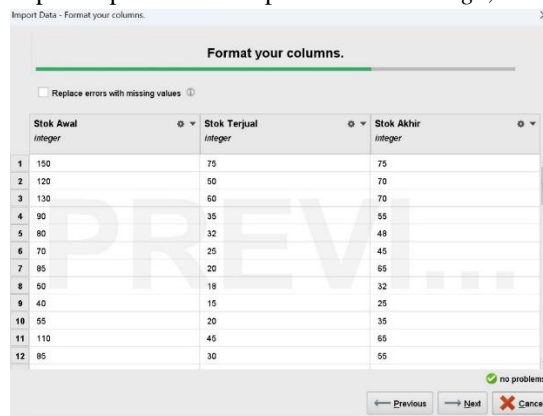
Gambar 3. Import Data

2. Pada gambar 4 tampilan proses setelah data di masukan kedalam rapid miner untuk di pilih *cell range* nya.



Gambar 4. Tampilan Data

3. Pada gambar 5 menampilkan proses setelah pemilihan *cell range*, dan yang akan diproses oprator.



Gambar 5. Data yang akan di pakai

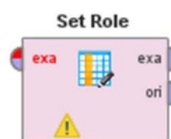
4. Pemanggilan Operator

Langkah selanjutnya menambahkan operator *retrieve* untuk membaca data set yang telah dimasukan sebelumnya, terlihat pada gambar 6 [8].



Gambar 6. Retrive

5. Pada gambar 7 oprator yang akan di masukan kedalam *rapid miner* adalah *set role* yang digunakan untuk melabelkan data yang akan di olah menggunakan rapid miner.



Gambar 7. Set Role

6. Setelah menetapkan *set role* kemudian oprator selanjutnya yang akan kita panggil adalah *K-Means* terlihat pada gambar 8 yang akan mengelompokan atau membuat *cluster*.



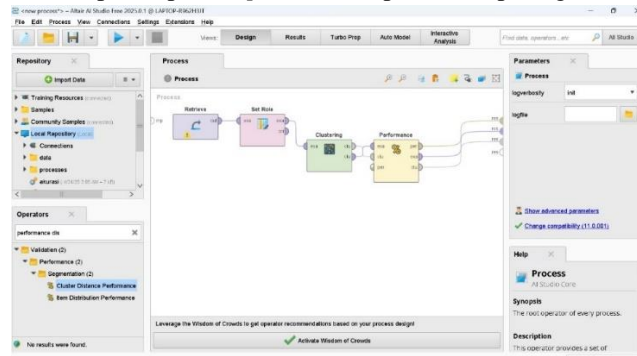
Gambar 8. K-Means

7. Pada gambar 9 terlihat operator *cluster distance performance* untuk menentukan hasil centeroid.



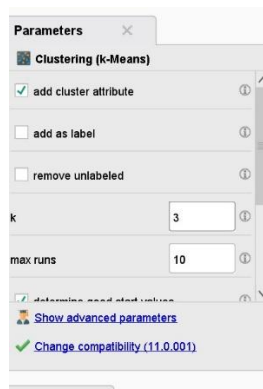
Gambar 9. Cluster Performance Distance

8. Tampilan keseluruhan operator pada *rapid miner* dapat dilihat pada gambar 10.



Gambar 10. Lembar kerja operator yang di gunakan pada *rapid miner*

9. Sebelum menjalankan proyek ubah jumlah *cluster* yang akan diproses dalam proyek dapat dilihat pada gambar 11.



Gambar 11. Jumlah Cluster

10. Penampilan Hasil Menjalankan *Rapid Miner*

Setelah semua sesuai jalankan proyek yang telah kita susun opratornya. Pada gambar 12 menampilkan gambar *runing* atau menjalankan proyek yang kita buat.



