



Implementasi Data Mining K-Means Clustering Untuk Pengelompokan Produk Keramik Berdasarkan Frekuensi, Volume, dan Jangkauan Penjualan

Ferdian Arya Dinata, Alwis Nazir*, Fadhilah Syafria, Teddie Darmizal, Eka Pandu Cynthia

Fakultas Sains dan Teknologi, Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ¹11950115065@student.uin-suska.ac.id, ^{2,*}alwis.nazir@uin-suska.ac.id, ³fadhilah.syafria@uin-suska.ac.id,

⁴teddie.darmizal@uin-suska.ac.id, ⁵eka.pandu.cynthia@uin-suska.ac.id

Email Penulis Korespondensi: alwis.nazir@uin-suska.ac.id

Abstrak—Pengelolaan stok keramik pada CV. Makmur Bersama selama ini umumnya masih dilakukan berdasarkan intuisi atau data penjualan yang dilihat secara parsial, tanpa mempertimbangkan pola perilaku pembelian secara menyeluruh. Pendekatan semacam ini menimbulkan dua risiko utama secara bersamaan, yaitu penumpukan produk yang jarang terjual sehingga membebani modal kerja dan ruang penyimpanan, serta kekosongan stok pada produk yang justru diminati pasar sehingga berpotensi menghilangkan peluang penjualan. Permasalahan ini diperburuk oleh keterbatasan data stok yang umumnya hanya memuat satu variabel kuantitatif, seperti jumlah unit terjual, sehingga belum mampu menggambarkan karakteristik permintaan produk secara komprehensif, misalnya seberapa sering produk dibeli atau seberapa luas basis pelanggannya. Akibatnya, keputusan restock dan strategi promosi sering kali tidak tepat sasaran. Penelitian ini bertujuan menerapkan algoritma K-Means untuk mengelompokkan produk keramik berdasarkan pola historis penjualan, sebagai solusi atas keterbatasan tersebut. Data historis penjualan periode 2025 pada CV. Makmur Bersama sebanyak 6.328 transaksi diolah menjadi 417 produk unik melalui pendekatan rekayasa fitur Frequency, Monetary, dan Reach (FMR), yaitu jumlah transaksi, total kuantitas terjual, dan jumlah pelanggan unik per produk. Setelah dilakukan deteksi outlier menggunakan metode Interquartile Range (IQR), tersisa 381 produk yang digunakan dalam proses clustering. Penentuan jumlah cluster optimal dilakukan melalui Elbow Method, menghasilkan $k=4$ sebagai jumlah cluster terbaik. Evaluasi menggunakan Davies-Bouldin Index (DBI) menghasilkan nilai 0,8954, yang dikategorikan baik, serta diuji stabilitasnya melalui lima kali iterasi dengan random state berbeda dan menunjukkan hasil yang konsisten (standar deviasi DBI sebesar 0,0034). Hasil pengelompokan menghasilkan Cluster 1 (190 produk, 49,9%) sebagai produk lambat terjual, Cluster 2 (34 produk, 8,9%) sebagai produk unggulan dengan rata-rata frekuensi transaksi 30,8 kali, Cluster 3 (93 produk, 24,4%) sebagai produk potensial, dan Cluster 4 (64 produk, 16,8%) sebagai produk dengan permintaan terbatas. Penelitian ini memberikan kontribusi praktis bagi perusahaan dalam menentukan prioritas restock, strategi promosi, dan efisiensi pengelolaan modal kerja berdasarkan pola penjualan aktual. Penelitian ini berkontribusi secara metodologis melalui adaptasi kerangka RFM menjadi FMR yang lebih sesuai dengan keterbatasan data riil, serta pengintegrasian Elbow Method, Davies-Bouldin Index, dan uji stabilitas sebagai mekanisme validasi yang komprehensif. Secara praktis, hasil segmentasi dapat langsung dimanfaatkan perusahaan sebagai dasar prioritas restocking, strategi promosi, dan efisiensi alokasi modal kerja berdasarkan pola penjualan aktual.

Kata Kunci: K-Means Clustering; Data Mining; FMR; Davies-Bouldin Index; Penjualan Keramik

Abstract—Ceramic inventory management at CV. Makmur Bersama has generally relied on intuition or partial sales data, without accounting for purchasing behavior patterns as a whole. This approach simultaneously creates two major risks: overstocking of slow-moving products, which burdens working capital and storage space, and stockouts of high-demand products, which can result in lost sales opportunities. This problem is further compounded by the limitation of stock data, which typically contains only a single quantitative variable such as the number of units sold and is therefore unable to comprehensively capture product demand characteristics, such as how frequently a product is purchased or how broad its customer base is. As a result, restocking decisions and promotional strategies are often poorly targeted. This research applies the K-Means algorithm to cluster ceramic products based on historical sales patterns as a solution to this limitation. Historical sales data from CV. Makmur Bersama for the 2025 period, consisting of 6,328 transactions, was processed into 417 unique products through a feature engineering approach using Frequency, Monetary, and Reach (FMR) namely transaction count, total quantity sold, and unique customer count per product. After outlier detection using the Interquartile Range (IQR) method, 381 products remained for the clustering process. The optimal number of clusters was determined using the Elbow Method, resulting in $k=4$ as the best cluster count. Evaluation using the Davies-Bouldin Index (DBI) produced a value of 0.8954, categorized as good, and stability testing across five iterations with different random states showed consistent results (DBI standard deviation of 0.0034). The clustering results produced Cluster 1 (190 products, 49.9%) as slow-moving products, Cluster 2 (34 products, 8.9%) as top-performing products with an average transaction frequency of 30.8 times, Cluster 3 (93 products, 24.4%) as potential products, and Cluster 4 (64 products, 16.8%) as products with limited demand. This research provides practical contributions for companies in determining restocking priorities, promotional strategies, and working capital efficiency based on actual sales patterns. This research contributes methodologically through the adaptation of the RFM framework into FMR to better suit real-world data constraints, as well as the integration of the Elbow Method, Davies-Bouldin Index, and stability testing as a comprehensive validation mechanism. Practically, the segmentation results can be directly utilized by the company as a basis for restocking priorities, promotional strategies, and working capital allocation efficiency based on actual sales patterns.

