JURNAL ILMIAH SISTEM INFORMASI (JUSI)
VOL 5, No. 1, Januari 2026, pp. 292 - 305
p-ISSN : 2803-1507 (media cetak) | e-ISSN : 2803-1531 (media online)
Link Publikasi: https://ejurnal.provisi.ac.id/index.php/JUISI/issue/view/169

Application of the K-Means Method for Grouping Product Data Based on Sales Level

Helson Matuan*[1], Esau Dude[2], Atius Mallo[3], Herlina Yowey[4], Yusuf Selius Patey[5], Heru Sutejo[6]
[1,2,3,4,5,6] Program Studi Sistem Informasi, Universitas Sepuluh Nopember Papua, 99224, Indonesia
Jl. Ardipura Raya No.22B, 99222, Kota Jayapura, Papua, Indonesia
e-mail: helsonmatuan94@gmail.com[1], esaudude19@gmail.com[2], atiusmallo@gmail.com[3], lenorayoweiuherlina@gmail.com[4], Zanpite45880@gmail.com[5], heru.sutejo01@gmail.com[6]

ARTICLE INFO

Article History:

Received: 10-11-2025
Revised: 09-12-2025
Accepted: 29-12-2025
Published: 21-01-2026

ABSTRACT

Modern retail in Indonesia is growing rapidly, requiring complex sales data to support managerial decisions. This study aims to group products at Indomaret Kotaraja by monthly sales patterns to support inventory, shelf layout, and promotional decisions. The method used is K-Means clustering on one year of packaged food and beverage sales data, with basic preprocessing and normalization to enable comparison of sales patterns across products. Several K values were evaluated using the Davies Bouldin Index, Silhouette coefficient, and SSE trend, and K = 4 was obtained as the best configuration. A total of 99 products were divided into four clusters with profiles: (1) high and stable sales (fast-moving), (2) strong spikes at the end of the year (seasonal/promotion-sensitive), (3) low but stable sales (slow-moving), and (4) fluctuating sales that require strict control. The main contribution of this research is the development of a repeatable clustering workflow and the presentation of actionable cluster profiles for replenishment prioritization, targeted promotion design, and slow-moving or volatile product stock control, thus supporting data-driven decision-making in modern retail.

Keywords: *K-Means Clustering, Retail Sales Data, Product Segmentation, Inventory Management, Fast-Moving Consumer Goods (FMCG).*

Abstrak

Ritel modern di Indonesia berkembang pesat sehingga data penjualan yang kompleks perlu diolah untuk mendukung keputusan manajerial. Penelitian ini bertujuan mengelompokkan produk di Indomaret Kotaraja berdasarkan pola penjualan bulanan untuk mendukung keputusan persediaan, penataan rak, dan promosi. Metode yang digunakan adalah klastering K-Means pada data penjualan makanan dan minuman kemasan selama satu tahun, dengan preprocessing dasar dan normalisasi agar pola penjualan antarproduk dapat dibandingkan. Beberapa nilai K dievaluasi menggunakan *Davies Bouldin Index*, koefisien *Silhouette*, dan tren SSE, dan diperoleh K = 4 sebagai konfigurasi terbaik. Sebanyak 99 produk terbagi ke dalam empat klaster dengan profil: (1) penjualan tinggi dan stabil (*fast moving*), (2) lonjakan kuat di akhir tahun (musiman/sensitif promosi), (3) penjualan rendah namun stabil (*slow moving*), dan (4) penjualan fluktuatif yang memerlukan pengendalian ketat. Kontribusi utama penelitian ini adalah penyusunan alur kerja klastering yang dapat diulang serta penyajian profil klaster yang mudah ditindaklanjuti untuk prioritas

DOI: https://doi.org/10.51903/53pfrd78

p-ISSN : 2803-1507 e-ISSN : 2803-1531

pengisian ulang, perancangan promosi terarah, dan pengendalian stok produk lambat atau volatil, sehingga mendukung pengambilan keputusan berbasis data di ritel modern.

Kata Kunci: *Klastering K-Means, Data Penjualan Ritel, Segmentasi Produk, Manajemen Persediaan, Barang Konsumsi Bergerak Cepat (FMCG).*

1. PENDAHULUAN

Pertumbuhan sektor ritel modern di Indonesia menunjukkan laju peningkatan yang sangat signifikan [1]. Indomaret merupakan salah satu jaringan ritel yang sudah sangat familiar bagi masyarakat, karena gerainya tersebar di berbagai kawasan permukiman dan pusat kegiatan masyarakat. Keberagaman jenis produk yang dipasarkan, ragam merek yang ditawarkan, serta variasi tingkat permintaan dari konsumen menjadikan pengelolaan data penjualan sebagai aspek yang sangat krusial dalam menunjang proses pengambilan keputusan manajerial [2]. Pada tingkat gerai, termasuk Indomaret Kotaraja, volume dan keragaman data penjualan yang terus meningkat menuntut adanya pendekatan analitis yang lebih sistematis, tidak hanya sekadar rekapitulasi angka penjualan agregat.

Sebagai salah satu outlet ritel modern, Indomaret Kotaraja dihadapkan pada kesulitan dalam mengelola data penjualan produk yang terus meningkat baik dari sisi jumlah maupun variasinya, sedangkan proses evaluasi penjualan yang dilakukan selama ini masih bersifat umum dan belum didukung oleh pengelompokan yang lebih rinci sesuai karakteristik penjualan masing-masing produk. Situasi ini membuat manajemen mengalami hambatan dalam mengenali produk-produk yang tergolong memiliki penjualan tinggi (fast moving), penjualan sedang, maupun penjualan rendah (*slow moving*) [3]. Dampaknya, keputusan yang berkaitan dengan penyediaan stok, penataan produk di rak (*display*), serta perumusan program promosi tidak dapat dilakukan secara maksimal [4]. Padahal, data penjualan yang sesungguhnya mengandung banyak informasi berharga selama ini hanya digunakan sebatas sebagai bahan rekap, bukan dimanfaatkan sebagai landasan untuk analisis yang lebih komprehensif. Oleh sebab itu, dibutuhkan suatu pendekatan analitis yang dapat mengelompokkan produk berdasarkan pola penjualannya, sehingga informasi yang dihasilkan dapat dimanfaatkan untuk menunjang pengambilan keputusan yang lebih akurat dan terfokus [5].

Penerapan teknik data mining, terutama metode clustering, memberikan suatu pendekatan yang terstruktur untuk membagi data produk ke dalam kelompok-kelompok berdasarkan pola tertentu. Di antara berbagai algoritma *clustering, K-Means* merupakan salah satu yang paling sering diaplikasikan [6]. Algoritma *K-Means* bekerja dengan mengelompokkan objek ke dalam sejumlah cluster berdasarkan kemiripan karakteristiknya, sehingga data produk dapat dipisahkan ke dalam kelompok-kelompok sesuai dengan tingkat penjualan masing-masing [7]. Melalui hasil pengelompokan tersebut, pihak manajemen lebih terbantu dalam menganalisis kategori produk yang menyumbangkan penjualan terbesar maupun terkecil, dan selanjutnya dapat menyusun strategi yang selaras dengan kondisi tersebut.

Kebaruan qpenelitian ini terletak pada penerapan metode K-Means untuk mengelompokkan produk di Indomaret Kotaraja pada tingkat satuan produk (SKU) berbasis pola penjualan bulanan selama satu periode tertentu, serta pada penyusunan alur kerja end-to-end yang mencakup preprocessing data, penentuan jumlah cluster terbaik, hingga interpretasi centroid yang secara eksplisit diterjemahkan ke dalam rekomendasi operasional terkait pengelolaan persediaan, penataan rak, dan strategi promosi di level gerai ritel modern. Berbeda dengan kajian yang umumnya hanya menyoroti total penjualan per kategori secara agregat, penelitian ini menekankan pengelompokan yang lebih rinci sesuai perilaku penjualan tiap produk sehingga menghasilkan segmentasi yang lebih aplikatif bagi pengambilan keputusan manajerial di satu outlet spesifik.

