



IMPLEMENTASI RAPIDMINER UNTUK KLASTERING DATA WHOLESALE CUSTOMER MENGGUNAKAN METODE K-MEANS

IMPLEMENTATION OF RAPIDMINER FOR WHOLESALE CUSTOMER DATA KLASTERING USING THE K-MEANS METHOD

Nizar Arfanni^{1*}, Hasbi Firmansyah²

Univeristas Pancasakti Tegal

Email: nizararfan22@gmail.com^{1*}, hasbifirmansyah@upstegal.ac.id²

Article Info

Article history :

Received : 17-12-2025

Revised : 18-12-2025

Accepted : 20-12-2025

Pulished : 22-12-2025

Abstract

This study aims to perform customer segmentation on wholesale customer data using the K-Means clustering method implemented in RapidMiner. Wholesale customer datasets often contain large and unstructured information, making it difficult for businesses to identify purchasing patterns and customer behavior. To address this issue, data preprocessing was conducted to ensure data cleanliness, followed by feature selection focusing on numerical attributes relevant to clustering. The K-Means algorithm was applied using several values of k to determine the most optimal cluster grouping, and the model performance was evaluated using internal validation metrics such as Davies-Bouldin Index. The results showed that the K-Means method successfully grouped wholesale customers into several distinct clusters, each characterized by different purchase volumes, spending behaviors, and product demand tendencies. These clusters provide meaningful insights for business decision-making, particularly in designing marketing strategies, managing inventory, and identifying high-value customer groups. Overall, the application of RapidMiner proved effective in simplifying the clustering workflow and generating interpretable segmentation results. This research demonstrates that K-Means clustering can be utilized as a supporting tool for wholesale business analytics and customer management optimization.

Keywords : Clustering, K-Means, RapidMiner

Abstrak

Penelitian ini bertujuan untuk melakukan segmentasi pelanggan pada data wholesale customer menggunakan metode klastering K-Means yang diimplementasikan melalui perangkat lunak RapidMiner. Dataset wholesale customer sering kali memiliki jumlah data yang besar dan bervariasi sehingga menyulitkan perusahaan dalam mengidentifikasi pola pembelian dan perilaku konsumen. Untuk mengatasi permasalahan tersebut, dilakukan proses pra-pemrosesan data, pemilihan atribut numerik yang relevan, serta pengujian beberapa nilai k untuk memperoleh jumlah cluster yang paling optimal. Evaluasi hasil klastering dilakukan menggunakan metrik internal seperti Davies-Bouldin Index. Hasil penelitian menunjukkan bahwa metode K-Means berhasil mengelompokkan pelanggan wholesale ke dalam beberapa cluster dengan karakteristik berbeda, seperti jumlah pembelian, total pengeluaran, dan kecenderungan permintaan produk. Informasi dari tiap cluster dapat dimanfaatkan sebagai dasar dalam penyusunan strategi pemasaran, pengelolaan persediaan, serta identifikasi pelanggan bernilai tinggi. Secara keseluruhan, penggunaan RapidMiner mempermudah proses analisis dan menghasilkan segmentasi yang lebih terstruktur. Penelitian ini menunjukkan bahwa metode K-Means dapat diterapkan sebagai alat pendukung pengambilan keputusan pada bisnis wholesale.

Kata Kunci: Klastering, K-Means, Rapid Miner



PENDAHULUAN

Pada era digital, perkembangan teknologi informasi telah membawa perubahan signifikan dalam cara bisnis mengelola data, khususnya data penjualan. Sejak awal 2000-an, transformasi digital mulai merambah sektor perdagangan, mendorong perusahaan untuk beralih dari pencatatan manual ke sistem berbasis data terstruktur (HASIPAH 2025). Pengelolaan data penjualan menjadi semakin krusial dalam memahami perilaku konsumen, mengidentifikasi tren pasar, dan mendukung pengambilan keputusan strategis yang berbasis bukti.

Teknik Klastering adalah metode analisis data tanpa label (unsupervised) yang mengelompokkan objek-objek serupa ke dalam kelompok (cluster) yang sama (Pane, Nasution, and Furqan 2024). K-Means merupakan salah satu algoritma klastering paling populer dan sederhana, yang bekerja secara iteratif dengan menentukan beberapa centroid awal dan menyesuaikannya hingga konvergen (Pane et al. 2024). Dalam konteks pemasaran, klastering dapat mengidentifikasi segmen pelanggan yang memiliki pola pengeluaran serupa. Semakin kecil jarak dalam masing-masing cluster dan semakin besar jarak antar cluster, semakin baik hasil klastering.

Namun, dalam praktiknya, banyak bisnis skala menengah seperti grosir (wholesale) masih menghadapi kendala dalam pencatatan transaksi. Data pembelian barang, stok, hingga transaksi penjualan sering kali tidak tercatat dengan baik, menyebabkan kesulitan dalam evaluasi performa penjualan dan perencanaan stok (Annisyah et al. 2025). Petugas penjualan kerap harus menelusuri ulang catatan transaksi secara manual untuk mengetahui tingkat penjualan pakaian atau produk lainnya, yang tentu tidak efisien.