Keywords: K-Means Clustering; Data Mining; FMR; Davies-Bouldin Index; Ceramic Sales

1. PENDAHULUAN

Penjualan merupakan instrumen krusial dalam memenuhi kebutuhan pelanggan yang dinamis serta menjadi indikator utama keberhasilan operasional perusahaan [1]. Dalam konteks industri material bangunan, bisnis keramik muncul sebagai salah satu peluang usaha yang menjanjikan akibat tingginya permintaan pasar yang dipicu oleh pesatnya pembangunan infrastruktur dan properti [2], [3]. Pertumbuhan sektor konstruksi yang terus berlanjut di berbagai daerah turut mendorong permintaan terhadap bahan bangunan, termasuk keramik, baik untuk kebutuhan perumahan baru maupun



renovasi bangunan lama. Sebagai elemen estetika sekaligus struktural bangunan, permintaan terhadap keramik terus meningkat, namun keragaman jenis, ukuran, dan merek produk yang sangat banyak membuat perusahaan kesulitan memantau pola penjualan setiap item secara manual. Kondisi ini menjadi semakin menantang ketika jumlah produk yang dikelola mencapai ratusan jenis, sementara sumber daya manusia dan waktu yang tersedia untuk melakukan analisis manual terhadap setiap item sangat terbatas.

CV. Makmur Bersama, sebagai salah satu pelaku usaha distribusi keramik, menghadapi tantangan dalam mengidentifikasi produk mana yang benar-benar laris, mana yang stagnan, dan mana yang berpotensi tumbuh, di tengah ratusan jenis produk yang dijual setiap periode. Tanpa adanya sistem analisis data yang memadai, keputusan terkait pengadaan barang, penentuan stok minimum, dan strategi promosi cenderung diambil berdasarkan pengalaman dan intuisi semata. Pendekatan semacam ini berisiko menimbulkan dua kondisi yang merugikan secara bersamaan, yaitu penumpukan produk yang jarang terjual sehingga membebani modal kerja dan ruang penyimpanan gudang, serta di sisi lain terjadinya kekosongan stok pada produk yang justru diminati pasar, yang berpotensi menghilangkan peluang penjualan dan menurunkan tingkat kepuasan pelanggan.

Permasalahan ini menjadi lebih kompleks ketika data yang tersedia di perusahaan hanya berupa catatan stok dengan satu variabel kuantitatif, yaitu jumlah unit barang. Ketergantungan pada satu variabel tunggal membuat metode clustering konvensional seperti K-Means yang idealnya bekerja pada ruang multidimensi menjadi kurang bermakna secara analitis, karena pengelompokan yang dihasilkan tidak dapat menggambarkan perilaku pasar yang sesungguhnya. Sebagai contoh, dua produk dengan jumlah unit terjual yang sama dapat memiliki karakteristik permintaan yang sangat berbeda satu produk mungkin dibeli dalam jumlah besar oleh satu pelanggan dalam satu transaksi, sementara produk lain dibeli secara rutin oleh banyak pelanggan dalam transaksi-transaksi kecil. Jika hanya mengandalkan variabel jumlah unit, kedua produk tersebut akan dianggap setara, padahal secara bisnis keduanya membutuhkan strategi pengelolaan yang berbeda. Padahal, data transaksi penjualan historis sejatinya menyimpan informasi yang jauh lebih kaya, seperti seberapa sering suatu produk dibeli, berapa total kuantitas yang terjual, dan seberapa luas jangkauan pelanggan yang membeli produk tersebut. Informasi semacam ini, jika diekstraksi dengan tepat, dapat menjadi dasar pengelompokan yang lebih representatif terhadap kondisi pasar aktual dibandingkan sekadar mengandalkan angka stok semata.

Implementasi data mining melalui teknik clustering menjadi solusi efektif dalam mengekstraksi informasi berharga dari data transaksi historis untuk mendukung keputusan manajerial [4], [5]. Teknik clustering berfungsi mengelompokkan data berdasarkan karakteristik tertentu untuk menemukan pola yang tersembunyi [6]. Algoritma K-Means sering dipilih sebagai metode utama karena efisiensi komputasi dan kecepatannya dalam memproses data skala besar [7] di berbagai sektor bisnis [8]. Selain efisiensinya, K-Means juga relatif mudah diinterpretasikan oleh pengguna non-teknis, sehingga hasil pengelompokan dapat langsung dimanfaatkan oleh manajemen perusahaan tanpa memerlukan keahlian analisis data yang mendalam. Karakteristik inilah yang membuat K-Means banyak diadopsi pada penelitian-penelitian terapan di lingkungan bisnis kecil dan menengah.

Pendekatan berbasis frekuensi dan nilai transaksi seperti kerangka Recency, Frequency, Monetary (RFM) telah banyak digunakan dalam segmentasi pelanggan maupun produk berbasis data transaksi [9]. Kerangka RFM pada dasarnya dirancang untuk menangkap tiga dimensi penting dari perilaku transaksi, yaitu seberapa baru (recency), seberapa sering (frequency), dan seberapa besar nilai (monetary) suatu transaksi dilakukan. Namun, ketika data historis tidak memuat informasi temporal yang memadai untuk menghitung Recency secara bermakna misalnya karena rentang waktu pencatatan yang terlalu singkat atau pola transaksi yang tidak mencerminkan siklus pembelian yang jelas beberapa peneliti mengadaptasi kerangka tersebut dengan variabel pengganti yang lebih relevan terhadap konteks data yang tersedia [10]. Adaptasi semacam ini penting dilakukan agar kerangka analisis tetap dapat diterapkan secara bermakna meskipun keterbatasan data tetap ada, tanpa mengorbankan validitas hasil segmentasi yang dihasilkan.