Penelitian ini memberikan batasan bahwa metode clustering yang digunakan hanya algoritma *K-Means* tanpa dilakukan perbandingan dengan algoritma clustering yang lain, dan ruang lingkup kajian difokuskan pada pengelompokan produk, bukan pada kegiatan peramalan (*forecasting*) penjualan di waktu mendatang [8]. Namun demikian, pemilihan jumlah cluster yang tepat dilakukan secara sistematis melalui evaluasi kualitas cluster sehingga hasil pengelompokan yang diperoleh tidak hanya bersifat deskriptif, tetapi juga dapat dipertanggungjawabkan secara metodologis. Dengan berpijak pada latar belakang tersebut, rumusan masalah penelitian ini mencakup bagaimana penerapan metode *K-Means* untuk mengelompokkan data produk di Indomaret Kotaraja berdasarkan tingkat penjualannya, berapa banyak cluster yang paling tepat untuk merepresentasikan tingkat penjualan produk, serta bagaimana pemanfaatan hasil pengelompokan tersebut oleh manajemen dalam proses pengambilan keputusan.

p-ISSN : 2803-1507 e-ISSN : 2803-1531

Selaras dengan rumusan masalah di atas, penelitian ini bertujuan untuk: (1) menerapkan metode K-Means dalam pengelompokan data produk Indomaret Kotaraja berdasarkan tingkat penjualan, (2) menetapkan cluster produk yang merepresentasikan kategori penjualan tinggi, sedang, dan rendah, serta (3) menyajikan informasi hasil pengelompokan yang dapat digunakan sebagai bahan pertimbangan manajemen dalam pengelolaan persediaan, penataan produk, dan penyusunan strategi promosi [9]. Diharapkan, penelitian ini dapat memberikan kontribusi bagi manajemen Indomaret Kotaraja dalam bentuk informasi yang tersusun secara sistematis mengenai pengelompokan produk, sehingga dapat membantu penentuan prioritas dalam pengadaan stok, penyusunan strategi promosi, serta pengaturan penempatan produk di rak. Di samping itu, penelitian ini juga memberikan manfaat bagi peneliti sebagai media untuk memperluas pemahaman mengenai penerapan metode data mining, khususnya algoritma *K-Means* pada studi kasus nyata di sektor ritel, serta dapat dijadikan rujukan bagi peneliti maupun akademisi lain yang ingin meneliti lebih lanjut mengenai penerapan metode clustering dalam pengelolaan data penjualan atau manajemen ritel [10]. Dengan demikian, kontribusi ilmiah utama penelitian ini tidak hanya pada penggunaan algoritma K-Means, tetapi pada cara hasil clustering diintegrasikan secara eksplisit ke dalam konteks pengambilan keputusan operasional di ritel modern Indonesia.

2. TINJAUAN PUSTAKA

2.1. *Data Penjualan dan Manajemen Ritel*

Berbagai penelitian menunjukkan bahwa informasi transaksi penjualan dalam periode tertentu dapat dimanfaatkan untuk memahami kinerja produk dan perilaku konsumen [11]. Data ini umumnya mencakup berbagai atribut seperti kode produk, nama produk, jumlah produk yang terjual, harga satuan, nilai penjualan, serta waktu terjadinya transaksi. Dalam konteks manajemen ritel, data penjualan memiliki peran yang sangat penting sebagai dasar untuk menganalisis kinerja produk, memahami pola pembelian konsumen, serta menilai efektivitas strategi pemasaran yang telah diterapkan [12]. Melalui pengolahan data penjualan yang baik, pihak manajemen dapat menentukan produk mana yang perlu ditambah atau dikurangi stoknya, mengidentifikasi produk yang layak diberikan program promosi, serta mengatur tata letak produk (product placement) agar lebih menarik dan sesuai dengan minat konsumen. Tanpa adanya analisis yang tepat, data penjualan hanya akan menjadi data mentah yang sulit dimanfaatkan dalam pengambilan keputusan strategis [13]. Oleh karena itu, diperlukan teknik analisis yang memadai, seperti data mining, untuk menggali pola dan informasi yang bermanfaat dari data penjualan tersebut.

2.2. *Data Mining*

Sejumlah studi membahas data mining sebagai pendekatan analitis yang digunakan untuk menggali pola dan hubungan tersembunyi dari data dalam jumlah besar guna mendukung pengambilan keputusan [14]. Proses ini melibatkan berbagai teknik statistika, pembelajaran mesin *(machine learning*), serta kecerdasan buatan untuk mengekstraksi pengetahuan (*knowledge*) yang tersembunyi di dalam data. Secara umum, tahapan dalam data mining meliputi proses pengumpulan dan pemahaman data, dilanjutkan dengan prapengolahan data (*preprocessing*) seperti pembersihan data (*data cleaning*) dan transformasi data agar data siap diolah [15]. Setelah itu, dilakukan penerapan metode atau algoritma data mining, misalnya teknik klasifikasi, clustering, asosiasi, dan lain sebagainya, sesuai dengan tujuan analisis. Tahap berikutnya adalah evaluasi hasil dan interpretasi, sehingga pola atau informasi yang dihasilkan dapat dimanfaatkan untuk mendukung pengambilan keputusan. Dalam konteks penelitian ini, teknik data mining yang digunakan adalah clustering, yaitu pengelompokan data berdasarkan kemiripan karakteristik antar data [16].

2.3. *Clustering*

Dalam berbagai penelitian, clustering digunakan sebagai metode untuk mengelompokkan data berdasarkan kemiripan karakteristik sehingga pola tertentu dapat diidentifikasi dengan lebih mudah [17]. Teknik clustering digunakan secara luas di berbagai bidang, antara lain dalam pemasaran untuk segmentasi pelanggan, pengelompokan dokumen, analisis citra, hingga pengelompokan produk berdasarkan pola penjualannya [18]. Pada penelitian ini, clustering dimanfaatkan untuk mengelompokkan produk di Indomaret Kotaraja berdasarkan tingkat penjualan. Secara umum, clustering memiliki beberapa karakteristik utama, yaitu tidak menggunakan label kelas karena termasuk dalam kategori unsupervised learning, mengandalkan ukuran kedekatan atau kemiripan seperti jarak Euclidean untuk mengukur kesamaan antar objek, serta jumlah cluster dapat ditentukan terlebih dahulu atau ditentukan secara otomatis sesuai dengan algoritma yang digunakan [19].

p-ISSN : 2803-1507 e-ISSN : 2803-1531

Pada metode *K-Means*, proses pengelompokan dilakukan dengan mengoptimalkan kedekatan data terhadap pusat cluster yang terbentuk.. Fungsi objektif yang digunakan dalam algoritma K-Means untuk mengukur kualitas pengelompokan ditunjukkan pada Persamaan (1).

$$J = \sum_{j=1}^{k} \sum_{x_i \in C_i} ||x_i - \mu_j||^2 \quad (1)$$

Keterangan:

J = nilai *error* (semakin kecil semakin baik)
k = jumlah cluster
x_i = data ke-i
μ_j = centroid cluster

Jika data memiliki atribut lebih dari satu (misal penjualan Januari–Desember), maka *centroid* dihitung per atribut menggunakan rumus yang ditunjukkan pada persamaan (2).

$$\mu_{jk} = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_{ik} \quad (2)$$

2.4. *Algoritma K-Means*

Algoritma *K-Means* banyak digunakan dalam penelitian sebelumnya karena kemampuannya dalam mengelompokkan data secara efisien dengan tingkat kompleksitas komputasi yang relatif rendah [20]. Algoritma ini bekerja dengan cara meminimalkan jumlah kuadrat jarak antara data dengan pusat cluster (*centroid*), sehingga objek-objek dalam satu cluster memiliki kedekatan karakteristik yang tinggi. Secara umum, langkah-langkah algoritma *K-Means* dimulai dengan menentukan jumlah cluster yang diinginkan, yaitu K, kemudian menginisialisasi pusat cluster (*centroid*) secara acak atau dengan metode tertentu [21]. Setelah itu, dihitung jarak setiap data ke masing-masing centroid, kemudian setiap data dikelompokkan ke dalam cluster yang memiliki jarak centroid paling dekat. Selanjutnya, posisi *centroid* dihitung ulang berdasarkan rata-rata data yang berada pada masing-masing cluster. Proses penghitungan jarak dan pembaruan centroid ini diulang hingga posisi centroid tidak banyak berubah (konvergen) atau jumlah iterasi maksimum tercapai [22].

Algoritma K-Means memiliki beberapa kelebihan, antara lain relatif mudah diimplementasikan dan memiliki waktu komputasi yang cepat untuk data berukuran besar. Namun, algoritma ini juga memiliki beberapa kelemahan, seperti sensitif terhadap pemilihan nilai awal centroid, mengharuskan penentuan jumlah cluster K terlebih dahulu sebelum proses clustering dilakukan, serta kurang sesuai untuk data dengan bentuk cluster yang tidak bulat atau data yang memiliki skala sangat berbeda jika tidak dilakukan normalisasi terlebih dahulu [23]. Dalam penelitian ini, algoritma *K-Means* digunakan untuk mengelompokkan produk berdasarkan atribut yang mencerminkan tingkat penjualan, seperti jumlah penjualan atau nilai penjualan dalam periode tertentu.

