



## IMPLEMENTASI RAPIDMINER UNTUK CLESTERING DATA PENJUALAN PAKAIAN MENGGUNAKAN METODE K-MEANS

Idham<sup>1\*</sup>, Herliana Rosika<sup>2</sup>, Yuliadi<sup>3</sup>

<sup>1</sup>Fakultas Teknik, Universitas Muhammadiyah Mataram

<sup>2</sup>Fakultas Teknik, Universitas Mataram

<sup>3</sup>Fakultas Rekayasa Sistem, Universitas Teknologi Sumbawa

Email: [idham@ummat.ac.id](mailto:idham@ummat.ac.id)\*, [herliana2014@gmail.com](mailto:herliana2014@gmail.com)2, [yuliadi@uts.ac.id](mailto:yuliadi@uts.ac.id)3

Informasi Artikel	ABSTRACT
<p>Riwayat artikel:            Disubmit : 1 Juni 2024            Direvisi : 16 Juni 2024            Diterima : 29 Juni 2024            Dipublikasi : 30 Juni 2024</p>	<p><i>In this clothing product sales research, it can be seen that sales transaction data is not well recorded. With this problem, we need to analyze group clothing categories. Grouping sales data focuses on applying the K-Means Clustering method to group sales transactions, to identify customer patterns and trends. The main aim of this study is to segment sales transactions to understand customer behavior better and help owners make more effective decisions. In this study, the K-Means Method is a manual calculation method, and its implementation uses RapidMiner. The application of the K-Means method succeeded in grouping sales transactions into several clusters. Each cluster exhibits different characteristics such as purchase frequency, transaction value, and product preferences. Clustering sales transactions using the K-Means method provides added value for owners in understanding customer behavior, marketing strategies, and market segmentation.</i></p>
<p>Keywords:            Method K-Means, Clustering, Rapidminer, Sales, Purchasing, Marketing</p>	
<p>Kata Kunci:            Metode k-means, clustering, rapidminer, penjualan, Pembelian, Pemasaran</p>	<p><b>ABSTRAK</b></p> <p>Pada penelitian penjualan produk pakaian ini, terlihat data transaksi penjualan tidak terdata dengan baik. Dengan adanya masalah ini maka kami perlu melakukan analisis untuk mengelompokkan katagori pakaian. Dalam proses pengelompokkan data penjualan berfokus pada penerapan metode K-Means Clustering untuk mengelompokkan transaksi penjualan, dengan tujuan mengidentifikasi pola dan tren pelanggan. Tujuan utama dalam penelitian ini adalah melakukan segmentasi transaksi penjualan guna memahami perilaku pelanggan dengan lebih baik dan membantu pemilik dalam pengambilan keputusan yang lebih efektif. Dalam penelitian ini, menggunakan Metode K-Means sebagai metode perhitungan secara manual adan dalam implementasinya menggunakan RapidMiner. Penerapan metode K-Means berhasil mengelompokkan transaski penjualan menjadi beberapa cluster. Setiap cluster menunjukkan karakteristik yang berbeda seperti frekuensi pembelian, nilai transaksi dan preferensi produk. Clustering transaksi penjualan menggunakan metode K-Means memberikan nilai tambah bagi pemilik dalam memahami perilaku pelanggan, strategi pemasaran dan segmentasi pasar.</p>



## PENDAHULUAN

Pada era digital perkembangan teknologi semakin berkembang khususnya pengelolaan data penjualan menjadi semakin krusial bagi bisnis untuk memahami perilaku konsumen, mengidentifikasi tren pasar, dan membuat keputusan strategis. Permasalahan yang terjadi dalam penelitian data transaksi penjualan adalah data tidak tercatat dan tersimpan dengan baik mulai data pembelian barang, stok barang sampai data transaksi penjualan. Untuk mengetahui tingkat transaksi penjualan petugas melihat kembali catatan transaksi penjualan pakaian, sehingga memerlukan metode analisis yang tepat agar dapat diolah dan dimanfaatkan secara efektif dan efisien. Salah satu metode yang efektif dalam analisis data transaksi penjualan adalah clustering, dimana data dikelompokkan berdasarkan karakteristik yang serupa (Normah et al. 2022).