Studi terdahulu telah membuktikan keberhasilan K-Means dalam meningkatkan efektivitas strategi penjualan pakaian [1], bahan bangunan [11], hingga klasifikasi tingkat laris pada penjualan mobil [9]. Penelitian lain juga menunjukkan bahwa kombinasi K-Means dengan kerangka RFM efektif digunakan untuk pemasaran presisi berbasis pola transaksi [9]. Selain itu, integrasi metode elbow telah terbukti akurat dalam menentukan jumlah kluster yang optimal melalui evaluasi grafik penurunan nilai kriteria secara objektif [12], sementara Davies-Bouldin Index (DBI) banyak digunakan sebagai metrik validasi internal karena kemudahannya dalam mengukur rasio kohesi dan separasi antar cluster [13], [14]. Penggunaan kombinasi Elbow Method dan DBI secara bersamaan memungkinkan proses penentuan jumlah cluster tidak hanya bergantung pada interpretasi visual yang bersifat subjektif, tetapi juga didukung oleh ukuran kuantitatif yang dapat dipertanggungjawabkan secara ilmiah.

Berbeda dengan penelitian sebelumnya yang umumnya mengandalkan data stok sebagai variabel tunggal [15], [16] penelitian ini mengisi gap tersebut dengan mengekstraksi tiga variabel baru dari data historis penjualan, yaitu Frequency (jumlah transaksi), Monetary (total kuantitas terjual), dan Reach (jumlah pelanggan unik), sebagai adaptasi dari kerangka RFM yang disesuaikan dengan ketersediaan data riil di lapangan. Pendekatan ini selanjutnya disebut sebagai kerangka FMR. Berdasarkan latar belakang tersebut, rumusan masalah yang diangkat dalam penelitian ini adalah bagaimana proses clustering data historis penjualan keramik pada CV. Makmur Bersama periode 2025 dapat diimplementasikan menggunakan algoritma K-Means berbasis variabel FMR untuk menghasilkan segmentasi produk yang bermakna secara bisnis, serta bagaimana hasil segmentasi tersebut dapat divalidasi secara objektif agar dapat dipercaya dan diandalkan dalam pengambilan keputusan manajerial.

Penelitian ini bertujuan untuk melakukan pengelompokan (clustering) produk keramik berdasarkan pola historis penjualan guna memberikan rekomendasi yang presisi dalam manajemen persediaan dan strategi pemasaran. Analisis ini



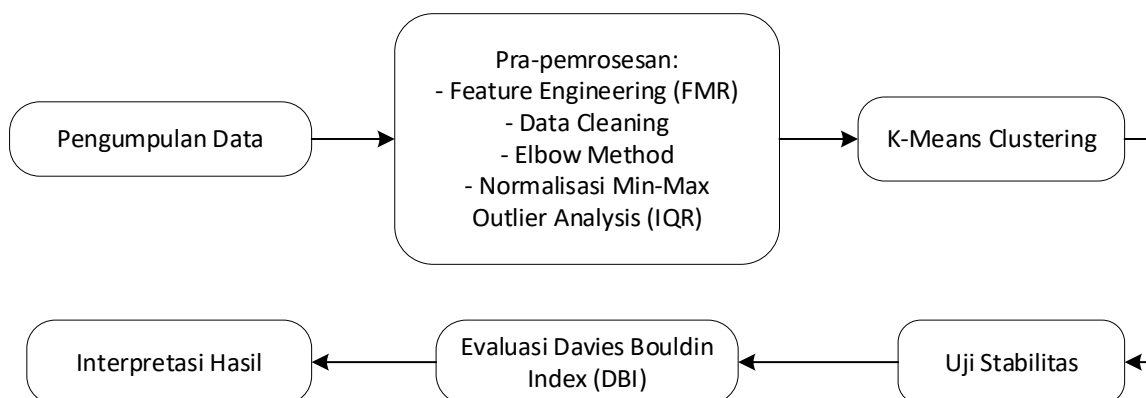
tidak hanya mengandalkan pengelompokan semata, tetapi juga memanfaatkan metode elbow untuk penentuan jumlah kluster, evaluasi menggunakan Davies-Bouldin Index (DBI), serta uji stabilitas cluster untuk memastikan hasil yang konsisten dan dapat dipertanggungjawabkan secara ilmiah. Melalui pendekatan ini, penelitian ini diharapkan dapat menjadi landasan bagi perusahaan dalam mengambil keputusan strategis terkait prioritas restock, evaluasi produk slow-moving, serta efisiensi alokasi modal kerja.