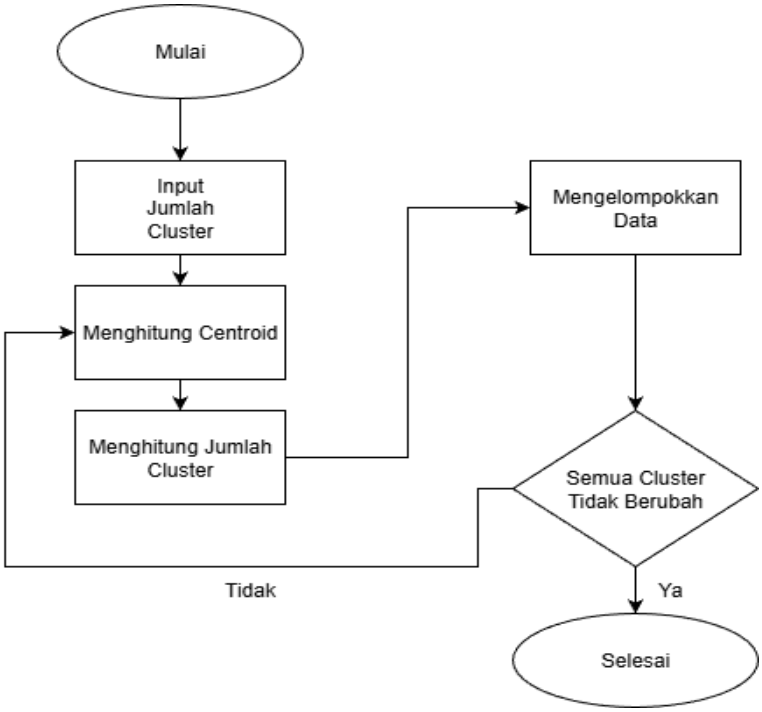
2.5. *Rapidminer*

RapidMiner telah banyak dimanfaatkan dalam penelitian analisis data sebagai perangkat bantu untuk mengimplementasikan berbagai metode data mining dan visualisasi hasil analisis [24]. *Rapidminer* menyediakan alat untuk mengumpulkan, mengekstrak, mengubah, dan menganalisis data, serta membangun model prediktif dan menghasilkan laporan analitik. Dalam pekerjaannya, *RapidMiner* memberi pengguna wawasan melalui berbagai metode prediktif dan deskriptif sehingga mereka dapat membuat keputusan terbaik [25].

3. METODOLOGI PENELITIAN

Penerapan algoritma *K-Means* dalam penelitian ini dilakukan menggunakan perangkat lunak *RapidMiner*. Pada *RapidMiner*, algoritma *K-Means* bekerja secara iteratif untuk membentuk beberapa kelompok data berdasarkan tingkat kemiripan atribut, khususnya atribut penjualan produk [26]. Proses pengelompokan dilakukan dengan cara menghitung kedekatan setiap data terhadap centroid, kemudian memperbarui posisi centroid hingga susunan cluster menjadi stabil dan tidak lagi mengalami perubahan. Sebagai contoh alur nya dapat dilihat di Gambar 1.

p-ISSN : 2803-1507 e-ISSN : 2803-1531



Gambar 1. Flowchart K-Means

Proses klastering dilakukan menggunakan algoritma *K-Means* pada *RapidMiner* dengan tahapan utama sebagai berikut. Pertama, jumlah cluster (K) ditetapkan melalui parameter *Number of Clusters* sesuai tujuan pemisahan produk, misalnya untuk membedakan kelompok fast-moving, medium-moving, dan slow-moving. Setelah *K* ditentukan, *RapidMiner* melakukan inisialisasi centroid awal secara otomatis berdasarkan pengaturan standar algoritma *K-Means*.

Setiap record data kemudian dihitung jaraknya terhadap seluruh centroid menggunakan *Euclidean Distance* sesuai kaidah *K-Means* [27], dan data dialokasikan ke cluster dengan jarak minimum. Rumus jarak antara data $x_i$dan centroid $\mu_j$dinyatakan pada persamaan (3).

$$d(x_i, \mu_j) = \sqrt{\sum_{k=1}^{n} (x_{ik} - \mu_{jk})^2} \quad (3)$$

dengan $n$merepresentasikan jumlah atribut, yakni fitur penjualan bulanan.

Berdasarkan keanggotaan *cluster* sementara tersebut, *centroid* setiap *cluster* diperbarui dengan menghitung rata-rata vektor seluruh data dalam cluster [28] menggunakan rumus yang ditunjukkan pada persamaan (4).

$$\mu_j = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i \quad (4)$$

Proses perhitungan jarak, pengelompokan, dan pembaruan centroid diulang hingga tidak ada lagi perubahan keanggotaan data (konvergen), sehingga diperoleh komposisi cluster dan centroid akhir yang stabil. *Output* RapidMiner berupa daftar cluster, nilai centroid, jumlah anggota *cluster*, dan visualisasi klastering kemudian dianalisis lebih lanjut untuk menginterpretasikan karakteristik tiap kelompok produk, misalnya kategori penjualan tinggi, sedang, dan rendah, serta implikasinya terhadap keputusan manajerial di gerai ritel.

4. HASIL DAN PEMBAHASAN

4.1. *Data Penjualan Produk*

Pada bab ini disajikan hasil pengolahan dan analisis data penjualan produk di Indomaret Kotaraja yang menjadi dasar penerapan metode *K-Means*. Data yang digunakan merupakan data historis penjualan berbagai produk makanan kemasan selama satu tahun, yang dicatat dalam bentuk penjualan tiap bulan serta total penjualan dalam satu periode. Setiap baris data merepresentasikan satu jenis produk dengan atribut utama

p-ISSN : 2803-1507 e-ISSN : 2803-1531

berupa kode produk (Kode_Produk), nama produk (Nama_Produk), kategori (Kategori), serta nilai penjualan pada bulan-bulan tertentu. Produk yang dianalisis seluruhnya berasal dari kategori makanan, antara lain berbagai varian Indomie Goreng, Indomie Kuah, Mie Sedaap, Pop Mie, Roma Marie Susu, Good Time, hingga Oreo dengan beragam ukuran kemasan. Data penjualan bulanan, misalnya untuk bulan Januari, September, Oktober, November, dan Desember, memberikan gambaran mengenai bagaimana pergerakan penjualan masing-masing produk dalam kurun waktu tertentu, sedangkan kolom total penjualan menunjukkan akumulasi unit yang terjual selama satu tahun dan membantu melihat kontribusi setiap produk terhadap keseluruhan penjualan dan data di sajikan dalam bentuk table bisa di liat di Tabel 1 .

Tabel 1. Data penjualan produk

Kode_Produk	Nama_Produk	Kategori	Jan	Sep	.....	.....	Okt	Nov	Des	Total	Rata-rata
P001	Indomie Goreng Original 75g	Makanan	281	312			735	318	485	5147	429
P002	Indomie Goreng Original 80g	Makanan	226	166			590	430	349	6309	526
P003	Indomie Goreng Rendang 75g	Makanan	677	431			730	657	249	4619	385
P004	Indomie Goreng Rendang 80g	Makanan	615	258			727	769	188	5649	471
P005	Indomie Goreng Ayam Geprek 75g	Makanan	472	445			496	296	191	5248	437
P006	Indomie Goreng Ayam Geprek 80g	Makanan	531	314			742	134	351	4633	386
P007	Indomie Kuah Kari Ayam 70g	Makanan	554	657			695	235	618	5621	468
P008	Indomie Kuah Kari Ayam 75g	Makanan	373	613			791	588	103	5306	442
P009	Indomie Kuah Soto Lamongan 70g	Makanan	465	371			263	289	527	4409	367
P010	Indomie Kuah Soto Lamongan 75g	Makanan	714	671			339	424	495	5877	490
P011	Mie Sedaap Goreng 90g	Makanan	642	216			758	640	409	5870	489
P012	Mie Sedaap Kari Spesial 90g	Makanan	668	419			368	452	318	6518	543
P013	Mie Sedaap Ayam Bawang 90g	Makanan	422	432			645	609	619	5677	473
P014	Sarimi Isi 2 Ayam Bawang 80g	Makanan	693	496			759	510	216	5192	433
P015	Pop Mie Rasa Baso 75g	Makanan	204	110			487	535	663	4427	369
……	……..	……..	……	…..	…	…	……	…..	…..	…..	…..