*Clustering* merupakan metode yang digunakan untuk mengelompokkan membagi data menjadi beberapa kelompok atau cluster memiliki kemiripan.(Al-Rizki, Widaningrum, and Buntoro 2020). *Clustering* adalah proses mengelompokkan sekumpulan data sehingga data dalam satu kelompok atau *cluster* memiliki kemiripan yang tinggi.(Anggraeni, Saputra, and Noranita 2013). Tujuan dilakukan clustering adalah untuk menentukan pola atau struktur dalam data, mengelompokkan data ke dalam kelompok lebih kecil, serta mengurangi kompleksitas data dengan mengelompokkan data menjadi beberapa *cluster*.

Metode yang digunakan dalam menganalisis data adalah data mining. Data mining adalah proses pengumpulan dan pengolahan data yang besar untuk di ekstrak menjadi informasi yang lebih spesifik (Budiyasari et al. 2017). *Rapidminer* salah perangkat lunak yang bersifat terbuka (open source) yang dapat digunakan dalam memberikan sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi (Aprilla Dennis 2013). Output data mining dapat digunakan untuk membantu pengambilan keputusan di masa mendatang (Normah et al. 2022).

Metode K-Means adalah salah satu teknik *clustering* yang paling populer dalam data mining. K-means bekerja dengan membagi data ke dalam sejumlah K cluster, dimana K adalah jumlah cluster yang ditentukan sebelumnya. (Melpa Metisen and Latipa Sari 2015). K-Means bertujuan untuk meminimalkan variasi dalam cluster dengan menugaskan data *point* ke *cluster* terdekat berdasarkan jarak ke *centroid cluster*.



## METODE PENELITIAN

Data mining merupakan proses pencarian pengetahuan yang menarik data berukuran besar yang disimpan dalam basis data, *warehouse* atau tempat penyimpanan informasi lainnya (Rafi Nahjan, Nono Heryana, and Apriade Voutama 2023). Teknik data mining digunakan untuk memeriksa basis data berukuran besar sebagai cara untuk menentukan pola yang baru dan berguna.

Metode K-Means digunakan untuk melakukan clustering pada data transaksi penjualan. Metode penelitian terdiri dari beberapa tahap yang sistematis untuk memastikan data yang digunakan bersih, akurat, dan hasil clustering yang diperoleh relevan dan dapat digunakan untuk analisis lebih lanjut. Adapun tahapan yang dilakukan dalam penelitian ini adalah:

1. Pengumpulan data: data transaksi penjualan dikumpulkan dari sistem manajemen penjualan.
2. Pra-pemrosesan data: dilakukan untuk memastikan kualitas data yang ditinggi dan mengurangi noise.
3. Pemilihan fitur : pemilihan fitur yang relevan sangat penting untuk mendapatkan hasil *clustering* yang bermakna.
4. Penentuan jumlah *cluster*: menentukan cluster yang optimal merupakan Langkah penting dalam metode K-means.
5. Implementasi K-Means: algoritma k-means diterapkan pada data yang telah diproses.
6. Evaluasi hasil: memastikan kualitas dan relevansi *cluster* yang dihasilkan.
7. Analisis dan Interpretasi: setelah memperoleh hasil *clustering*, analisis lebih lanjut dilakukan untuk memahami karakteristik setiap *cluster*.

Kesimpulan dan rekomendasi: hasil penelitian ini di simpulkan dengan menyoroti temuan utama dan implikasi bisnis. Rekomendasi untuk perusahaan disusun berdasarkan wawasan yang diperoleh dari hasil *clustering*.

## HASIL DAN PEMBAHASAN

### a. Data Sheet

Data yang digunakan pada penelitian adalah data penjualan pakaian pada bulan februari 2024 sampai dengan April 2024 pada toko *Quache*. Data berikut adalah data *sheet* yang digunakan sebagai data penelitian. Berikut adalah datasheet dari penjualan produk pakaian pada toko *Quache* yang terdiri dari 9 atribut.