Kontribusi penelitian ini dapat diidentifikasi pada tiga aspek. Pertama, pada aspek metodologis, penelitian ini mengusulkan adaptasi kerangka RFM menjadi FMR (Frequency, Monetary, Reach) sebagai solusi atas keterbatasan data yang umum dijumpai pada usaha distribusi berskala menengah, di mana dimensi Recency tidak dapat dihitung secara bermakna karena keterbatasan rentang waktu pencatatan. Kedua, pada aspek teknis, penelitian ini menggabungkan Elbow Method dan Davies-Bouldin Index sebagai mekanisme validasi ganda, serta memperkuatnya dengan uji stabilitas pada lima random state berbeda, sehingga hasil clustering yang diperoleh tidak hanya optimal secara statistik tetapi juga terbukti konsisten dan dapat direproduksi. Ketiga, pada aspek praktis, penelitian ini menghasilkan segmentasi produk yang dapat langsung diterapkan oleh manajemen CV. Makmur Bersama dalam menentukan prioritas restocking, merancang strategi promosi yang berbeda untuk setiap segmen, serta mengoptimalkan alokasi modal kerja dan ruang penyimpanan gudang.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini dilakukan melalui serangkaian tahapan yang terstruktur, dimulai dari pengumpulan data historis penjualan, dilanjutkan dengan rekayasa fitur (feature engineering), analisis outlier, normalisasi data, penentuan jumlah cluster optimal, implementasi algoritma K-Means, hingga evaluasi dan interpretasi hasil. Tahapan tersebut digambarkan secara ringkas pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1.1 Pengumpulan Data

Data penelitian ini bersumber dari catatan historis penjualan keramik CV. Makmur Bersama periode tahun 2025 yang diperoleh dalam format Microsoft Excel, mencakup 6.328 baris transaksi. Atribut yang digunakan meliputi nama pelanggan, nama produk, tanggal transaksi, nomor nota, kuantitas (Qty), satuan, kode barang, dan merek.

2.1.2 Pra-pemrosesan

a. Rekayasa Fitur (Feature Engineering)

Karena data stok asli hanya memuat satu variabel kuantitatif yang tidak memadai untuk clustering multidimensi, penelitian ini mengekstraksi tiga variabel baru dari data historis transaksi berdasarkan pendekatan FMR sebagai berikut:

1. Frequency, dihitung dari jumlah nomor nota (transaksi) unik per produk.
2. Monetary, dihitung dari total kuantitas (Qty) yang terjual per produk.
3. Reach, dihitung dari jumlah pelanggan unik yang membeli produk tersebut.

Ketiga variabel ini diperoleh melalui proses agregasi data transaksi berdasarkan nama produk.

b. Analisis Outlier

Deteksi outlier dilakukan menggunakan metode Interquartile Range (IQR) pada masing-masing variabel FMR. Data yang berada di luar rentang $[Q1 - 1,5 \times IQR, Q3 + 1,5 \times IQR]$ dikategorikan sebagai outlier dan dikeluarkan dari proses clustering utama agar tidak menarik posisi centroid secara tidak proporsional.

c. Normalisasi Data

Data yang telah bersih dari outlier dinormalisasi menggunakan metode Min-Max Scaling agar seluruh variabel memiliki rentang nilai yang setara $[0,1]$, mengingat skala asli Monetary jauh lebih besar dibandingkan Frequency dan Reach.



2.2 Algoritma K-Means

K-Means merupakan salah satu metode clustering non-hirarki yang mempartisi data ke dalam sejumlah cluster, sehingga data dengan karakteristik yang sama dikelompokkan ke dalam cluster yang sama [17]. Berikut adalah tahapan pada algoritma K-Means:

- Menentukan jumlah cluster (k) menggunakan Elbow Method.
- Menginisialisasi centroid awal menggunakan metode k-means++.
- Menghitung jarak setiap data ke centroid menggunakan rumus Euclidean Distance[18]:

$$d(x_i, y_i) = \sqrt{\sum_{i=1}^n |x_i - y_i|^2} \quad (1)$$

Mengelompokkan data berdasarkan jarak minimum terhadap centroid.

- Memperbarui nilai centroid dengan menghitung rata-rata anggota setiap cluster[19]:

$$C_k = \frac{1}{n_k} \sum d_i \quad (2)$$

dimana n_k adalah jumlah data dalam cluster, dan d_i adalah nilai data yang menjadi anggota cluster tersebut.

- Mengulangi langkah 3–5 hingga tidak ada lagi perpindahan anggota cluster (konvergen).

Adapun parameter K-Means yang digunakan dalam penelitian ini dijelaskan pada Tabel 1.

Tabel 1. Parameter K-Means yang Digunakan

Parameter	Nilai	Keterangan
n_clusters (k)	4	Jumlah cluster hasil Elbow Method
init	k-means++	Inisialisasi centroid cerdas untuk mempercepat konvergensi
max_iter	300	Batas maksimum iterasi per proses run
n_init	10	Jumlah pengulangan algoritma dengan centroid awal berbeda
random_state	42	Seed untuk menjamin reproduktifitas hasil
algorithm	lloyd	Algoritma standar K-Means

2.3 Elbow Method

Elbow Method adalah salah satu metode yang paling umum digunakan untuk membedakan jumlah cluster yang optimal, diskriminan jumlah cluster tergantung pada identifikasi titik siku secara manual pada kurva visualisasi [20]. Metode elbow mudah diimplementasikan dengan melihat grafik nilai k ideal dengan posisi pada siku beserta SSE (Sum of Square Error) yang kurang dari 1. Hasil k cluster terbaik akan menjadi dasar pengelompokan. Semakin kecil nilai SSE dan semakin kecil grafik siku maka hasil cluster semakin baik. Berikut adalah persamaan untuk menghitung nilai SSE.

$$SSE = \sum_{i=1}^n (d)^2 \quad (3)$$

Dimana d adalah jarak antara data dan pusat klaster

2.4 Davies-Bouldin Index (DBI)

Davies-Bouldin Index (DBI) digunakan sebagai metrik evaluasi internal untuk menilai kualitas hasil clustering, dengan mengukur rasio antara kekompakan dalam cluster (intra-cluster) dan pemisahan antar cluster (inter-cluster) [15]. Nilai DBI dihitung melalui tahapan berikut:

- Menghitung Sum of Square Within-Cluster (SSW) untuk mengetahui kohesi dalam cluster ke- i [20]:

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(X_j, C_i) \quad (4)$$

Dimana m_i adalah jumlah data dalam cluster ke- i dan C_i adalah centroid cluster ke- i

- Menghitung Sum of Square Between-Cluster (SSB) untuk mengetahui separasi antar cluster:

$$SSB_{ij} = d(X_i, X_j) \quad (5)$$

Dimana $d(X_i, X_j)$ adalah jarak antara data ke- i dengan data ke- j di cluster lain

- Menghitung rasio R_{ij} antar cluster:

$$R_{ij \dots n} = \frac{SSW_i + SSW_j + \dots + SSW_n}{SSB_{i,j} + \dots + SSB_{ni,nj}} \quad (6)$$

Dimana SSW_i adalah Sum of Square Within-Cluster pada centroid i dan SSB_i adalah Sum of Square Between Cluster data ke- i dengan j pada cluster yang berbeda.

- Menghitung nilai DBI akhir:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j,\dots,k}) \quad (7)$$

Dimana, $(R_{i,j,\dots,k})$ merupakan ratio dari nilai SSW dan SSB melalui persamaan 5 maka dapat diketahui k adalah jumlah cluster. Dari perhitungan Davies Bouldin Index (DBI) dapat disimpulkan bahwa jika semakin kecil nilai Davies Bouldin Index (DBI) yang diperoleh (non negatif ≥ 0) maka cluster tersebut semakin baik.

3. HASIL DAN PEMBAHASAN

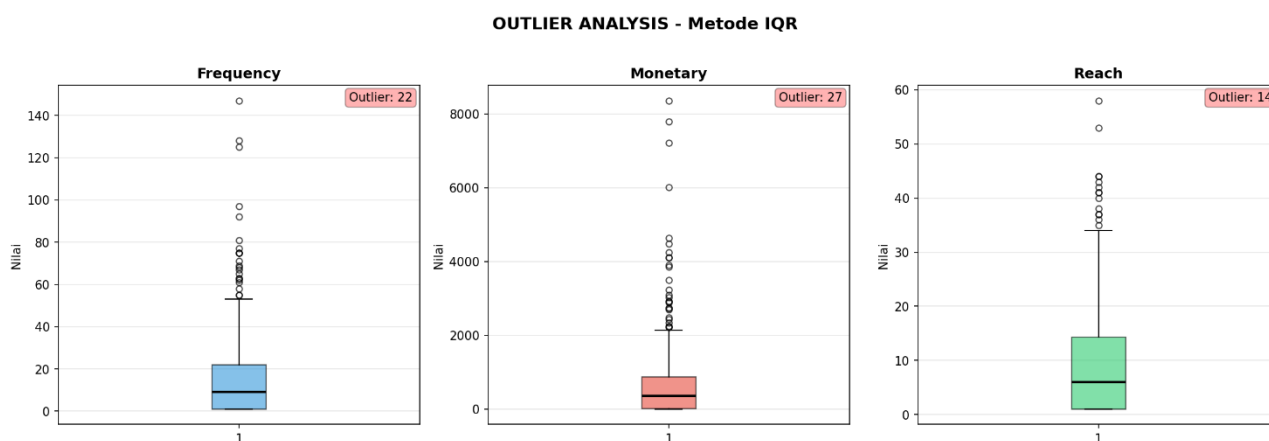
3.1 Analisis Deskriptif dan Pra-pemrosesan Data

Data penelitian ini bersumber dari 6.328 baris transaksi penjualan keramik CV. Makmur Bersama sepanjang tahun 2025. Melalui proses agregasi berdasarkan nama produk, diperoleh 417 produk unik dengan tiga variabel hasil rekayasa fitur, yaitu Frequency, Monetary, dan Reach.

Proses agregasi ini dilakukan melalui beberapa tahapan teknis yang berurutan. Pertama, data transaksi mentah yang masih dalam bentuk baris per item per nota dikelompokkan (grouped) berdasarkan nama produk menggunakan operasi groupby. Pada tahap ini, nilai Frequency dihitung dengan menghitung jumlah nomor nota unik (distinct count) yang memuat produk tersebut, sehingga satu nota yang memuat beberapa item produk yang sama tetap dihitung sebagai satu transaksi untuk produk itu. Nilai Monetary diperoleh dengan menjumlahkan seluruh nilai kolom kuantitas (Qty) pada setiap baris transaksi yang memuat produk tersebut, terlepas dari berapa kali produk itu muncul di nota yang berbeda.

Sementara itu, nilai Reach dihitung dengan menghitung jumlah nilai unik pada kolom nama pelanggan yang pernah membeli produk tersebut sepanjang periode pengamatan. Ketiga proses agregasi ini menghasilkan satu baris data per produk, dengan tiga kolom numerik yang merepresentasikan perilaku transaksi historisnya secara ringkas.

Sebelum dilakukan clustering, ketiga variabel ini dianalisis terlebih dahulu untuk mendeteksi keberadaan outlier yang dapat mengganggu validitas pengelompokan. Hasil analisis outlier menggunakan metode Interquartile Range (IQR) ditampilkan pada Gambar 2.



Gambar 2. Boxplot Analisis Outlier pada Variabel Frequency, Monetary, dan Reach

Berdasarkan Gambar 2, ditemukan 22 produk outlier pada variabel Frequency, 27 produk pada variabel Monetary, dan 14 produk pada variabel Reach. Setelah dilakukan union terhadap ketiga deteksi tersebut, terdapat 36 produk yang dikategorikan sebagai outlier dan dikeluarkan dari proses clustering utama.