Untuk mengatasi permasalahan tersebut, diperlukan metode analisis yang mampu mengelompokkan data secara sistematis agar dapat diolah dan dimanfaatkan secara efektif. Salah satu metode yang terbukti efisien dalam analisis data penjualan adalah klastering, khususnya algoritma K-Means. Metode ini memungkinkan pengelompokan data berdasarkan karakteristik yang serupa, seperti frekuensi pembelian, jumlah transaksi, atau jenis produk (Kiagus Ardelansyah Pratama, Zufar Rifa, Nur Aini Setiyawati 2024). Dengan bantuan tools seperti RapidMiner, proses klastering dapat dilakukan secara visual dan modular, memudahkan interpretasi hasil oleh tim bisnis maupun teknisi data.

Penelitian ini bertujuan untuk menerapkan algoritma K-Means dalam mengelompokkan data transaksi penjualan pada pelanggan grosir (wholesale customer), sehingga dapat membantu perusahaan dalam menentukan strategi pengadaan barang dan segmentasi pasar yang lebih akurat.

METODE PENELITIAN

Data transaksi penjualan grosir yang tersedia secara publik, seperti pada UCI Machine Learning Repository, sering kali belum dianalisis secara optimal untuk memahami perilaku pelanggan. Padahal, data tersebut menyimpan informasi penting terkait pola pembelian berdasarkan kategori produk seperti Fresh, Milk, Grocery, dan lainnya. Oleh karena itu, penelitian ini dilakukan untuk menerapkan algoritma K-Means Klastering guna mengelompokkan pelanggan grosir berdasarkan karakteristik pembelian mereka, sehingga dapat digunakan untuk strategi pemasaran dan pengadaan barang yang lebih tepat sasaran (Nugraha et al. 2025).



Data diperoleh dari UCI Machine Learning Repository, yang berisi 440 data pelanggan grosir dengan delapan atribut numerik: Channel, Region, Fresh, Milk, Grocery, Frozen, Detergents_Paper, dan Delicassen. Data bersifat open-source dan digunakan sebagai simulasi data transaksi penjualan untuk keperluan analisis.

Analisis dilakukan menggunakan algoritma K-Means Klastering di RapidMiner. Tahapan analisis meliputi:

1. Pra-pemrosesan data: memastikan format numerik konsisten dan menghilangkan outlier.
2. Pemilihan fitur: seluruh atribut digunakan karena sudah dalam bentuk numerik dan relevan untuk segmentasi pelanggan.
3. Penentuan jumlah cluster: dilakukan dengan evaluasi Davies-Bouldin Index untuk memilih jumlah cluster optimal.
4. Implementasi K-Means: algoritma diterapkan pada data yang telah diproses menggunakan RapidMiner.
5. Evaluasi hasil: menilai kualitas cluster berdasarkan jarak antar centroid dan kepadatan data.
6. Analisis dan interpretasi: memahami karakteristik tiap cluster untuk menyusun rekomendasi bisnis yang relevan.

HASIL DAN PEMBAHASAN

Data Sheet

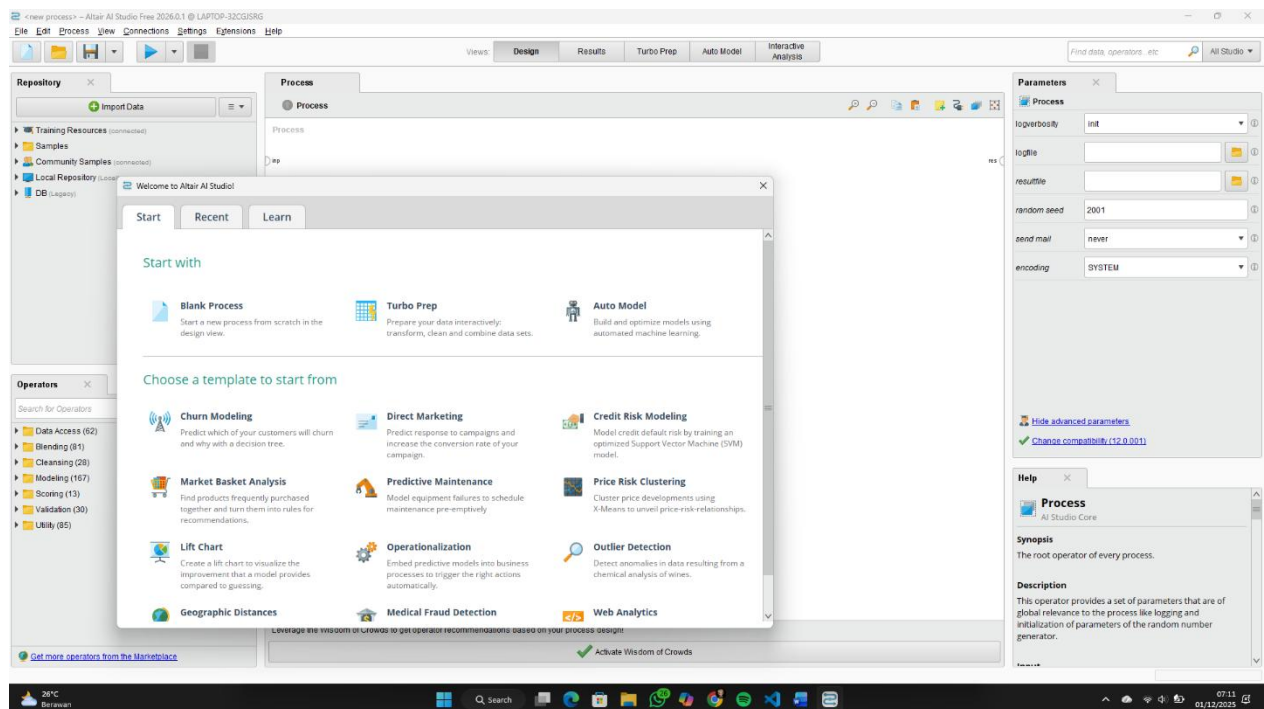
Data yang digunakan pada penelitian ini adalah data transaksi penjualan grosir yang diperoleh dari UCI Machine Learning Repository. Dataset ini berisi 440 pelanggan grosir dengan periode data yang bersifat historis. Setiap entri data memiliki delapan atribut numerik yang merepresentasikan kategori pembelian, yaitu Channel, Region, Fresh, Milk, Grocery, Frozen, Detergents_Paper, dan Delicassen. Data ini digunakan sebagai dasar untuk proses klastering dengan algoritma K-Means.