4.2. *Pengolahan Awal Data (Preprocessing)*

Bagian ini menguraikan tahap preprocessing sebagai fondasi sebelum data penjualan Indomaret Kotaraja dianalisis dengan algoritma K Means di RapidMiner, dengan tujuan memastikan data bersih, bertipe konsisten, dan berada pada skala yang sebanding agar perhitungan jarak Euclidean benar benar merefleksikan kemiripan pola penjualan antar produk. Pratinjau dataset pada Tabel 1 menampilkan tabel berisi Nama Produk, Kategori yang pada contoh seluruhnya Makanan dan minuman, serta kolom penjualan bulanan. Setiap baris merepresentasikan satu produk, misalnya berbagai varian Indomie, Mie Sedaap, Sarimi, dan Pop Mie, sedangkan setiap kolom penjualan memuat unit terjual per bulan. Variasi nilai antar bulan menunjukkan adanya fluktuasi permintaan sehingga diperlukan penyeragaman skala sebelum proses clustering. Secara prosedural, data diimpor melalui Read CSV atau Retrieve dengan pengecekan delimiter dan encoding agar kolom penjualan terbaca numerik.

Header dirapikan dan diseragamkan penamaannya, misalnya Penjualan Jan hingga Penjualan Des, kemudian peran atribut ditetapkan. Identitas seperti Kode Produk diset sebagai id, sementara Nama Produk dan Kategori dipertahankan untuk interpretasi hasil tetapi tidak digunakan sebagai fitur. Kualitas data diperiksa untuk mendeteksi nilai kosong, karakter non angka, duplikasi, dan nilai yang tidak wajar. Nilai hilang ditangani secara konsisten, contohnya nol untuk bulan tanpa penjualan, dan baris duplikat dibersihkan. Seluruh kolom penjualan bulanan kemudian dinormalisasi menggunakan operator Normalize dengan metode Z Transformation atau Min Max agar skala sebanding. Terakhir, dipilih atribut numerik yang relevan, umumnya vektor penjualan Januari sampai Desember, sebagai fitur K Means, sedangkan atribut non numerik

p-ISSN : 2803-1507 e-ISSN : 2803-1531

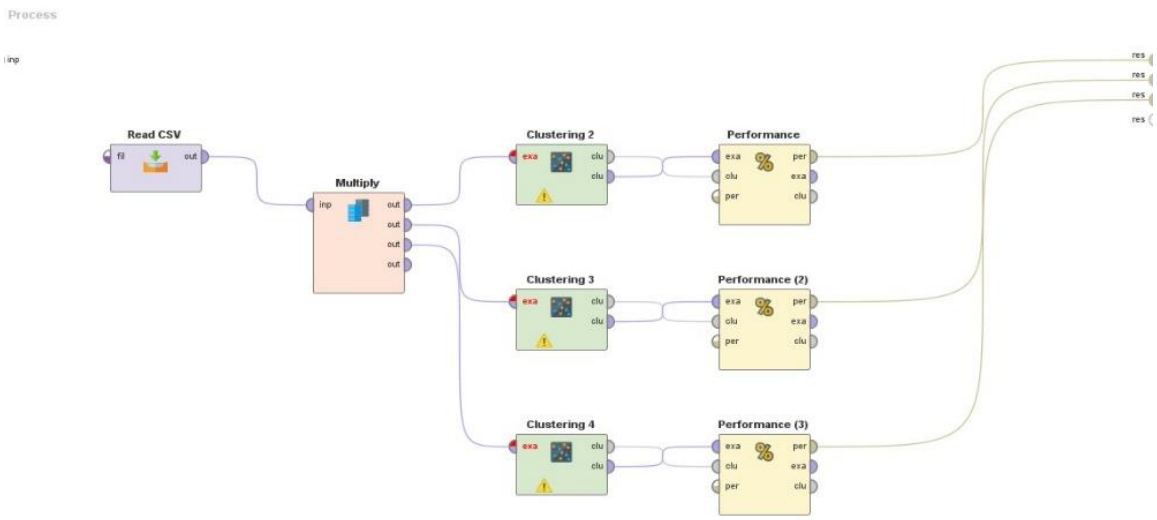
hanya dipakai pada tahap interpretasi. Hasil preprocessing menghasilkan ExampleSet yang rapi, konsisten, dan siap diproses oleh operator K Means sehingga pengelompokan produk berdasarkan kemiripan pola penjualan menjadi valid dan dapat dimanfaatkan untuk mendukung keputusan operasional toko, untuk contoh nya bisa di liat di Gambar 2.

1	Nama_P...	Kategori	Penjuala...	Penjuala...	Penjuala...	Penjuala...	Penjuala...	Penjuala...	Penjuala...	Penjuala...	Penjuala...	Penjuala...
2	Indomie ...	Makanan	281	122	771	349	517	444	422	391	312	735
3	Indomie ...	Makanan	226	689	725	789	698	660	404	583	166	590
4	Indomie ...	Makanan	677	477	324	127	346	194	126	281	431	730
5	Indomie ...	Makanan	615	281	611	263	483	638	354	462	258	727
6	Indomie ...	Makanan	472	686	676	439	240	633	359	315	445	496
7	Indomie ...	Makanan	531	463	108	291	721	525	291	162	314	742
8	Indomie ...	Makanan	554	276	162	696	638	463	483	144	657	695
9	Indomie ...	Makanan	373	110	590	429	346	771	463	129	613	791
10	Indomie ...	Makanan	465	388	310	264	120	712	289	411	371	263
11	Indomie ...	Makanan	714	574	189	509	677	277	447	561	671	339
12	Mie Sed...	Makanan	642	348	161	494	392	481	575	754	216	758
13	Mie Sed...	Makanan	668	763	402	574	674	440	652	788	419	368
14	Mie Sed...	Makanan	498	125	601	578	459	615	312	260	432	645
15	Sarimi Is...	Makanan	693	610	439	294	122	149	328	576	496	759
16	Pop Mie ...	Makanan	204	288	436	359	198	269	174	704	110	487

Gambar 2. Data Bersih

4.3. *Proses dan Hasil Clustering K-Means*

Pada tahap awal analisis *clustering*, penentuan jumlah *cluster* (K) menjadi aspek penting karena berpengaruh langsung terhadap struktur pengelompokan data yang terbentuk. Nilai K yang berbeda dapat menghasilkan tingkat homogenitas dan pemisahan antar *cluster* yang bervariasi sehingga perlu diuji secara sistematis. Oleh sebab itu, dilakukan beberapa percobaan dengan variasi jumlah K untuk melihat karakteristik hasil pengelompokan berdasarkan indikator evaluasi tertentu. Visualisasi proses pemilihan jumlah *cluster* disajikan pada Gambar 3.



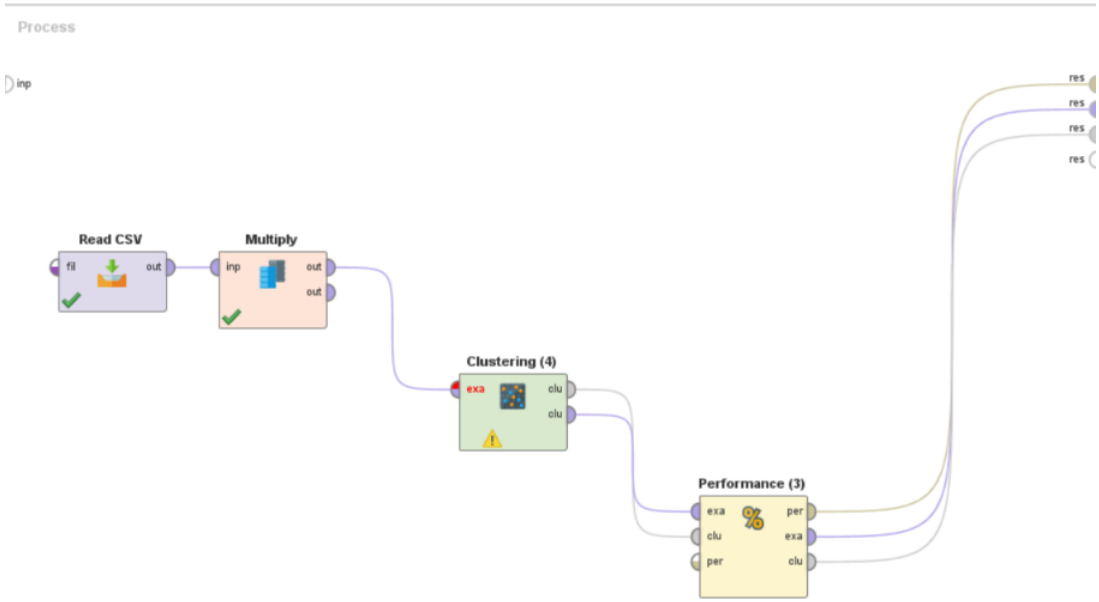
Gambar 3. Pemilihan Jumlah Kluster

Workflow RapidMiner pada penelitian ini terdiri dari dua tahap utama, yaitu eksperimen penentuan jumlah cluster terbaik dan pemodelan akhir dengan K yang terpilih. Pada tahap eksperimen, data penjualan dimuat melalui operator Read CSV menjadi *ExampleSet*, kemudian digandakan dengan Multiply ke tiga jalur yang masing-masing menjalankan algoritma K-Means dengan K = 2, K = 3, dan K = 4. Setiap jalur menghasilkan data yang sudah berlabel cluster (port *exa*) dan skema penugasan cluster (port *clu*), kemudian dievaluasi

p-ISSN : 2803-1507 e-ISSN : 2803-1531

menggunakan operator Performance dengan metrik klasterisasi seperti Davies Bouldin Index (semakin kecil semakin baik) dan Silhouette Coefficient (semakin besar semakin baik). Secara opsional diamati pula nilai SSE/WCSS dan jarak antarkluster untuk menilai kompaksi dan pemisahan kluster. Ketiga keluaran ini dibandingkan secara simultan pada panel *Results* sehingga pemilihan jumlah cluster tidak hanya bertumpu pada pertimbangan visual, tetapi pada indikator kuantitatif yang terukur.