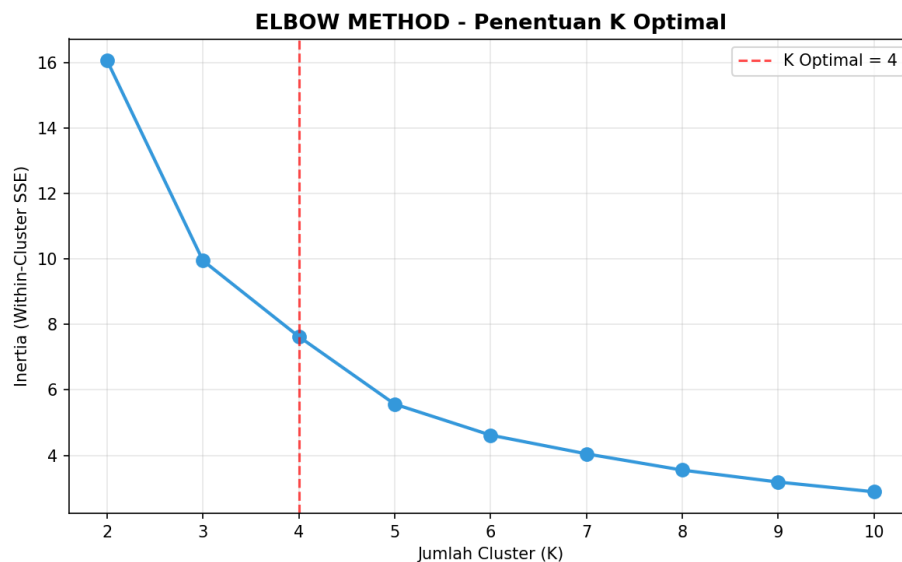
Secara teknis, batas outlier pada masing-masing variabel ditentukan dengan menghitung kuartil pertama (Q1) dan kuartil ketiga (Q3) dari distribusi data, kemudian menghitung rentang interkuartil ($IQR = Q3 - Q1$). Suatu data dianggap outlier apabila nilainya berada di bawah batas bawah ($Q1 - 1,5 \times IQR$) atau di atas batas atas ($Q3 + 1,5 \times IQR$). Karena deteksi dilakukan secara independen pada masing-masing dari ketiga variabel (Frequency, Monetary, Reach), satu produk dapat terdeteksi sebagai outlier pada lebih dari satu variabel sekaligus. Oleh karena itu, dilakukan operasi union (penggabungan tanpa duplikasi) terhadap ketiga himpunan outlier tersebut, sehingga total 36 produk yang dikeluarkan dari proses clustering merupakan gabungan unik dari ketiga deteksi, bukan penjumlahan langsung $22+27+14$. Pendekatan union dipilih agar produk yang ekstrem pada satu dimensi pun meskipun normal pada dimensi lain tetap dikeluarkan, mengingat tujuan utama tahap ini adalah menjaga agar posisi centroid tidak tertarik secara tidak proporsional oleh segelintir produk yang nilainya jauh berbeda dari mayoritas. Produk-produk ini umumnya merupakan item dengan volume transaksi yang sangat ekstrem dibandingkan mayoritas produk lainnya, sehingga apabila tetap disertakan akan menarik posisi centroid secara tidak proporsional dan menurunkan kualitas interpretasi cluster. Dengan demikian, proses clustering selanjutnya dilakukan terhadap 381 produk yang tersisa.

Ketiga variabel kemudian dinormalisasi menggunakan metode Min-Max Scaling agar berada pada rentang $[0,1]$ yang setara, mengingat skala asli Monetary (mencapai ribuan unit) jauh lebih besar dibandingkan Frequency dan Reach yang hanya berkisar puluhan. Normalisasi Min-Max dilakukan dengan rumus $x' = (x - x_{\min}) / (x_{\max} - x_{\min})$ pada

masing-masing variabel secara independen, sehingga setiap variabel memiliki rentang nilai 0 hingga 1 setelah transformasi. Langkah ini bersifat krusial dalam algoritma K-Means karena perhitungan jarak Euclidean yang digunakan pada tahap clustering sangat sensitif terhadap skala data. Tanpa normalisasi, variabel Monetary yang memiliki rentang nilai hingga ribuan akan mendominasi perhitungan jarak secara tidak proporsional dibandingkan Frequency dan Reach yang hanya berkisar puluhan, sehingga cluster yang terbentuk pada akhirnya hanya akan merefleksikan pola Monetary saja dan mengabaikan kontribusi dua variabel lainnya. Dengan menyamakan skala ketiga variabel pada rentang [0,1], setiap variabel memiliki bobot pengaruh yang setara dalam menentukan posisi suatu produk pada ruang fitur tiga dimensi, sehingga hasil pengelompokan benar-benar mencerminkan kombinasi pola Frequency, Monetary, dan Reach secara seimbang.

3.2 Penentuan Cluster Optimal

Penentuan jumlah cluster (k) dilakukan dengan menggunakan Elbow Method pada rentang $k=2$ hingga $k=10$, sebagaimana ditampilkan pada Gambar 3. Proses Elbow Method dilakukan dengan menjalankan algoritma K-Means secara berulang untuk setiap nilai k pada rentang yang ditentukan, kemudian mencatat nilai inerti (Sum of Squared Error/SSE) yang dihasilkan pada masing-masing nilai k tersebut. Inerti merepresentasikan total jarak kuadrat antara setiap titik data dengan centroid cluster tempatnya berada; semakin kecil nilai inerti, semakin kompak data di dalam clusternya. Namun, nilai inerti akan selalu menurun seiring bertambahnya jumlah cluster k , sehingga diperlukan identifikasi titik "siku" (elbow point) pada grafik, yaitu titik di mana penurunan inerti mulai melandai secara signifikan. Titik ini merepresentasikan jumlah cluster yang memberikan keseimbangan terbaik antara kompleksitas model (jumlah cluster) dan kualitas pengelompokan (nilai inerti).