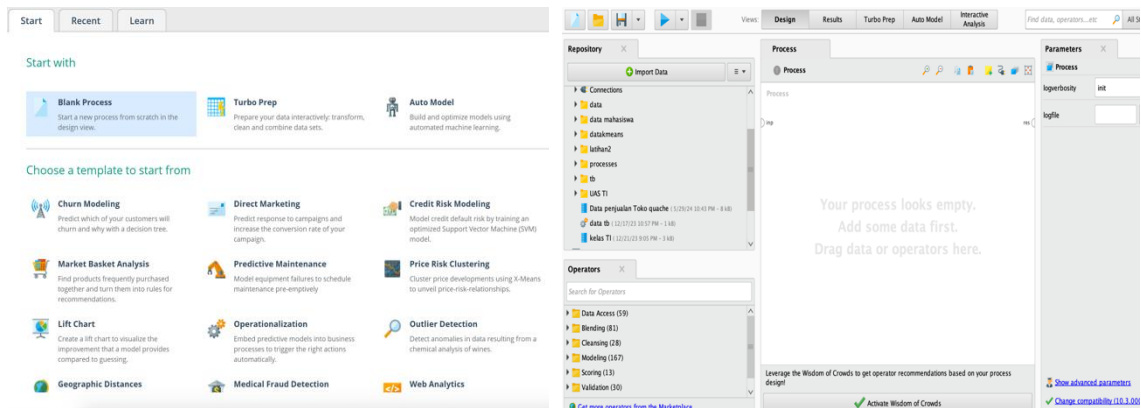


Tabel 1. Data Sheet Penjualan

No	Atribut	Deskripsi	Tipe
1	Kategori	Jenis kategori dari produk pakaian yang dijual	String
2	Jenis Produk	Jenis produk dari pakaian yang dijual	String
3	Nama Produk	Nama dari produk pakaian yang dial	String
4	Jumlah Awal	Total stok dari produk	Integer
5	Terjual	banyaknya produk yang telah terjual	Integer
6	Sisa Stok	Jumlah sisa dari stok pakain yang telah terjual	Integer
7	Harga Jual	harga jual produk	Integer
8	Total Penjualan	Jumlah yang diperoleh dari produk terjual	Integer
9	Status	Keterang apakah produk tersedia atau tidak	String

**b. Implementasi dan pengujian**

Dalam implementasi dan pengujian disini menggunakan sebuah *software* Rapidminer Studio versi 10, dengan pengujian data menggunakan *software* akan melakukan membandingkan bagaimana hasil pengahan data secara manula dengan hasil pengolahan data menggunakan *software*. Gambar 1 merupakan tampilan halaman utama dari aplikasi *rapidminer* studio versi 10.



Gambar 1. Tampilan New Proses Rapidminer



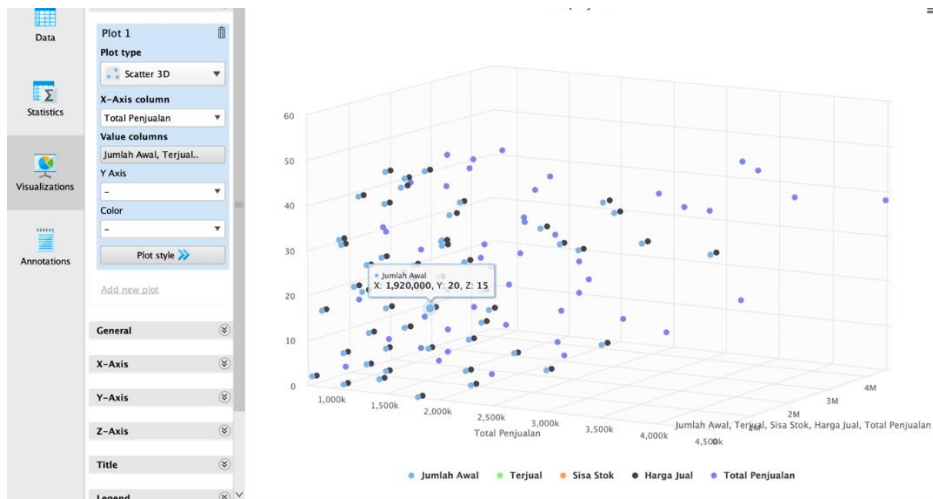
Result History ExampleSet (Retrieve Data penjualan Toko quache)