Table 1.Data Sheet Wholesale Customers

No	Attribute	Deskripsi	Type
1	Channel	Jenis saluran	Integer
2	Region	Wilayah pelanggan	Integer
3	Fresh	produk segar	Integer
4	Milk	produk susu	Integer
5	Grocery	Kebutuhan harian	Integer
6	Frozen	Produk beku	Integer
7	Detergents_Paper	Pembersih/kertas	Integer
8	Delicassen	Makanan olahan	Integer

Implementasi Dan Pengujian

Proses analisis data dilakukan menggunakan aplikasi RapidMiner. Gambar berikut menunjukkan tampilan New Process RapidMiner, yaitu antarmuka awal yang digunakan untuk membangun alur kerja (workflow) analisis. Pada tampilan ini, peneliti dapat menambahkan operator, mengatur parameter, serta menyusun langkah-langkah pemrosesan data secara modular.



Gambar 1.Tampilan Awal Rapid Miner

Data yang digunakan dalam penelitian ini adalah dataset Wholesale Customers dari UCI Machine Learning Repository. Dataset ini terdiri dari 440 baris data pelanggan grosir dengan 8 atribut numerik yang merepresentasikan kategori pembelian. Data ini digunakan sebagai dasar implementasi algoritma K-Means Klastering untuk mengelompokkan pelanggan berdasarkan pola pembelian.



ExampleSet (/Local Repository/data/Wholesale customers data)

Filter (440 / 440 examples): all

Row No.	Channel	Region	Fresh	Milk	Grocery	Frozen	Detergents_...	Delicassen
1	2	3	12669	9656	7561	214	2674	1338
2	2	3	7057	9810	9568	1762	3293	1776
3	2	3	6353	8808	7684	2405	3516	7844
4	1	3	13265	1196	4221	6404	507	1788
5	2	3	22615	5410	7198	3915	1777	5185
6	2	3	9413	8259	5126	666	1795	1451
7	2	3	12126	3199	6975	480	3140	545
8	2	3	7579	4956	9426	1669	3321	2566
9	1	3	5963	3648	6192	425	1716	750
10	2	3	6006	11093	18881	1159	7425	2098
11	2	3	3366	5403	12974	4400	5977	1744
12	2	3	13146	1124	4523	1420	549	497
13	2	3	31714	12319	11757	287	3881	2931
14	2	3	21217	6208	14982	3095	6707	602
15	2	3	24653	9465	12091	294	5058	2168
16	1	3	10253	1114	3821	397	964	412
17	2	3	1020	8816	12121	134	4508	1080
18	1	3	5876	6157	2933	839	370	4478
19	2	3	18601	6327	10099	2205	2767	3181
20	1	3	7780	2495	9464	669	2518	501
21	2	3	17546	4519	4602	1056	2259	2124
22	1	3	5567	871	2010	3383	375	569
23	1	3	31276	1917	4469	9408	2381	4334
24	2	3	26373	36423	22019	5154	4337	16523
25	2	3	22647	9776	13792	2915	4482	5778

ExampleSet (440 examples, 0 special attributes, 8 regular attributes)