Gambar 3. Grafik Elbow Method untuk Penentuan Jumlah Cluster Optimal

Berdasarkan Gambar 3, terlihat penurunan nilai inerti yang tajam dari $k=2$ (inerti=16,07) hingga $k=4$ (inerti=7,62), namun penurunan menjadi lebih landai pada $k=5$ ke atas. Titik siku (elbow point) yang teridentifikasi berada pada $k=4$, yang menandakan bahwa penambahan cluster lebih dari 4 tidak lagi memberikan penurunan inerti yang signifikan secara proporsional. Oleh karena itu, penelitian ini menetapkan $k=4$ sebagai jumlah cluster optimal untuk proses K-Means selanjutnya.

Pemilihan $k=4$ ini juga sejalan dengan kebutuhan interpretasi bisnis pada penelitian ini. Jumlah cluster yang terlalu sedikit (misalnya $k=2$ atau $k=3$) berisiko menggabungkan produk dengan karakteristik permintaan yang sebenarnya berbeda ke dalam satu kelompok yang sama, sehingga rekomendasi strategis yang dihasilkan menjadi terlalu umum. Sebaliknya, jumlah cluster yang terlalu banyak (misalnya $k=8$ atau lebih) akan menghasilkan segmentasi yang terlalu granular dan menyulitkan manajemen dalam menyusun strategi yang berbeda untuk setiap kelompok kecil tersebut. Dengan $k=4$, hasil segmentasi yang diperoleh cukup representatif untuk membedakan tingkatan performa produk dari yang paling lambat terjual hingga yang paling unggul namun tetap praktis untuk diterjemahkan ke dalam kebijakan operasional perusahaan.

3.3 Implementasi K-Means dan Evaluasi DBI

Implementasi K-Means dilakukan dengan parameter yang telah dijelaskan pada Tabel 1 ($n_clusters=4$, $init=k-means++$, $max_iter=300$, $n_init=10$, $random_state=42$). Secara teknis, proses clustering K-Means pada data 381 produk yang telah dinormalisasi dijalankan melalui langkah-langkah berikut. Pertama, algoritma menginisialisasi 4 titik centroid awal menggunakan metode $k-means++$, yaitu strategi inisialisasi yang memilih centroid awal secara probabilistik berdasarkan jarak antar titik data, sehingga centroid awal cenderung tersebar dan tidak saling berdekatan berbeda dengan inisialisasi

acak murni yang berisiko menghasilkan konvergensi pada solusi lokal yang buruk. Kedua, setiap dari 381 produk dihitung jaraknya (Euclidean distance) terhadap keempat centroid tersebut pada ruang fitur tiga dimensi (Frequency, Monetary, Reach yang telah dinormalisasi), kemudian setiap produk dialokasikan ke cluster dengan centroid terdekat. Ketiga, posisi centroid diperbarui dengan menghitung rata-rata nilai seluruh produk yang menjadi anggota masing-masing cluster pada iterasi tersebut. Keempat, langkah kedua dan ketiga diulang secara iteratif hingga tidak ada lagi perpindahan anggota antar cluster (konvergen) atau hingga mencapai batas maksimum 300 iterasi. Proses keseluruhan ini diulang sebanyak 10 kali ($n_{init}=10$) dengan titik inialisasi yang berbeda-beda, dan hasil dengan nilai inertiaterkecil dipilih sebagai solusi akhir, guna meminimalkan risiko algoritma terjebak pada solusi lokal optimum yang kurang baik. Proses clustering menghasilkan nilai inertiakhir sebesar 7,6135 dengan Davies-Bouldin Index (DBI) sebesar 0,8954.

Nilai DBI ini dihitung melalui beberapa tahap perhitungan sebagaimana telah dijelaskan pada bagian Metodologi. Tahap pertama adalah menghitung nilai Sum of Square Within-Cluster (SSW) untuk masing-masing dari keempat cluster, yang merepresentasikan rata-rata jarak antara setiap anggota cluster dengan centroidnya sendiri nilai ini mencerminkan tingkat kekompakan (cohesion) internal cluster tersebut. Tahap kedua adalah menghitung jarak antar centroid dari setiap pasangan cluster (SSB), yang merepresentasikan tingkat pemisahan (separation) antar cluster. Tahap ketiga adalah menghitung rasio antara penjumlahan SSW dari sepasang cluster dengan jarak SSB antara keduanya, untuk setiap kemungkinan pasangan cluster, kemudian diambil nilai rasio maksimum untuk masing-masing cluster terhadap cluster pembandingnya yang paling mirip (worst-case pairing). Tahap terakhir adalah merata-ratakan keempat nilai rasio maksimum tersebut untuk memperoleh nilai DBI akhir. Nilai DBI yang rendah mengindikasikan bahwa setiap cluster memiliki anggota yang kompak (SSW kecil) sekaligus terpisah jauh dari cluster lain (SSB besar relatif terhadap SSW), yang merupakan kondisi ideal dalam evaluasi clustering.

Berdasarkan kategori interpretasi DBI, nilai di bawah 1,0 dikategorikan sangat baik, nilai 1,0–2,0 dikategorikan baik, dan nilai di atas 2,0 dikategorikan kurang baik. Nilai DBI sebesar 0,8954 yang diperoleh pada penelitian ini menunjukkan bahwa skema pengelompokan dengan $k=4$ menghasilkan kualitas cluster yang baik, dengan tingkat kekompakan intra-cluster yang memadai serta pemisahan antar cluster yang cukup jelas.