Open in Turbo Prep Auto Model Interactive Analysis Filter (51 / 51 examples): all

Row No.	Kategori	Jenis Produk	Nama Prod...	Jumlah Awal	Terjual	Sisa Stok	Harga Jual	Total Penju...	Status
1	Atasan	Jaket	Denim jacket	20	15	5	120000	1800000	Tersedia
2	Atasan	Kemeja	Kemeja leng...	10	10	0	110000	1100000	Tidak tersedia
3	Atasan	Kemeja	Kemeja leng...	12	8	4	100000	800000	Tersedia
4	Atasan	Kemeja	Kemeja leng...	15	12	3	120000	1440000	Tersedia
5	Atasan	Kemeja	Kemeja leng...	20	20	0	115000	2300000	Tidak tersedia
6	Atasan	Kaos	Kaos lengan...	20	15	5	100000	1500000	Tersedia
7	Atasan	Sweater	Sweater	24	12	12	110000	1320000	Tersedia
8	Bawahan	Rok	Rok Plisket	20	18	2	125000	2250000	Tersedia
9	Bawahan	Celana	Celana Jean...	15	10	5	110000	1100000	Tersedia
10	Bawahan	Celana	Celana Jean...	30	25	5	120000	3000000	Tersedia
11	Atasan	Kemeja	Tuxedo	20	15	5	100000	1500000	Tersedia
12	Bawahan	Celana	Celana plisket	25	20	5	95000	1900000	Tersedia
13	Atasan	Kaos	Kaos lengan...	35	30	5	90000	2700000	Tersedia
14	Atasan	Tunik	Tunik abaya	15	11	4	122000	1342000	Tersedia
15	Atasan	Tunik	Tunik vest	25	19	6	120000	2280000	Tersedia

Gambar 2. Data Sheet Penjualan

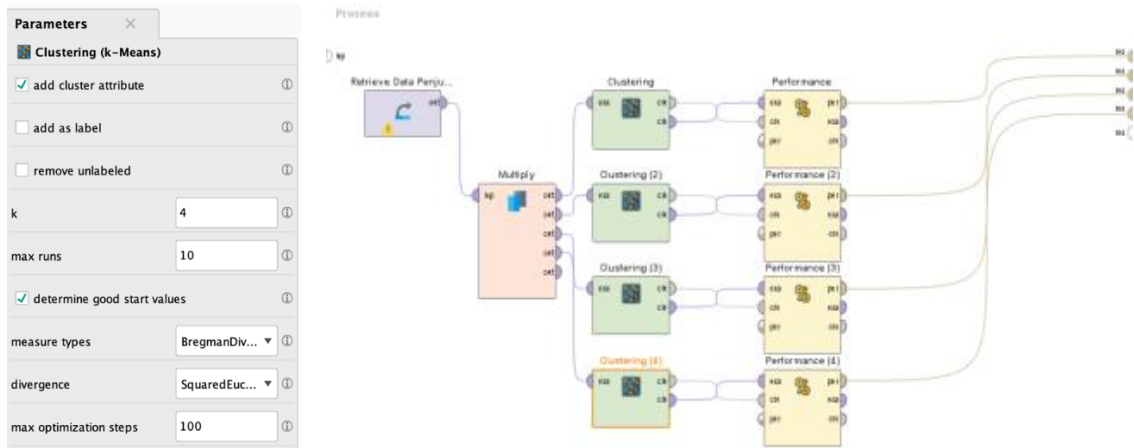
Pada gambar 2 adalah hasil *import* data yang berasal data data *excel* yang kemudian di proses sehingga menghasilkan data *sheet* penjualan.