Gambar 2.Data Sheet Wholesale Customers

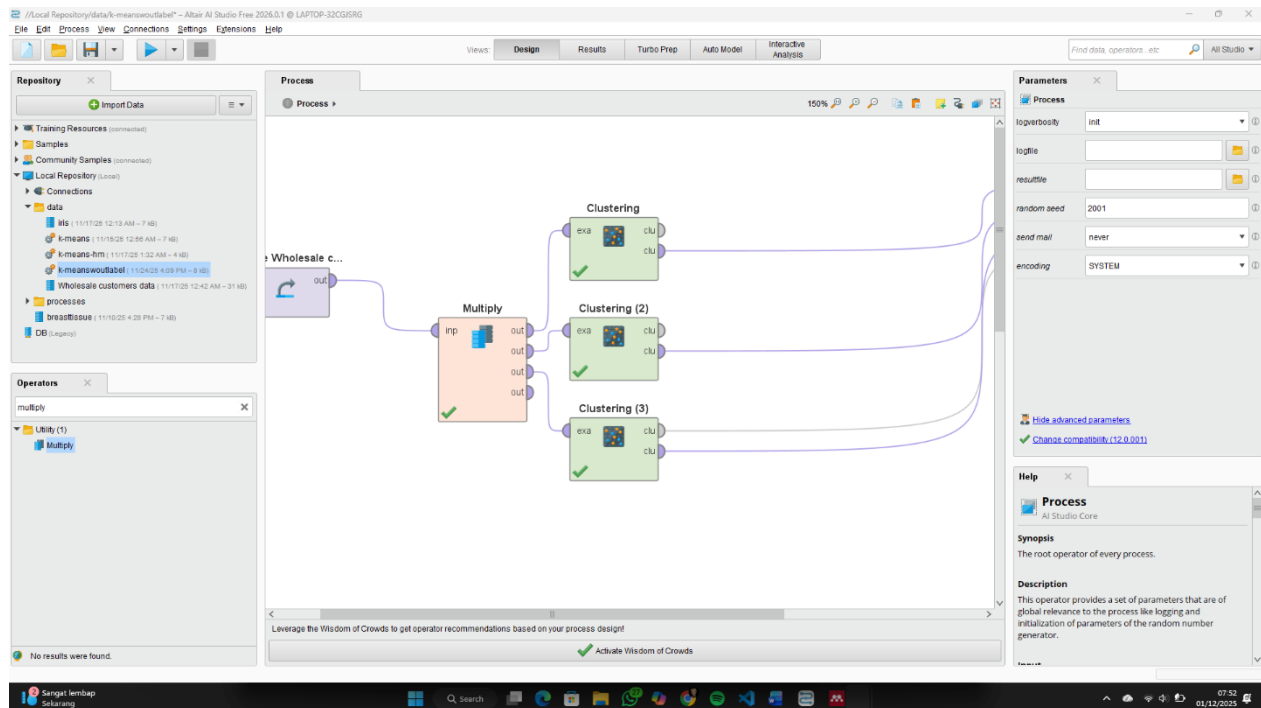


Gambar 3.Operator Multiply

Setelah data berhasil di-import ke RapidMiner, tahap berikutnya adalah menambahkan operator Multiply. Operator ini berfungsi untuk menduplikasi dataset sehingga dapat digunakan pada beberapa proses analisis secara paralel. Dengan adanya Multiply, peneliti dapat melakukan pengujian lebih dari satu skenario klastering tanpa harus mengulang proses input data. Pendekatan ini juga digunakan dalam penelitian (Susilowati and Wicaksono 2024), yang menerapkan algoritma K-Means di RapidMiner untuk klasterisasi data pencari kerja, di mana operator Multiply



dimanfaatkan untuk membagi dataset ke beberapa jalur analisis sehingga proses evaluasi lebih efisien.



Gambar 4.Operator Clustering

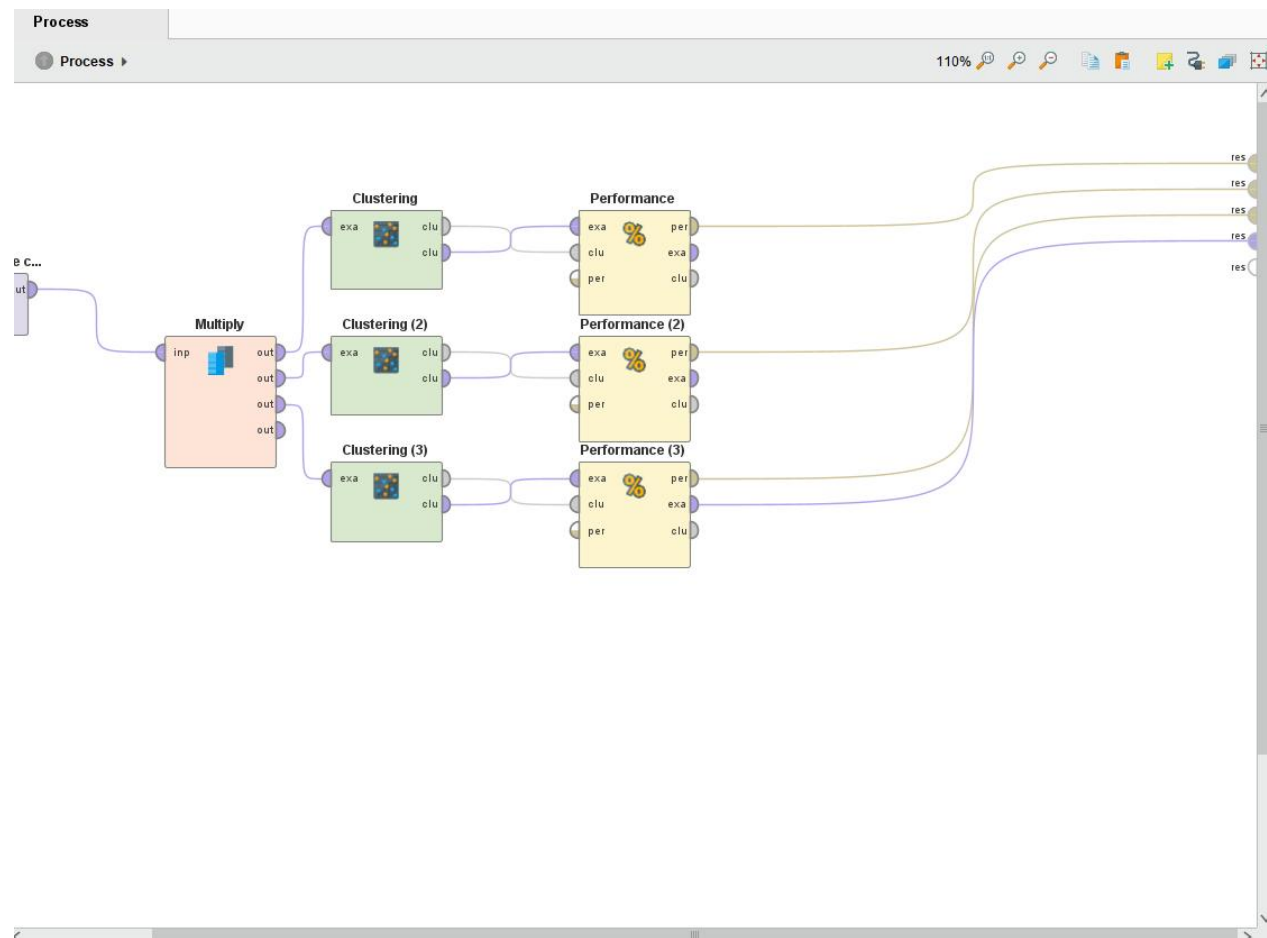
Tahap selanjutnya adalah penerapan algoritma K-Means Klastering. Pada operator ini ditentukan jumlah cluster (K) serta parameter max runs. Penelitian ini menggunakan beberapa variasi nilai K, yaitu K=2, K=4 dan K=6, dengan max runs = 10. Tujuan dari pengujian beberapa nilai K adalah untuk menemukan jumlah cluster yang paling optimal berdasarkan evaluasi performa. Pendekatan serupa dilakukan oleh (Faidah and Fatah 2025), yang menggunakan RapidMiner untuk mengelompokkan produk terlaris dengan K-Means, di mana variasi jumlah cluster diuji untuk mendapatkan hasil segmentasi yang paling relevan.