Setelah proses evaluasi jumlah *cluster* dilakukan, tahap selanjutnya adalah menerapkan algoritma *K-Means* menggunakan nilai K yang telah ditentukan pada *workflow* akhir. Pada tahap ini, fokus analisis diarahkan pada proses pengelompokan data penjualan serta pola yang terbentuk pada masing-masing *cluster*. Implementasi *clustering* dilakukan secara terstruktur menggunakan perangkat lunak *RapidMiner* sehingga setiap tahapan dapat diamati dengan jelas. Alur proses *clustering* dengan empat *cluster* tersebut ditunjukkan pada Gambar 4.



Gambar 4. Proses Clustering K-Means dengan 4 Kluster

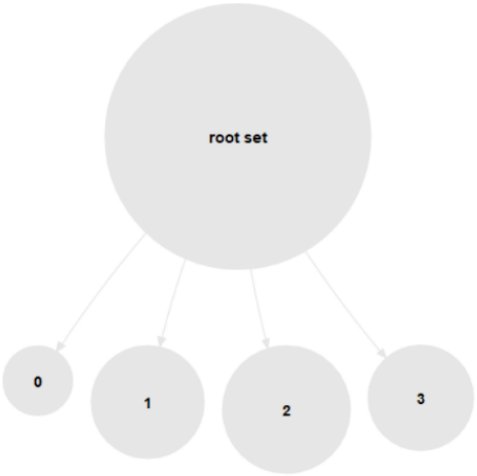
Berdasarkan perbandingan nilai metrik untuk K = 2, K = 3, dan K = 4, konfigurasi K = 4 dipilih karena menghasilkan kompromi terbaik antara kompaksi dan pemisahan kluster, dengan nilai Davies Bouldin Index dan/atau SSE yang lebih rendah serta Silhouette yang relatif lebih tinggi dan ukuran kluster yang masih wajar. Workflow akhir yang digunakan kemudian dipadatkan menjadi satu jalur: data penjualan dimuat melalui Read CSV, diteruskan oleh Multiply sebagai penghubung ke operator K-Means dengan K = 4. Output K-Means berupa label cluster untuk tiap produk dan empat centroid yang merepresentasikan pola penjualan yang berbeda. Tahap ini diikuti kembali oleh operator Performance untuk mendokumentasikan kualitas klasterisasi K = 4. Pada panel *Results*, peneliti mengamati data berlabel cluster, struktur penugasan cluster, serta ringkasan metrik performa sebelum menginterpretasikan karakteristik masing-masing kluster dalam konteks bisnis (fast moving, medium moving, slow moving, dan kluster khusus).

Secara konseptual, workflow ini tidak hanya menjelaskan langkah teknis di RapidMiner, tetapi juga merepresentasikan sebuah prototipe subsistem analitik dalam kerangka Decision Support System (DSS) untuk ritel. Pertama, penggunaan beberapa metrik validitas (DBI, Silhouette, SSE) menjadikan penentuan jumlah cluster sebagai proses pengambilan keputusan yang terstruktur dan terdokumentasi, bukan sekadar subjektif. Kedua, label cluster dan centroid yang dihasilkan dapat diintegrasikan ke dalam sistem informasi penjualan sebagai atribut tambahan pada master produk, sehingga mendukung aturan keputusan operasional seperti prioritas pengadaan stok, penataan rak, dan penargetan promosi berbasis segmen penjualan. Ketiga, alur yang bersifat *repeatable* ini memungkinkan manajemen melakukan *re-run* klasterisasi saat terjadi perubahan pola permintaan (misalnya musiman atau pasca promosi), sehingga DSS yang dibangun bersifat dinamis dan adaptif terhadap data baru. Dengan demikian, workflow RapidMiner dalam penelitian ini

p-ISSN : 2803-1507 e-ISSN : 2803-1531

berkontribusi pada pengembangan pengelolaan data ritel berbasis sistem informasi yang lebih analitis, transparan, dan berorientasi pada dukungan keputusan manajerial.

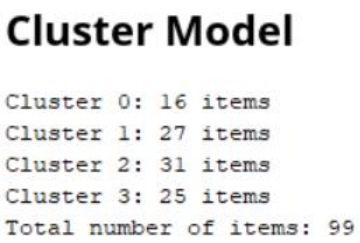
Visualisasi hasil *clustering* digunakan untuk membantu memahami struktur pembagian data yang dihasilkan oleh algoritma *K-Means*. Melalui grafik, hubungan antara himpunan data awal dan kluster hasil pengelompokan dapat diamati secara lebih intuitif. Representasi grafis ini memudahkan peneliti dalam menelusuri bagaimana setiap objek data dialokasikan ke dalam kluster tertentu. Tampilan grafik kluster hasil proses pengelompokan ditunjukkan pada Gambar 5.



Gambar 5. Graph Cluster

Gambar 5 menampilkan sebuah grafik kluster yang menunjukkan bagaimana satu himpunan asal atau root set dipisah menjadi empat kelompok bernomor 0, 1, 2, dan 3. Lingkaran besar di bagian atas mewakili keseluruhan data sebelum dikelompokkan, sedangkan empat lingkaran di bawahnya menggambarkan hasil pemisahan ketika jumlah kluster ditetapkan empat, dengan arah keterhubungan mengindikasikan perpindahan keanggotaan dari himpunan asal ke tiap kluster. Dalam kerangka K-Means, visualisasi ini menegaskan bahwa seluruh objek berasal dari satu kumpulan yang sama lalu dibagi menjadi empat kelompok yang saling terpisah berdasarkan kemiripan atribut penjualan; ukuran atau posisi simpul, bila diterapkan, dapat dipahami sebagai petunjuk banyaknya anggota atau tingkat keterkaitan. Singkatnya, gambar tersebut merepresentasikan struktur hierarkis sederhana: root set sebagai sumber semua objek dan kluster 0, 1, 2, 3 sebagai keluaran pengelompokan ke empat kelompok berbeda.

Model *clustering* tidak hanya menggambarkan proses pengelompokan, tetapi juga memberikan informasi mengenai distribusi jumlah data pada setiap *cluster* yang terbentuk. Informasi ini penting untuk melihat keseimbangan ukuran antar kelompok serta potensi dominasi pola tertentu dalam data penjualan. Dengan mengetahui jumlah anggota di setiap *cluster*, peneliti dapat memahami seberapa banyak produk yang memiliki karakteristik penjualan yang serupa. Visualisasi jumlah anggota pada masing-masing *cluster* ditampilkan pada Gambar 6.



Gambar 6. Model Kluster

p-ISSN : 2803-1507 e-ISSN : 2803-1531

Gambar 6 ini menampilkan jumlah anggota pada setiap kelompok hasil K Means. Dari 99 produk, cluster 0 berisi 16, cluster 1 berisi 27, cluster 2 berisi 31, dan cluster 3 berisi 25 item. Angka ini belum menunjukkan mana yang paling laku, melainkan berapa banyak produk dengan pola penjualan yang mirip. Untuk memahami makna setiap cluster, hitung ringkasan penjualan di dalamnya seperti rata rata bulanan, total tahunan per produk, ukuran kestabilan melalui deviasi standar atau selisih maksimum sampai minimum, serta tren dari awal hingga akhir tahun. Nilai tinggi dan stabil dapat digolongkan fast moving, nilai sedang masuk medium, nilai rendah dan konsisten sebagai slow, sedangkan pola rendah namun naik pada bulan tertentu sebagai niche atau musiman.