Gambar 3. Visualization data penjualan

Pada gambar 3 terlihat hasil visualialisasi data penjualan berdasarkan hasil total penjualan. Dengan aribut jumlah awal, terjual, sisa stok, harga jual dan total penjualan. Pemilihan fitur, pada tahap ini, dilakukan pemilihan fitur yang relevan untuk mendapatkan hasil *clustering* yang bermakna. Adapun fitur yang digunakan adalah parameter multiplay, metode k-means untuk menentukan berapa jumlah kluster yang akan digunakan untuk mendapatkan hasil yang di paling tepat. Pada gambar 4. Model proses *clustering*, menggunakan parameter k-means sejumlah 4

clustering. Clustering 1 dengan K=4, clustering 2 dengan K= 5, clustering 3 dengan K=6, dan clustering 4 dengan K=8. Dengan max runs= 10.



Gambar 4. Model Proses clustering



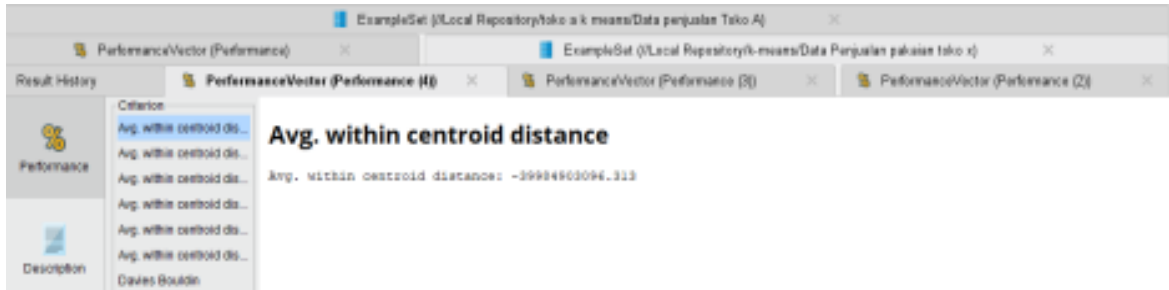
Gambar 5. hasil performance 1



Gambar 6. hasil performance 2