Tahap berikutnya adalah evaluasi hasil klastering menggunakan operator Performance. Operator ini menampilkan metrik evaluasi seperti Davies-Bouldin Index (DBI) dan Within-Cluster Distance untuk menilai kualitas cluster yang terbentuk. Evaluasi ini penting untuk menentukan jumlah cluster yang paling optimal, karena semakin kecil nilai DBI maka semakin baik kualitas pemisahan antar cluster, sedangkan nilai Within-Cluster Distance yang rendah menunjukkan homogenitas data dalam satu cluster.

Selain itu, evaluasi performa juga membantu peneliti dalam membandingkan hasil dari beberapa variasi nilai K yang telah diuji sebelumnya. Dengan cara ini, pemilihan jumlah cluster tidak hanya berdasarkan asumsi, tetapi didukung oleh metrik kuantitatif yang dapat dipertanggungjawabkan secara ilmiah. Penelitian oleh Irwansyah et al. (2025) juga menekankan penggunaan DBI dalam RapidMiner untuk menilai kualitas klastering pada data pendidikan dasar, di mana nilai DBI terkecil menunjukkan hasil klastering terbaik. Hal ini sejalan dengan praktik



umum dalam analisis data, di mana evaluasi performa menjadi tahap krusial untuk memastikan hasil klastering benar-benar merepresentasikan pola yang ada di dalam data.



Gambar 5.Operator Cluster Distance Performance

Hasil

Gambar di bawah menunjukkan hasil dari proses K-Means Klastering yang telah diterapkan pada dataset Wholesale Customers. Tampilan ini merupakan output dari tab Cluster Model di RapidMiner, yang menampilkan jumlah item dalam masing-masing cluster. Berdasarkan hasil tersebut, data terbagi ke dalam enam cluster dengan distribusi sebagai berikut: Cluster 0 berisi 286 item, Cluster 1 berisi 2 item, Cluster 2 berisi 5 item, Cluster 3 berisi 1 item, Cluster 4 berisi 62 item, dan Cluster 5 berisi 84 item, dengan total keseluruhan 440 data pelanggan.

Distribusi ini menunjukkan bahwa sebagian besar data terkonsentrasi pada Cluster 0, sementara Cluster 1 hingga Cluster 3 memiliki jumlah anggota yang sangat kecil. Hal ini dapat mengindikasikan adanya outlier atau kelompok pelanggan dengan pola pembelian yang sangat berbeda dari mayoritas. Visualisasi ini juga dilengkapi dengan fitur tambahan seperti Centroid Table, Plot, dan Annotations yang dapat digunakan untuk analisis lebih lanjut terhadap karakteristik masing-masing cluster.

Penelitian oleh (Florentin and Santoso 2022) menunjukkan bahwa visualisasi Cluster Model dalam RapidMiner sangat membantu dalam mengidentifikasi segmentasi pelanggan dan mendeteksi



anomali data, terutama ketika jumlah anggota dalam suatu cluster jauh lebih kecil dibandingkan cluster lainnya.