Gambar 12. Run Proses

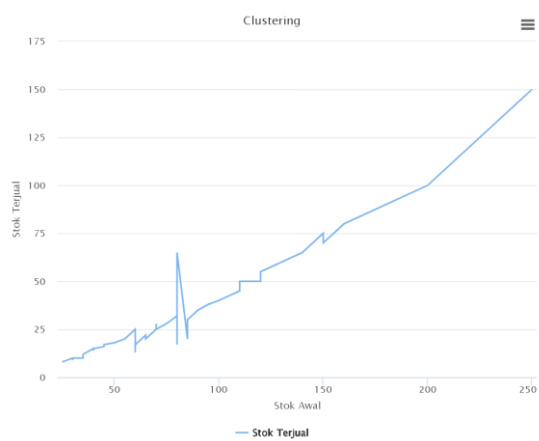
11. Setelah menjalankan proyek maka akan tampil hasil dari rapid miner seperti pada gambar 13.

Cluster Model

Cluster 0: 40 items
 Cluster 1: 16 items
 Cluster 2: 4 items
 Total number of items: 60

Gambar 13. Hasil Cluster

12. Dari hasil cluster model di dapatkan 3 cluster produk yang terlaris, laris dan yang kurang laris. Produk yang terlaris mendapatkan 4, yang laris 16 produk sedangkan yang kurang laris mendapatkan 40 produk.



Gambar 14. Hasil Grafik Rapid Miner

13. Penentuan jumlah nilai K Optimal yang dilakukan menggunakan operator cluster distance performance untuk menghasilkan nilai perbandingan dari nilai Davis Bouldin. Hasil operasi tersebut menghasilkan nilai 0,602.