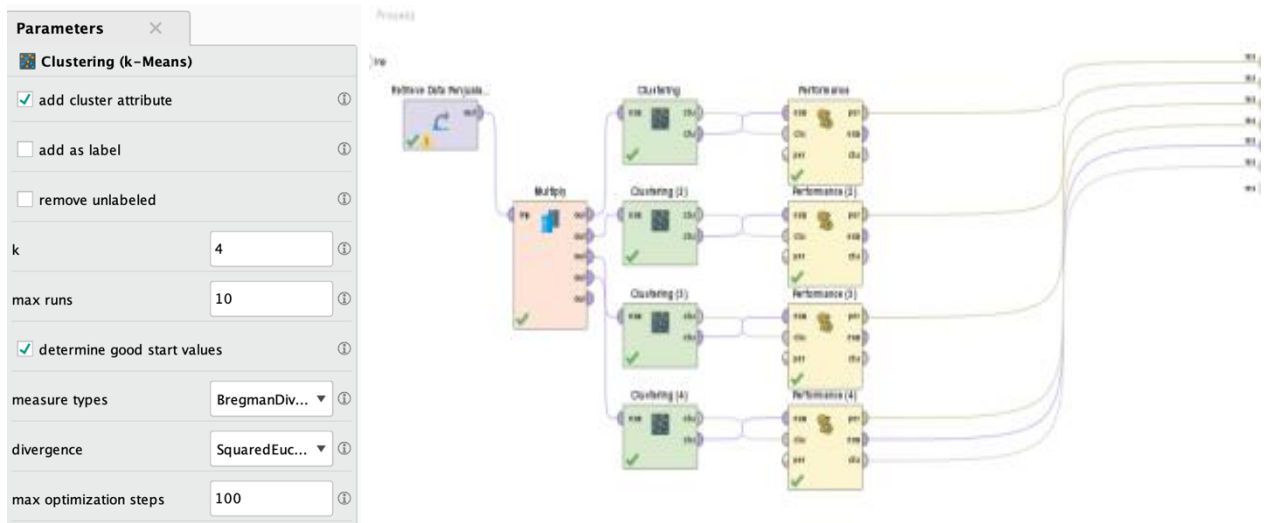


Gambar 7. hasil performance 3



Gambar 8. hasil performance 4

Pemberian *clustering performance* pada *rapidminer* merujuk pada evaluasi kinerja algoritma *clustering* yang digunakan adalah untuk mengelompokkan data. Berdasarkan data hasil *performance* diatas maka dapat kita simpulkan cluster yang digunakan adalah performace ke-4 karena nilai yang diperoleh paling kecil. Tahap selanjutnya, menentukan jumlah klaster berdasarkan data hasil *performance* pada gambar 5, 6, 7, dan 8 maka didapatkan.



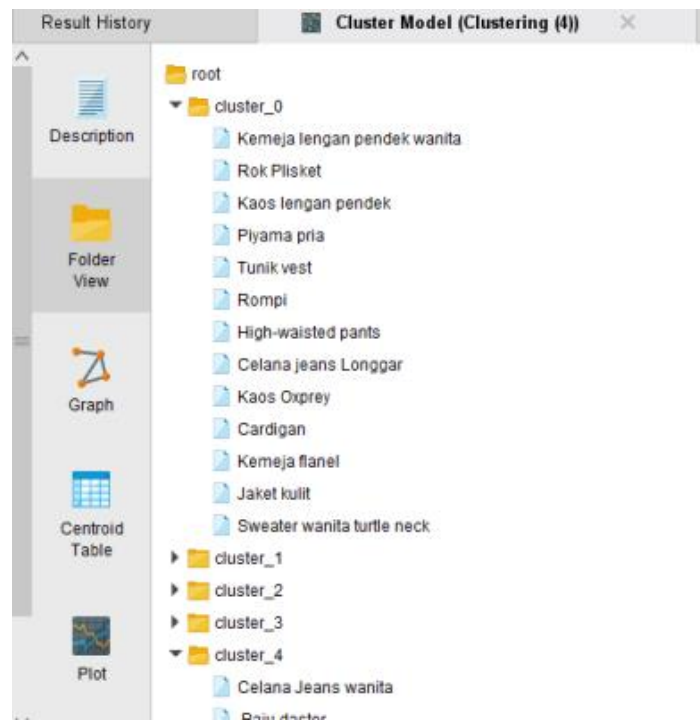
Gambar 9. Model Proses Hasil Performace Terkecil



Gambar 10 Tampilan View Text



Pada *cluster* model *clustering* dapat dilihat beberapa tampilan hasil *cluster*, yaitu *text view* yang merupakan tampilan hasil pengelompokan berdasarkan *cluster* dan jumlah anggotanya.