Cluster Model

Description

Cluster 0: 286 items
 Cluster 1: 2 items
 Cluster 2: 5 items
 Cluster 3: 1 items
 Cluster 4: 62 items
 Cluster 5: 94 items
 Total number of items: 440

Folder View

Graph

Centroid Table

Plot

Annotations

Gambar 6.Cluster Model

Row No.	id	cluster	Channel	Region	Fresh	Milk	Grocery	Frozen	Detergents...	Delicassan
1	1	cluster_0	2	3	12659	9656	7561	214	2674	1338
2	2	cluster_0	2	3	7057	9810	9568	1762	3293	1776
3	3	cluster_0	2	3	6353	8808	7884	2405	3516	7844
4	4	cluster_0	1	3	13265	1196	4221	6404	507	1788
5	5	cluster_4	2	3	22615	5410	7198	3915	1777	5185
6	6	cluster_0	2	3	9413	8259	5126	666	1795	1451
7	7	cluster_0	2	3	12126	3199	6975	480	3140	545
8	8	cluster_0	2	3	7579	4956	9426	1669	3321	2566
9	9	cluster_0	1	3	5963	3648	6192	425	1716	750
10	10	cluster_5	2	3	6006	11093	18881	1159	7425	2098
11	11	cluster_5	2	3	3366	5403	12974	4400	5977	1744
12	12	cluster_0	2	3	13146	1124	4523	1420	549	497
13	13	cluster_4	2	3	31714	12319	11757	287	3881	2931
14	14	cluster_4	2	3	21217	6208	14982	3095	6707	602
15	15	cluster_4	2	3	24653	9465	12091	284	5058	2108
16	16	cluster_0	1	3	10253	1114	3821	397	964	412
17	17	cluster_5	2	3	1020	8816	12121	134	4508	1080
18	18	cluster_0	1	3	5876	6157	2933	839	370	4478
19	19	cluster_0	2	3	18601	6327	10099	2205	2767	3181
20	20	cluster_0	1	3	7780	2495	9464	669	2518	501
21	21	cluster_0	2	3	17546	4519	4602	1066	2259	2124
22	22	cluster_0	1	3	5567	871	2010	3383	375	569
23	23	cluster_4	1	3	31276	1917	4459	9408	2381	4334
24	24	cluster_5	2	3	26373	36423	22019	5154	4337	16523

ExampleSet (440 examples, 2 special attributes, 8 regular attributes)

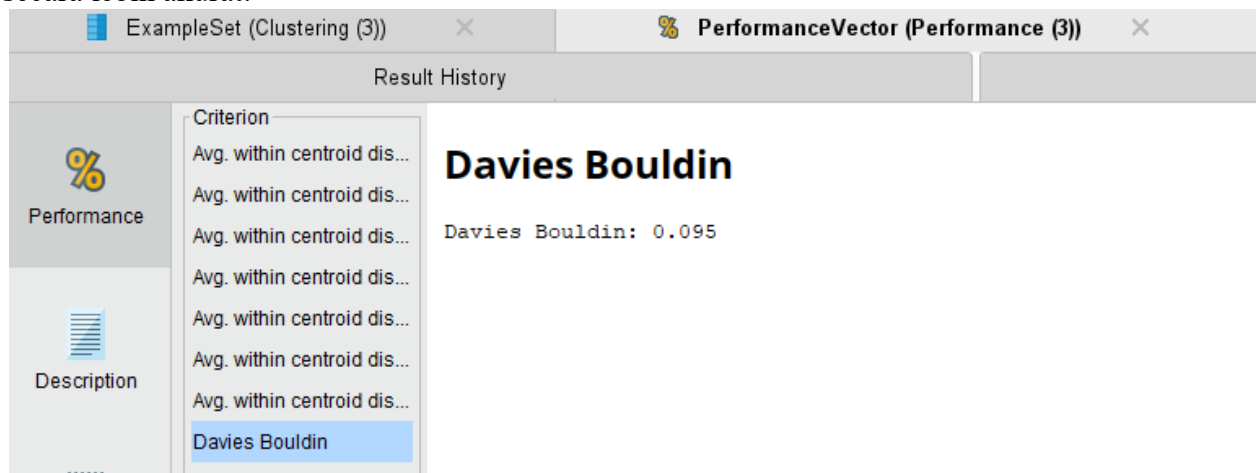
Gambar 7.Example Set

Gambar di atas menunjukkan hasil Example Set dari proses klastering menggunakan algoritma K-Means di RapidMiner. Tampilan ini memperlihatkan setiap record pelanggan yang telah dikelompokkan ke dalam cluster tertentu, lengkap dengan atribut input dan label cluster hasil pengelompokan. Dengan adanya Example Set, peneliti dapat melakukan analisis lebih mendalam