Analisis lanjutan terhadap hasil *clustering* dilakukan dengan mengamati nilai pusat *cluster* atau *centroid* yang dihasilkan oleh algoritma *K-Means*. Nilai *centroid* merepresentasikan karakteristik rata-rata dari seluruh anggota dalam satu *cluster* sehingga dapat digunakan untuk memahami pola penjualan secara agregat. Melalui ringkasan nilai ini, perbandingan perilaku penjualan antar *cluster* dapat dilakukan untuk setiap periode waktu. Penyajian nilai pusat *cluster* dalam bentuk tabel ditampilkan pada Gambar 7.

Attribute	cluster_0	cluster_1	cluster_2	cluster_3
Penjualan_Jan	473	335.852	491	520.280
Penjualan_Feb	513.062	250.889	515.548	434.440
Penjualan_Mar	553.938	464	317.258	390.520
Penjualan_Apr	579.625	456.926	476.290	364.320
Penjualan_Mei	567.750	449.481	555.161	210.800
Penjualan_Jun	424.312	483.444	385.355	482.680
Penjualan_Jul	432.250	397.630	373.484	458.600
Penjualan_Agu	632.500	557.296	282.226	365.400
Penjualan_Sep	406.500	334.111	521.968	363.880
Penjualan_Okt	458.375	573.222	405.968	427
Penjualan_Nov	250.500	611.407	525.710	323.480
Penjualan_Des	436.812	404.519	527.935	391.800

Gambar 7. Pusat Kluster Tabel

Gambar 7 ini menampilkan model cluster berupa ringkasan nilai pusat kluster atau centroid untuk setiap bulan penjualan; kolom cluster_0 hingga cluster_3 menunjukkan rata rata penjualan tipikal anggota tiap kluster pada Januari sampai Desember, sedangkan baris Penjualan_Jan hingga Penjualan_Des adalah atribut waktu yang diringkas. Angka pada tiap sel merepresentasikan pola bulanan khas dari kluster tersebut, sehingga bisa dibandingkan mana kluster yang cenderung tinggi, sedang, atau rendah pada bulan tertentu serta melihat stabilitas atau fluktuasinya sepanjang tahun. Dengan kata lain, tabel ini bukan daftar penjualan produk satu per satu, melainkan profil rata rata per bulan yang mewakili perilaku penjualan masing masing kluster; jika nilai pada satu kolom konsisten lebih tinggi dibanding kolom lain, kluster itu dapat ditafsirkan sebagai kelompok dengan performa penjualan lebih kuat, sedangkan kolom dengan nilai lebih rendah menggambarkan kelompok yang bergerak lebih lambat.

4.4. *Analisis Cluster*

Evaluasi kualitas hasil *clustering* merupakan tahap penting untuk menilai seberapa baik algoritma *K-Means* mengelompokkan data penjualan. Evaluasi ini dilakukan menggunakan metrik performa yang mengukur tingkat kerapatan data dalam kluster dan tingkat pemisahan antar kluster. Melalui metrik tersebut, karakteristik setiap kluster dapat dianalisis secara kuantitatif tanpa bergantung pada visualisasi semata. Ringkasan nilai evaluasi performa *clustering* disajikan pada Gambar 8.