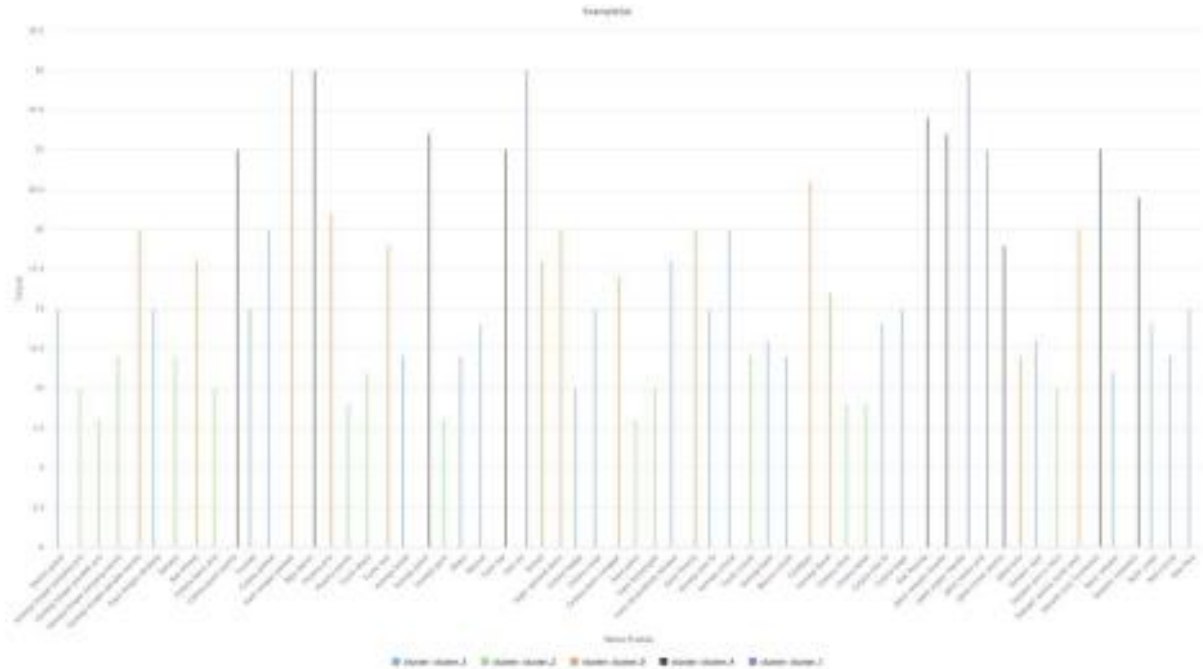


Gambar 12. Tampilan Folder View

Row No.	Nama Prodak	cluster	Jumlah Awal	Terjual	Sisa Stok	Harga Jual	Total Penjua...
1	Denim jacket	cluster_3	20	15	5	120000	1800000
2	Kemeja leng...	cluster_2	10	10	0	110000	1100000
3	Kemeja leng...	cluster_2	12	8	4	100000	800000
4	Kemeja leng...	cluster_2	15	12	3	120000	1440000
5	Kemeja leng...	cluster_0	20	20	0	115000	2300000
6	Kaos lengan ...	cluster_3	20	15	5	100000	1500000
7	Sweater	cluster_2	24	12	12	110000	1320000
8	Rok Plisket	cluster_0	20	18	2	125000	2250000
9	Celana Jean...	cluster_2	15	10	5	110000	1100000
10	Celana Jean...	cluster_4	30	25	5	120000	3000000
11	Tuxedo	cluster_3	20	15	5	100000	1500000
12	Celana plisket	cluster_3	25	20	5	95000	1900000
13	Kaos lengan ...	cluster_0	35	30	5	90000	2700000

Gambar 13. Tampilan hasil *cluster* pada data view





Gambar 14 Tampilan Hasil *Example Set Visualization*

Proses *clustering* data melibatkan serangkaian langkah yang dapat dijabarkan sebagai berikut. Awalnya, data diimpor dari lembar data menggunakan aplikasi Excel dan dimuat ke dalam repository. Setelah tahap import selesai, langkah selanjutnya adalah memilih metode clustering, dengan menggunakan metode K-Means pada kasus ini. Metode K-Means dipilih untuk membagi data menjadi beberapa cluster.

Dalam langkah-langkah K-Means, awalnya dipilih opsi *multiply* untuk melakukan pengelompokan data. Evaluasi cluster density performance dilakukan untuk mengukur kinerja sejumlah data *clustering*, memberikan gambaran tentang kepadatan dan konsistensi pembentukan *cluster*. Tahap ini diikuti oleh evaluasi performa umum untuk menilai perbedaan antara hasil *clustering* yang berbeda.