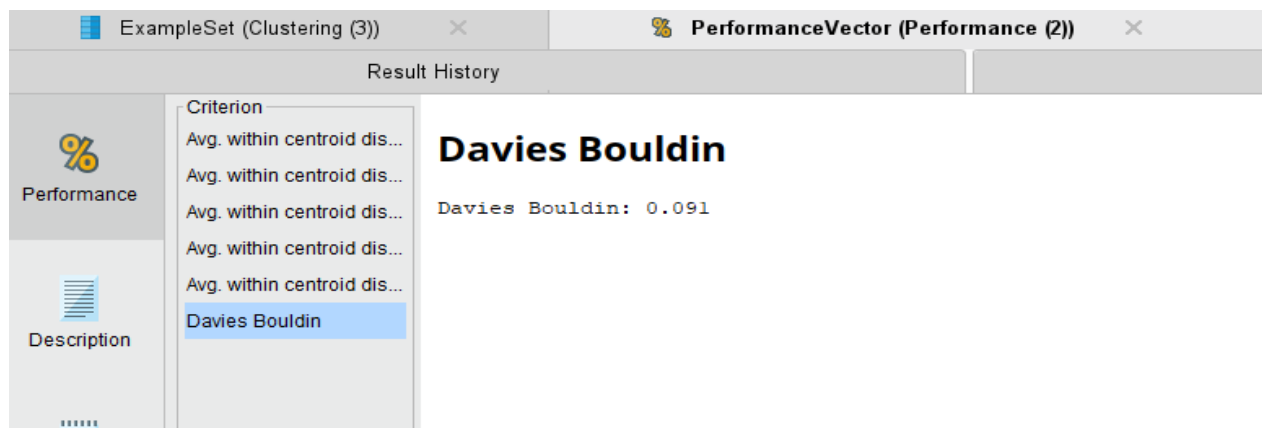


terhadap karakteristik masing-masing anggota cluster, serta memverifikasi apakah hasil pengelompokan sesuai dengan pola pembelian yang diharapkan.

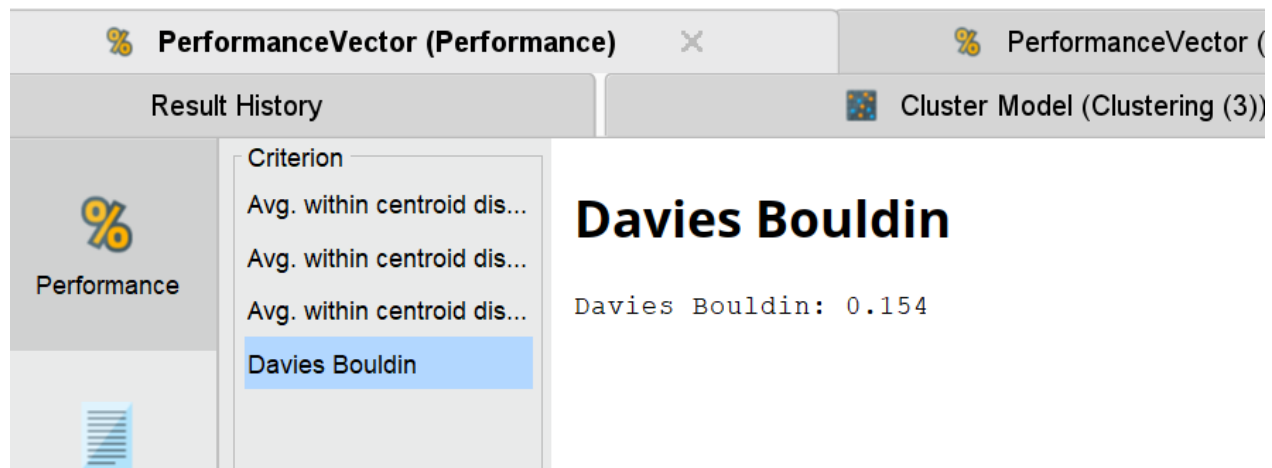
Selain itu, tampilan ini memungkinkan identifikasi outlier atau data yang tidak sesuai dengan mayoritas pola dalam cluster. Misalnya, jika terdapat pelanggan dengan nilai pembelian yang ekstrem namun masuk ke cluster yang padat, maka perlu dilakukan evaluasi ulang terhadap parameter klustering. Example Set juga menjadi dasar untuk analisis lanjutan seperti profiling pelanggan, pengembangan strategi pemasaran, atau penentuan prioritas layanan. Penelitian oleh (Informatika et al. 2022) menunjukkan bahwa penggunaan Example Set dalam RapidMiner sangat membantu dalam proses klasifikasi gangguan jaringan, di mana setiap entri data dapat ditelusuri dan dianalisis berdasarkan cluster yang terbentuk untuk mendukung pengambilan keputusan teknis secara lebih akurat.



Gambar 8.Hasil Performance 3



Gambar 9.Hasil Performance 2



Gambar 10.Hasil Performance 1

Pada gambar 8,9,10 menunjukkan hasil evaluasi klastering menggunakan operator Performance Vector di RapidMiner. Evaluasi dilakukan terhadap tiga skenario klastering dengan variasi jumlah cluster, dan metrik utama yang digunakan adalah Davies-Bouldin Index (DBI). Nilai DBI digunakan untuk menilai kualitas pemisahan antar cluster, di mana semakin kecil nilai DBI maka semakin baik hasil klastering.

Berdasarkan hasil yang diperoleh, Performance 2 menghasilkan nilai DBI terkecil yaitu 0.091, diikuti oleh Performance 3 dengan DBI 0.095, dan Performance 1 dengan DBI 0.154. Hasil ini menunjukkan bahwa skenario pada Performance 2 memiliki kualitas klastering terbaik karena pemisahan antar cluster lebih optimal dan homogenitas dalam cluster lebih tinggi. Perbedaan nilai DBI ini juga memperkuat pentingnya pengujian beberapa nilai K dalam proses klastering, agar pemilihan jumlah cluster tidak hanya berdasarkan asumsi, tetapi didukung oleh metrik kuantitatif.