p-ISSN : 2803-1507 e-ISSN : 2803-1531

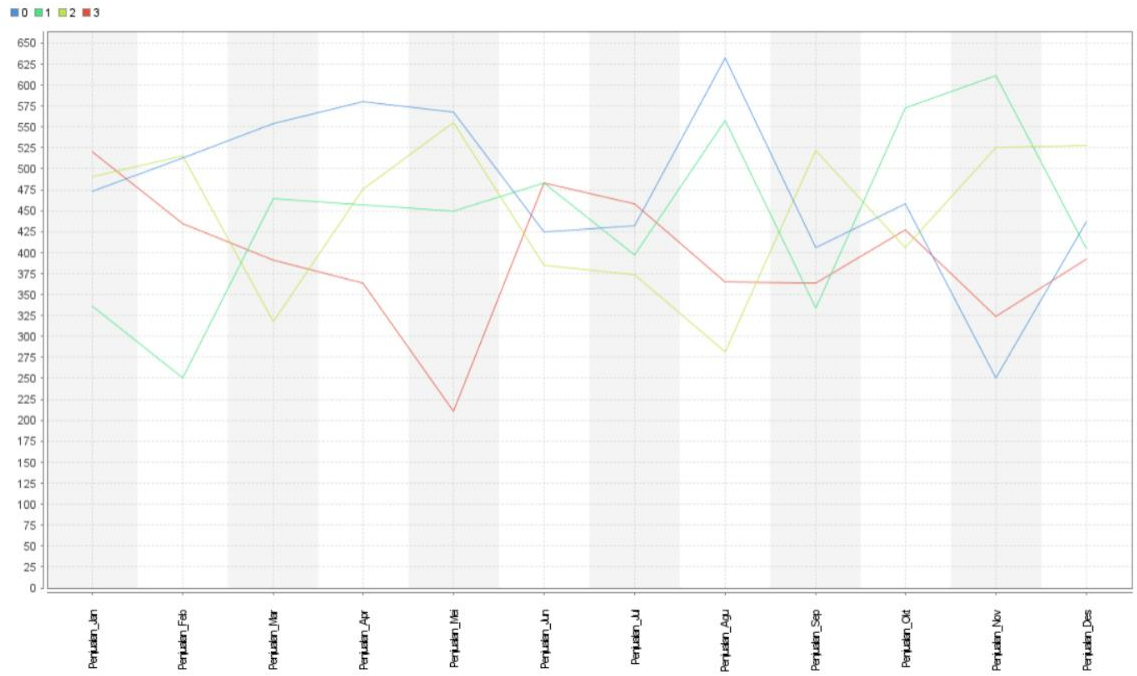
PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: -377616.858
Avg. within centroid distance_cluster_0: -360878.203
Avg. within centroid distance_cluster_1: -405963.402
Avg. within centroid distance_cluster_2: -385342.795
Avg. within centroid distance_cluster_3: -348135.168
Davies Bouldin: -2.336
```

Gambar 8. Performance Vector

Gambar 8 PerformanceVector ini merangkum kualitas hasil pengelompokan K-Means. Nilai "Avg. within centroid distance" menunjukkan rata-rata jarak data terhadap pusat klusternya; semakin kecil jarak (dalam praktik perhatikan besar absolutnya), semakin kompak kluster tersebut. Dari output yang tampak, rata-rata keseluruhan adalah −377616,858, dengan rincian per kluster: kluster 0 sebesar −360878,203, kluster 1 sebesar −405963,402, kluster 2 sebesar −385342,795, dan kluster 3 sebesar −348135,168. Karena tandanya negatif akibat cara ukur yang digunakan, bandingkan berdasarkan besarannya: kluster 3 memiliki jarak dalam kluster paling kecil sehingga paling rapat, disusul kluster 0, lalu kluster 2, sedangkan kluster 1 paling menyebar. Baris "Davies Bouldin: −2,336" adalah indeks pemisahan kluster; prinsipnya semakin kecil nilainya semakin baik karena menandakan kluster makin kompak dan saling berjauhan. Dengan demikian, keluaran ini menyiratkan struktur kluster yang relatif terpisah dengan tingkat kerapatan berbeda; kluster 3 paling kompak, kluster 1 paling renggang, dan nilai Davies-Bouldin yang rendah mengindikasikan kualitas pemisahan yang cukup baik jika dibandingkan alternatif konfigurasi K lain yang telah diuji. Jika ingin menuliskan di naskah, Anda bisa menyatakan: "Rata-rata jarak dalam kluster menunjukkan kluster 3 paling kompak, kluster 1 paling tersebar, dan indeks Davies-Bouldin yang rendah mendukung pemilihan konfigurasi ini."

Interpretasi hasil *clustering* tidak hanya didasarkan pada nilai metrik dan pusat *cluster*, tetapi juga diperkuat melalui visualisasi sebaran data dalam bentuk plot. Plot *cluster* memberikan gambaran mengenai perbedaan pola penjualan antar kelompok secara lebih intuitif, khususnya dalam melihat kecenderungan nilai pada periode tertentu. Melalui visualisasi ini, dinamika perubahan penjualan dari awal hingga akhir tahun dapat diamati untuk setiap *cluster*. Penyajian sebaran data hasil pengelompokan divisualisasikan pada Gambar 9.



Gambar 9. Plot Kluster

p-ISSN : 2803-1507 e-ISSN : 2803-1531

Kesimpulannya, cluster 2 merupakan kandidat utama produk fast moving pada periode akhir tahun karena menunjukkan penjualan yang konsisten tinggi di kuartal empat. Cluster 0 juga memiliki kinerja penjualan kuat, namun pola yang lebih volatil menuntut pengelolaan stok yang lebih adaptif, terutama untuk mengantisipasi penurunan pada bulan November. Cluster 1 dapat dikategorikan sebagai medium moving dengan momentum peningkatan yang besar menjelang akhir tahun, sehingga menjadi sasaran yang potensial untuk program promosi musiman, bundling, dan penawaran bertema akhir tahun. Sementara itu, cluster 3 termasuk slow moving yang memerlukan pengendalian persediaan yang ketat, evaluasi varian produk, serta penempatan pada rak yang tidak mengambil ruang utama agar biaya penyimpanan dan risiko penumpukan stok dapat ditekan.

Dari sisi konseptual, keempat cluster ini dapat dipandang sebagai basis segmentasi keputusan dalam kerangka Decision Support System (DSS) ritel. Label cluster berfungsi sebagai atribut tambahan pada master data produk di sistem informasi, yang kemudian dapat dikaitkan dengan seperangkat aturan keputusan, misalnya aturan pengisian ulang stok yang lebih agresif untuk fast moving, aturan penempatan rak prioritas untuk produk dengan kontribusi penjualan tinggi, serta aturan promosi terarah bagi kelompok medium dan musiman. Dengan demikian, hasil klasterisasi tidak hanya berhenti pada deskripsi pola penjualan, tetapi diintegrasikan sebagai komponen *model base* dan *knowledge* dalam DSS yang membantu manajemen merumuskan strategi stok, display, dan promosi secara lebih sistematis, terdokumentasi, dan dapat diulang ketika data penjualan diperbarui.

5. DISKUSI

Hasil pengelompokan data penjualan menggunakan metode *K-Means* menghasilkan empat kluster dengan karakteristik pola penjualan yang berbeda antarproduk. Setiap kluster memperlihatkan variasi tingkat penjualan, mulai dari produk dengan volume tinggi dan stabil hingga produk dengan volume rendah dan pola penjualan yang tidak konsisten sepanjang periode pengamatan. Perbedaan karakteristik ini mencerminkan adanya segmentasi perilaku permintaan konsumen terhadap produk ritel pada tingkat *SKU*. Pola tersebut memberikan gambaran mengenai heterogenitas kinerja produk yang tidak dapat dijelaskan hanya melalui nilai penjualan rata-rata.

Jika dikaitkan dengan penelitian terdahulu, pola kluster yang terbentuk memiliki kesesuaian dengan temuan studi yang mengelompokkan produk ritel ke dalam kategori *fast moving* dan *slow moving* menggunakan teknik *clustering* berbasis data penjualan [4], [5]. Namun, keberadaan kluster tambahan dengan pola fluktuatif memperlihatkan perbedaan dari penelitian yang hanya mengidentifikasi tiga kelompok utama [9], [10]. Perbedaan ini dapat dikaitkan dengan cakupan data yang lebih rinci serta periode observasi yang memungkinkan variasi temporal penjualan muncul secara lebih jelas. Kondisi tersebut memperlihatkan bahwa struktur kluster sangat dipengaruhi oleh karakteristik data yang digunakan dalam proses analisis.

Selain kluster dengan penjualan stabil, hasil analisis juga memperlihatkan kelompok produk dengan pola penjualan yang meningkat pada periode tertentu dan menurun pada periode lainnya. Pola semacam ini sejalan dengan temuan penelitian sebelumnya yang mengaitkan fluktuasi penjualan dengan faktor musiman, promosi, serta perubahan preferensi konsumen [11], [12]. Keberadaan variasi temporal tersebut menunjukkan bahwa sebagian produk memiliki sensitivitas yang lebih tinggi terhadap konteks waktu dibandingkan produk lain. Hal ini menjelaskan mengapa pendekatan pengelolaan produk yang seragam kurang efektif untuk seluruh kelompok produk.

Dari perspektif praktis, perbedaan karakteristik kluster memberikan dasar analitis bagi pengelolaan persediaan dan strategi pemasaran yang lebih terarah. Produk dengan volume penjualan tinggi memerlukan pengelolaan stok yang berbeda dibandingkan produk dengan pergerakan lambat atau tidak stabil [6], [13]. Pada sisi teoritis, penggunaan metode *unsupervised learning* dalam konteks ritel memperlihatkan kemampuannya dalam mengungkap pola laten yang tidak terlihat melalui analisis deskriptif sederhana. Adapun keterbatasan penelitian ini terletak pada penggunaan satu algoritma klasterisasi dan keterbatasan variabel penjualan, sehingga penelitian selanjutnya dapat memperkaya analisis dengan membandingkan metode lain serta menambahkan atribut seperti harga, diskon, dan promosi.

6. KESIMPULAN DAN SARAN

6.1. *Kesimpulan*

p-ISSN : 2803-1507 e-ISSN : 2803-1531

Penelitian berhasil menerapkan K-Means di RapidMiner untuk mengelompokkan produk Indomaret Kotaraja berdasarkan pola penjualan bulanan. Dari uji jumlah kluster, konfigurasi empat kluster dipilih karena memberi kinerja terbaik menurut metrik evaluasi (Davies–Bouldin lebih rendah, separasi antarkluster lebih baik) dan tetap masuk akal secara bisnis. Pembagian anggota menunjukkan total 99 produk tersebar ke empat kluster dengan komposisi 16, 27, 31, dan 25 produk. Profil centroid bulanan memperlihatkan perbedaan karakter: satu kluster berperforma tinggi dan relatif stabil pada kuartal empat sehingga layak diposisikan sebagai fast moving; satu kluster berkinerja sedang namun meningkat kuat menjelang akhir tahun yang cocok ditangani dengan promosi musiman; satu kluster cenderung rendah tetapi stabil yang menunjukkan slow moving; dan satu kluster lain berfluktuasi lebih tajam sehingga memerlukan pengelolaan yang lebih hati-hati. Evaluasi PerformanceVector menegaskan kualitas pemisahan, dengan variasi kerapatan antarkluster yang tetap terpisah dengan baik. Keseluruhan proses preprocessing yang mencakup pembersihan, penetapan tipe, dan normalisasi memastikan hasil klasterisasi valid serta siap digunakan untuk keputusan operasional toko.

6.2. *Saran*