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 0.599
Avg. within centroid distance_cluster_0: 0.602
Avg. within centroid distance_cluster_1: 0.601
Avg. within centroid distance_cluster_2: 0.595
Davies Bouldin: 0.602
```

Gambar 15. Performance Vector

Semakin kecil nilai DBI, semakin baik kualitas Clustering yang dihasilkan, menunjukkan cluster yang padat dan terpisah satu sama lain [15].

4. KESIMPULAN

Penelitian ini mengimplementasikan metode K-Means Clustering untuk mengeliminasi aspek subjektivitas dalam pengelompokan performa produk. Hasil analisis menunjukkan bahwa Cluster 2 (Sangat Laris) didominasi oleh produk dengan volume penjualan tertinggi, antara lain *T-shirt Persib Edisi Juara 2024* (150 unit), *T-shirt Persib Home Jersey Style* (100 unit), dan *T-shirt Persib Away Jersey Style* (90 unit). Temuan ini mengindikasikan bahwa produk bertema klub sepak bola lokal merupakan penggerak ekonomi utama pada Distro Engkus, dengan rata-rata penjualan yang signifikan melampaui Cluster 0 (Kurang Laris) yang hanya mencapai 17,28 unit. Validasi kualitas pengelompokan menggunakan nilai Davies-Bouldin Index (DBI) sebesar 0,392 mengonfirmasi bahwa pemisahan antar klaster memiliki performa yang sangat signifikan. Hal ini membuktikan adanya diferensiasi yang objektif antara komoditas yang bersifat *fast-moving* dengan stok yang memiliki perputaran lambat (*slow-moving*). Sebagai implikasi strategis, direkomendasikan penguatan manajemen inventori melalui penambahan stok awal pada produk di Cluster 2 secara periodik guna memitigasi risiko *stock-out*. Sebaliknya, terhadap produk dalam Cluster 0, seperti varian kemeja dan jaket tertentu, diperlukan langkah akselerasi likuidasi modal melalui strategi promosi agresif, skema *bundling*, atau penyesuaian harga (diskon). Untuk pengembangan penelitian selanjutnya, integrasi data ke dalam sistem *Point-*

of-Sale (POS) serta penambahan variabel demografi pelanggan sangat disarankan guna membangun model prediktif yang lebih akurat dalam memproyeksikan tren permintaan masa depan[14].

REFERENSI

- [1] H. R. Y. Idham, "IMPLEMENTASI RAPIDMINER UNTUK CLESTERING DATA PENJUALAN PAKAIAN MENGGUNAKAN METODE *K-MEANS*," *Journal Education and Technology*, vol. 5, pp. 221-231, 2024.
- [2] N. N. d. A. M. M. M. Wijana, "Analisis Kualitas Layanan dan Fasilitas terhadap Kepuasan Siswa Menggunakan Data Mining," *Jurnal Maps (Manajemen Perbankan Syariah)*, vol. 8, pp. 115-123, 2025.
- [3] D. P. d. S. S. Y. N. Rumapea, "Analisis Segmentasi Pelanggan Ritel Online Menggunakan *K-Means Clustering* Berdasarkan Model Recency, Frequency, Monetary (RFM)," *Jurnal Sains dan Teknologi*, vol. 6, pp. 292-299, 2024.
- [4] A. I. G. A. D, "Visualisasi Segmentasi Pelanggan Berdasarkan Atribut RFM Menggunakan Algoritma *K-Means* untuk Memahami Karakteristik Pelanggan pada Toko Retail Online," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 12, pp. 283-292, 2025.
- [5] A. W. Sukri Illaihi Wahyudi, "Implementasi Metode *K-Means Clustering* Untuk Pengelompokan Data Stok Produk Toko Online Perdagangan Kaos," *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)*, pp. 470-478, 2022.
- [6] N. S. N. R. N. Afiasari, "Implementasi Data Mining Transaksi Penjualan Menggunakan Algoritma *Clustering* dengan Metode *K-Means*," *Jurnal SAINTEKOM*, vol. 13, pp. 100-110, 2023.
- [7] d. K. U. M. H. Fakhriza, "Analisis Produk Terlaris Menggunakan Metode *K-Means Clustering* pada PT. Sukanda Djaya," *JIKA (Jurnal Informatika)*, vol. 5, pp. 8-15, 2021.
- [8] B. i. Febri Fahrizal, "ANALISIS PRODUK TERLARIS DAN PENGUJIAN *K-MEANS* UNTUK "UMKM CETOM"," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, pp. 3055 - 3061, 2024.
- [9] N. R. S. D. Prasetiani, "Penerapan Data Mining Untuk *Clustering* Menu Favorit Menggunakan Algoritma *K-Means* (Studi Kasus Kedai Expo)," *JINACS (Journal of Informatics and Computer)*, vol. 3, pp. 278-286, 2022.
- [10] A. R. S. Dewi, "Klasterisasi Pemilihan Paket Umrah Berdasarkan Musim untuk Menentukan Strategi Promosi Menggunakan Algoritma *K-Means*," *Jurnal Dimamu*, vol. 4, pp. 83-96, 2024.
- [11] S. Meirisah, "ANALISA KASUS KEJAHATAN SIBER DENGAN MENGGUNAKAN VISUALISASI DATA," *(Jurnal Informatika Teknologi dan Sains)*, vol. 5, pp. 32-37, 2023.
- [12] S. H. Fathurrahman, "EVALUASI *CLUSTERING K-MEANS* DAN *K-MEDOID* PADA PERSEBARAN COVID-19 DI INDONESIA DENGAN METODE *DAVIES-BOULDIN INDEX (DBI)*," *Jurnal MNEMONIC*, vol. 6, pp. 117-128, 2023.
- [13] H. R. Y. Idham, "IMPLEMENTASI RAPIDMINER UNTUK CLESTERING DATA PENJUALAN PAKAIAN MENGGUNAKAN METODE *K-MEANS*," *Journal Education and Technology*, vol. 5, pp. 221-231, 2024.
- [14] N. N. d. A. M. M. M. Wijana, "Analisis Kualitas Layanan dan Fasilitas terhadap Kepuasan Siswa Menggunakan Data Mining," *Jurnal Maps (Manajemen Perbankan Syariah)*, vol. 8, pp. 115-123, 2025.
- [15] D. P. d. S. S. Y. N. Rumapea, "Analisis Segmentasi Pelanggan Ritel Online Menggunakan *K-Means Clustering* Berdasarkan Model Recency, Frequency, Monetary (RFM)," *Jurnal Sains dan Teknologi*, vol. 6, pp. 292-299, 2024.