Penelitian oleh (Ananda et al. 2023) menegaskan bahwa penggunaan DBI dalam RapidMiner merupakan pendekatan yang efektif untuk mengevaluasi kualitas klastering, terutama dalam konteks data pelanggan ritel. Evaluasi berbasis DBI memungkinkan peneliti untuk memilih model klastering yang paling representatif terhadap pola data yang dianalisis.

KESIMPULAN

Penelitian ini menunjukkan bahwa penerapan algoritma K-Means Klastering pada dataset Wholesale Customers dengan bantuan RapidMiner mampu menghasilkan segmentasi pelanggan yang terstruktur dan bermakna. Hasil Cluster Model memperlihatkan distribusi data yang tidak merata, di mana sebagian besar pelanggan terkonsentrasi pada cluster tertentu sementara sebagian kecil membentuk cluster dengan jumlah anggota yang sangat sedikit. Hal ini mengindikasikan adanya kelompok pelanggan dengan pola pembelian yang unik atau berpotensi sebagai outlier.

Melalui Example Set, setiap record pelanggan dapat ditelusuri sesuai cluster yang terbentuk, sehingga analisis karakteristik individu maupun kelompok dapat dilakukan secara lebih detail. Evaluasi menggunakan Performance Vector dengan metrik Davies-Bouldin Index (DBI) memperkuat temuan bahwa variasi jumlah cluster berpengaruh terhadap kualitas pemisahan data. Nilai DBI terkecil (0.091 pada Performance 2) menunjukkan model klastering yang paling optimal, sehingga tujuan penelitian untuk menemukan jumlah cluster yang tepat telah tercapai.



Dengan demikian, penelitian ini memberikan makna bahwa analisis berbasis klastering tidak hanya mampu mengelompokkan data secara matematis, tetapi juga dapat digunakan sebagai dasar pengambilan keputusan strategis, misalnya dalam segmentasi pelanggan ritel atau perencanaan distribusi produk.

DAFTAR PUSTAKA

- Ananda, Pingky Septiana, Eko Sediono, Irwan Sembiring, Magister Sistem Informasi, Fakultas Teknik Informasi, Universitas Kristen, and Satya Wacana. 2023. "KMeans Klastering Menggunakan RapidMiner Dalam Segmentasi Pelanggan Dengan Evaluasi Davies Bouldin Index Untuk Menentukan Jumlah Cluster Paling Optimal." 6(2):8–13.
- Annisyah, Sari, Fakultas Sains, D. A. N. Teknologi, Universitas Islam, Negeri Sultan, and Syarif Kasim. 2025. "KLAUSTERING DATA PENJUALAN DENGAN PENERAPAN METODE K MEANS UNTUK."
- Faidah, Mutmainnah Ilmiatul, and Zaehol Fatah. 2025. "Klastering K-Means Dengan Rapidminer Untuk Identifikasi Produk Terlaris." 4(1):25–33.
- Florentin, Sara Famayla, and Agus Santoso. 2022. "ANALISIS SEGMENTASI PELANGGAN MENGGUNAKAN K- MEANS." 26(2):446–57. doi:10.46984/sebatik.v26i2.2134.
- HASIPAH, SARIPA. 2025. "IMPLEMENTASI DATA MINING TRANSAKSI PENJUALAN MENGGUNAKAN ALGORITMA KLAUSTERING DENGAN METODE K-MEAN."
- Informatika, Jurnal, Perangkat Lunak, Julius Widiyanto, Pratama Putra, Erik Arfan Suganda, and Intan Purnamasari. 2022. "Penerapan RapidMiner Dengan Metode K-Means Dalam Penentuan Kluster Gangguan Jaringan WIFI Provider PT . XYZ Di Daerah Karawang." 4(1):31–35.
- Kiagus Ardelansyah Pratama, Zufar Rifa, Nur Aini Setiyawati, Ayuni Asistiyasari. 2024. "IMPLEMENTASI KLAUSTERING DATA PENJUALAN MENGGUNAKAN ALGORITMA K-MEANS." 4(1):1–9.
- Nugraha, Agung, Yutika Amelia Effendi, Zejin Tao, Mokh Afifuddin, and Nania Nuzulita. 2025. "K-Means Klastering Interpretation Using Recency , Frequency , and Monetary Factor for Retail Customers Segmentation." 23(2):435–46. doi:10.12928/TELKOMNIKA.v23i2.26044.
- Pane, Putri Pratiwi, Yusuf Ramadhan Nasution, and Mhd Furqan. 2024. "Implementasi Data Mining Dengan K-Means Klastering Untuk Memprediksi Pengadaan Obat." 5(2):286–96. doi:10.47065/josyc.v5i2.4920.
- Susilowati, Desy, and Yanuar Wicaksono. 2024. "Klasterisasi Data Pencari Kerja Di Dinas Tenaga Kerja Dan Transmigrasi Kabupaten Bantul Menggunakan Algoritma K-Means." 11(September):54–58. doi:10.33369/pseudocode.11.2.54-58.