Pertama, gunakan hasil klaster untuk kebijakan stok: tingkatkan ketersediaan dan kecepatan pengisian untuk fast moving, terapkan stok ketat untuk slow moving, dan atur stok adaptif untuk kluster yang volatil. Kedua, rancang strategi promosi dan penataan rak yang berbeda per kluster: promosi bundling dan display menonjol untuk kelompok menengah yang memiliki potensi kenaikan, promosi musiman untuk kluster yang memuncak di akhir tahun, serta evaluasi harga dan penempatan rak sekunder untuk kelompok berpenjualan rendah. Ketiga, lakukan pemantauan berkala terhadap metrik performa kluster dan perbarui model minimal tiap kuartal atau saat ada perubahan pola musiman, agar label fast, medium, dan slow tetap relevan. Keempat, perluas atribut pada dataset dengan variabel pendukung seperti harga, diskon, stok, dan kalender promosi agar pemisahan kluster makin informatif; pada tahap berikutnya dapat dipertimbangkan perbandingan metode (misalnya K-Medoids atau DBSCAN) dan integrasi uji elbow atau silhouette secara sistematis. Kelima, bangun laporan atau dashboard yang menampilkan ringkasan centroid, jumlah anggota, serta rekomendasi tindakan per kluster agar tim toko dapat segera mengeksekusi keputusan harian.

DAFTAR PUSTAKA

[1] M. Handoko, H. Mubarok, R. K. Shaura, H. Lesmana, R. Widyastuti, R. Swastika, W. Haryanto, dan D. Hartini, "The Architecture of Intelligent Transportation System Based on Sensor Monitoring (Implementation in Jakarta Area)," *Journal of Technology Informatics and Engineering*, vol. 4, no. 2, pp. 190–203, 2025, doi:10.51903/jtie.v4i2.357.

[2] F. Roni, M. Handoko, R. D. B. Parancika, M. Aris, Y. M. Ardi, dan Syabrinildi, "Determination of Employee Performance: Work Environment and Leadership Style (Case Study at PT. MPIW Jakarta)," *Journal of Management and Informatics*, vol. 4, no. 2, pp. 773–790, 2025, doi:10.51903/jmi.v4i2.216.

[3] A. N. Amanta, Migunani, dan M. U. Dewi, "Design and Development of STEKOM Learning Management System (SIAKAD) to Optimize User Experience," *Jurnal Ilmiah Sistem Informasi*, vol. 5, no. 1, pp. 1–16, 2026, doi:10.51903/8gbfzh39.

[4] E. Sholekha, B. Irawan, and A. Bahtiar, "Analisis Penjualan Produk Snack Dan Minuman Menggunakan Metode K-Means Pada Dataset Transaksi Penjualan," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 3, pp. 2533–2539, Jun. 2024, doi: 10.36040/jati.v8i3.9310.

[5] T. Wahyudi *et al.*, "Penerapan Metode K-Means pada Data Penjualan untuk Mendapatkan Produk Terlaris di PT. Titian Nusantara Boga: Penerapan Metode K-Means Pada Data," *ejournal.sisfokomtek.org*, vol. 5, no. 1, pp. 228–236, 2023, doi: 10.55338/saintek.v5i1.1379.

[6] F. Zafira, B. Irawan, and A. Bahtiar, "Penerapan Data Mining untuk Estimasi Stok Barang dengan Metode K-Means Clustering," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 156–161, Feb. 2024, doi: 10.36040/jati.v8i1.8319.

[7] T. Amalina, D. B. A. Pramana, and B. N. Sari, "Metode K-Means Clustering dalam Pengelompokan Penjualan Produk Frozen Food," *Jurnal Ilmiah Wahana Pendidikan*, vol. 8, no. 15, pp. 574–583, Sep. 2022, doi: 10.5281/zenodo.7052276.

[8] L. Azzahra and A. Yasir, "Metode K-Means Clustering dalam Pengelompokan Penjualan Produk

p-ISSN : 2803-1507 e-ISSN : 2803-1531

Frozen Food," *Jurnal Ilmu Komputer dan Sistem Informasi*, vol. 3, no. 1, pp. 1–10, Jan. 2024, doi: 10.70340/jirsi.v3i1.88.

[9] Agung Nugraha, Odi Nurdiawan, and Gifthera Dwilestari, "Penerapan Data Mining Metode K-Means Clustering untuk Analisa Penjualan pada Toko Yana Sport," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 6, no. 2, pp. 849–855, 2022, doi: 10.36040/jati.v6i2.5755.

[10] Nur Afiasari, Nana Suarna, and Nining Rahaningsi, "Implementasi Data Mining Transaksi Penjualan Menggunakan Algoritma Clustering dengan Metode K-Means," *Jurnal Saintekom: Sains, Teknologi, Komputer dan Manajemen*, vol. 13, no. 1, pp. 100–110, Mar. 2023, doi: 10.33020/saintekom.v13i1.402.

[11] Normah, S. Nurajizah, and A. Salbinda, "Penerapan Data Mining Metode K-Means Clustering untuk Analisa Penjualan pada Toko Fashion Hijab Banten," *Jurnal Teknik Komputer*, vol. 7, no. 2, pp. 158–163, Jul. 2021, doi: 10.31294/jtk.v7i2.10553.

[12] S. A. Rahmah, "Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means Clustering (Studi Kasus di Toko Juanda Tani Kecamatan Hutabayu Raja)," *Djtechno: Jurnal Teknologi Informasi*, vol. 1, no. 1, pp. 1–5, 2020, doi: 10.46576/djtechno.v1i1.964.

[13] S. Pujiono, R. Astuti, and F. Muhamad Basysyar, "Implementasi Data Mining untuk Menentukan Pola Penjualan Produk Menggunakan Algoritma K-Means Clustering," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 615–620, Feb. 2024, doi: 10.36040/jati.v8i1.8360.

[14] M. Bui and A. Bahtiar, "Implementasi metode algoritma K-Means Clustering untuk mengelompokkan transaksi penjualan barang di Toko Arino," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, 2024, Accessed: Nov. 23, 2025. [Online]. Available: https://mail.ejournal.itn.ac.id/index.php/jati/article/view/8975

[15] N. P. Gantara, I. Ali, K. Akuntansi, and R. P. Lunak, "Penerapan Metode K-Means Clustering Pada Penjualan Barang Di Sports Station," *E-Link: Jurnal Ilmu Ekonomi dan Informatika*, vol. 18, pp. 28–39, 2023, Accessed: Nov. 23, 2025. [Online]. Available: https://journal.umg.ac.id/index.php/e-link/article/view/5339

[16] F. Sandova, R. Kurniawan, and T. Supratati, "Penerapan Data Mining Menggunakan Metode K-Means Clustering pada Penjualan Tas di Asia Toserba Cirebon," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 245–251, 2024, doi: 10.36040/jati.v8i1.8330.

[17] M. R. Nahjan, N. Heryana, and A. Voutama, "Implementasi Rapidminer dengan Metode Clustering K-Means untuk Analisa Penjualan pada Toko Oj Cell," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 1, pp. 101–104, 2023, doi: 10.36040/jati.v7i1.6094.

[18] V. Alvianatinova, I. Ali, and N. R. J., "Penerapan Algoritma K-Means Clustering dalam Pengelompokan Data Penjualan Supermarket Berdasarkan Cabang (Branch)," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, 2024, Accessed: Nov. 24, 2025. [Online]. Available: https://mail.ejournal.itn.ac.id/index.php/jati/article/view/8993

[19] D. Abid, R. Wirya Adikusuma, A. Mufti, A. Fikri, D. Rinci, and K. Hapsari, "Penerapan Metode K-Means Clustering Untuk Analisa Penjualan Komoditas Toko Tani Indonesia," *Jurnal Ilmiah Teknik Informatika*, vol. 3, no. 2, 2022, Accessed: Nov. 24, 2025. [Online]. Available: http://ejurnal.itats.ac.id/kernel/article/view/4076

[20] F. D. Agustiar, B. N. Sari, and I. Maulana, "Penerapan Data Mining untuk Pengelompokan Produk Penjualan Menggunakan Algoritma K-Means," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 1, 2025, doi: 10.36040/jati.v9i1.12178.

[21] E. Erlinda, H. Nopriandi, and D. P. S. Journal, "Penerapan Data Mining Metode K-Means Clustering untuk Analisa Penjualan pada Toko Nusa Indah," *Startup Journal*, vol. 2, no. 1, 2025, Accessed: Nov. 24, 2025. [Online]. Available: https://journal.unilak.ac.id/index.php/Startup/article/view/27384

[22] U. Surapati and M. Jannah, "Penerapan Data Mining Menggunakan Metode K-Means Untuk Mengetahui Minat Customer dalam Pembelian Merchandise Kpop," *Jurnal Sains dan Teknologi*, vol. 5, no. 3, pp. 875–884, 2024, doi: 10.55338/saintek.v5i3.2739.

p-ISSN : 2803-1507 e-ISSN : 2803-1531

[23] I. Pii, N. Suarna, and N. R. J., "Penerapan Data Mining pada Penjualan Produk Pakaian Dameyra Fashion Menggunakan Metode K-Means Clustering," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 423-430, 2023, Accessed: Nov. 23, 2025. [Online]. Available: https://www.ejournal.itn.ac.id/jati/article/view/6336

[24] I. Anikah, A. Surip, N. Puji Rahayu, M. H. Al-Musa, and E. Tohidi, "Pengelompokan Data Barang dengan Menggunakan Metode K-Means untuk Menentukan Stok Persediaan Barang," *KOPERTIP: Jurnal Ilmiah Manajemen Informatika dan Komputer*, vol. 4, no. 2, pp. 58–64, Jun. 2022, doi: 10.32485/kopertip.v4i2.120.

[25] Nur Astuti, J. N. Utamajaya, and A. Pratama, "Penerapan Data Mining pada Penjualan Produk Digital Konter Leppangeng Cell Menggunakan Metode K-Means Clustering," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 3, pp. 754–760, 2022, doi: 10.30865/jurikom.v9i3.4351.

[26] N. M.-P. J. I. Komputer, "Penerapan Data Mining Pada Penjualan Produk Menggunakan Metode K-Means Clustering," *Repository UMB*, 2024, Accessed: Nov. 24, 2025. [Online]. Available: https://repository.umb.ac.id/846

[27] M. Sulistio, N. S., and J. T. I. K., "Analisa Penerapan Metode Clustering X-Means Dalam Pengelompokan Penjualan Barang," *Jurnal Teknologi Informasi dan Komunikasi (JTIK)*, vol. 1, no. 2, pp. 37–42, 2023, doi: 10.56854/jtik.v1i2.49.

[28] F. Amin, D. Anggraeni, and Q. A.-I. S., "Penerapan Metode K-Means dalam Penjualan Produk Souq. Com," *AISM (Advances in Information Systems and Management)*, vol. 5, no. 1, pp. 7-14, 2022, doi: 10.15408/aism.v5i1.22534.