Proses berlanjut dengan pengujian *cluster distance performance* untuk mengukur jarak antar *cluster*. Jumlah *cluster distance performance* disesuaikan dengan jumlah clustering yang diinginkan. Dalam uji performa, parameter k (jumlah *cluster*) diubah sesuai dengan jumlah data yang akan di-*cluster*. Misalnya, *performance 1* dengan k=2, *performance 2* dengan k=3, *performance 3* dengan k=4, dan *performance 4* dengan k=5. Sebagai indikator evaluasi performa, digunakan nilai Davies Bouldin, di mana nilai yang lebih kecil dianggap sebagai performa yang lebih baik. Setelah mendapatkan hasil terbaik, hasil *clustering*, contoh, dan detail clustering



ditampilkan. Semua parameter dan performa terhubung dengan hasil akhir, memastikan bahwa data telah terklasifikasi secara optimal. Proses ini memungkinkan pengguna untuk memahami dan mengevaluasi cluster terbaik yang sesuai dengan karakteristik data yang dimiliki.

## SIMPULAN

Penelitian ini bertujuan untuk menganalisis penjualan produk pakaian toko menggunakan metode K-Means. Melalui studi literatur, dipahami bahwa data mining menjadi kunci untuk mengungkap pola-pola tersembunyi dalam data penjualan. Tahapan data mining, seperti preprocessing dan transformasi, menjadi langkah awal yang penting dalam mempersiapkan data sebelum dianalisis. Metode K-Means digunakan untuk mengelompokkan produk atau pelanggan dengan karakteristik serupa. Dengan menguji setiap komponen data, K-Means membentuk pusat cluster yang memungkinkan toko untuk menyesuaikan strategi pemasaran dengan lebih terarah. Proses ini membantu mengoptimalkan manajemen stok dan meningkatkan respons terhadap perubahan pasar.

## DAFTAR RUJUKAN

- Al-Rizki, Muhammad Farid Iqbal, Ida Widaningrum, and Ghulam Asrofi Buntoro. 2020. "Prediksi Penyebaran Penyakit TBC Dengan Metode K-Means Clustering Menggunakan Aplikasi Rapidminer." *JTERA (Jurnal Teknologi Rekayasa)* 5(1): 1.
- Anggraeni, Hapsari Dita, Ragis Saputra, and Beta Noranita. 2013. "Aplikasi Data Mining Analisis Data Transaksi Penjualan Obat Menggunakan Algoritma Apriori." *Journal of Informatics and Technology* 2(2): 22–28.  
[https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book\\_part](https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book_part).
- Aprilla Dennis. 2013. "Belajar Data Mining Dengan RapidMiner." *Innovation and Knowledge Management in Business Globalization: Theory & Practice, Vols 1 and 2* 5(4): 1–5.  
[http://esjournals.org/journaloftechnology/archive/vol1no6/vol1no6\\_6.pdf%5Cnhttp://www.airccse.org/journal/nsa/5413nsa02.pdf](http://esjournals.org/journaloftechnology/archive/vol1no6/vol1no6_6.pdf%5Cnhttp://www.airccse.org/journal/nsa/5413nsa02.pdf).
- Budiyasari, Vernanda Novrini et al. 2017. "Implementasi Data Mining Pada Penjualan Kacamata Dengan Menggunakan Algoritma Apriori." *Indonesian Journal on Computer and Information Technology* 2(2): 31–39.
- Melpa Metisen, Benri, and Herlina Latipa Sari. 2015. "Analisis Clustering Menggunakan Metode K-Means Dalam Pengelompokkan Penjualan Produk Pada Swalayan Fadhila." *Jurnal Media Infotama* 11(2): 110–18.
- Normah, Bakhtiar Rifai, Satrio Vambudi, and Rifki Maulana. 2022. "Analisa Sentimen Perkembangan



Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE.” *Jurnal Teknik Komputer AMIK BSI* 8(2): 174–80. <https://ejournal.bsi.ac.id/ejurnal/index.php/jtk/article/view/13041/pdf>.

Rafi Nahjan, Muhammad, Nono Heryana, and Apriade Voutama. 2023. “Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell.” *JATI (Jurnal Mahasiswa Teknik Informatika)* 7(1): 101–4.