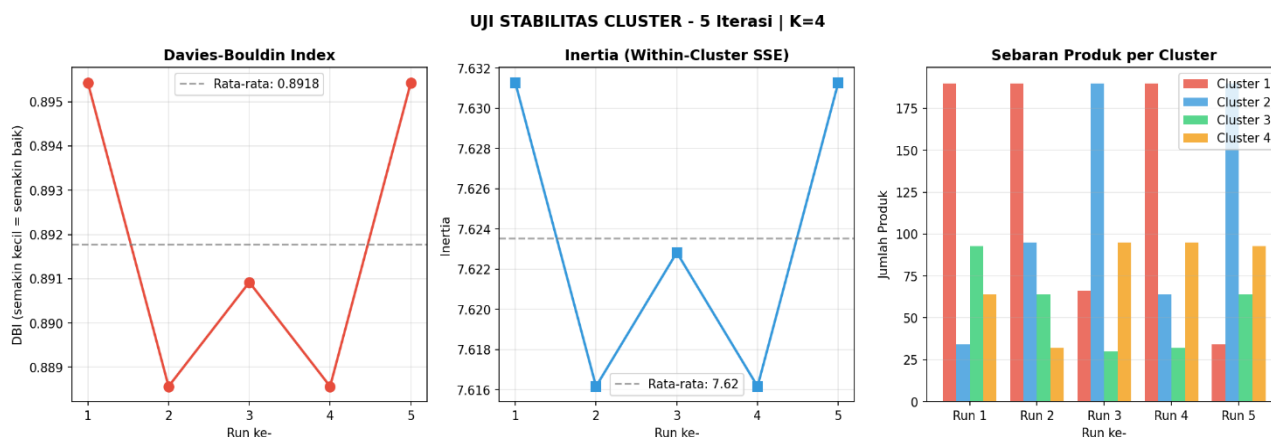
3.4 Uji Stabilitas Cluster

Untuk membuktikan bahwa hasil clustering tidak bergantung pada posisi inialisasi centroid yang bersifat acak, dilakukan uji stabilitas dengan menjalankan algoritma K-Means sebanyak 5 kali menggunakan random state berbeda. Hasil uji stabilitas ditampilkan pada Tabel 2 dan Gambar 4.

Pengujian stabilitas ini dilakukan dengan prosedur sebagai berikut: seluruh proses clustering pada bagian 3.3 mulai dari inialisasi centroid, iterasi pembaruan posisi centroid, hingga perhitungan nilai inertiadan DBI diulang sebanyak lima kali secara independen, dengan satu-satunya perbedaan adalah nilai `random_state` yang digunakan pada masing-masing run (42, 7, 13, 99, dan 2024). Penggunaan `random_state` yang berbeda menyebabkan titik centroid awal yang dipilih oleh metode `k-means++` juga berbeda pada setiap run, sehingga jika hasil clustering yang diperoleh konsisten meskipun titik awalnya berbeda-beda, hal ini menjadi bukti bahwa algoritma telah mencapai konvergensi yang stabil dan bukan sekadar kebetulan dari satu kali proses komputasi semata. Pada setiap run, dicatat nilai inertiadan DBI, serta distribusi jumlah anggota pada masing-masing cluster untuk kemudian dibandingkan satu sama lain.

Tabel 2. Hasil Uji Stabilitas Cluster pada 5 Iterasi

Run	Random State	Inertia	DBI	Distribusi Cluster (1 2 3 4)
1	42	7,6313	0,8954	190 34 93 64
2	7	7,6161	0,8885	190 95 64 32
3	13	7,6230	0,8909	66 189 30 96
4	99	7,6161	0,8885	189 64 32 96
5	2024	7,6313	0,8954	35 189 64 93



Gambar 4. Grafik Uji Stabilitas Cluster



Berdasarkan Tabel 2 dan Gambar 4, nilai DBI pada kelima percobaan berkisar antara 0,8885 hingga 0,8954 dengan standar deviasi sebesar 0,0034, sementara nilai inerti berkisar antara 7,6161 hingga 7,6313 dengan standar deviasi sebesar 0,0070. Variasi yang sangat kecil pada kedua metrik ini mengindikasikan bahwa algoritma K-Means telah mencapai konvergensi yang stabil. Meskipun label numerik cluster (1, 2, 3, 4) dapat bertukar posisi antar percobaan akibat sifat acak inialisasi K-Means, pola sebaran jumlah anggota pada kelima percobaan secara konsisten menunjukkan komposisi yang sama, yaitu satu cluster besar beranggotakan sekitar 189–190 produk, satu cluster beranggotakan sekitar 64–66 produk, satu cluster beranggotakan sekitar 93–96 produk, dan satu cluster beranggotakan sekitar 30–35 produk. Konsistensi pola distribusi inilah yang menjadi bukti ilmiah bahwa hasil cluster yang diperoleh bersifat stabil dan dapat dipercaya, bukan hasil kebetulan dari satu kali proses komputasi.

Fenomena pertukaran label numerik ini wajar terjadi pada algoritma K-Means karena penomoran cluster (Cluster 1, 2, 3, dst.) ditentukan secara arbitrer berdasarkan urutan inialisasi centroid, bukan berdasarkan karakteristik tertentu dari cluster tersebut. Misalnya pada Run 1, kelompok produk unggulan diberi label "Cluster 2" dengan 34 anggota, sementara pada Run 3 kelompok dengan karakteristik serupa (34 anggota dengan performa tertinggi) justru muncul sebagai "Cluster 3". Oleh karena itu, validasi stabilitas yang tepat tidak dilakukan dengan membandingkan label cluster secara langsung, melainkan dengan membandingkan pola distribusi ukuran (jumlah anggota) dan karakteristik statistik masing-masing kelompok pada setiap run. Konsistensi pola distribusi inilah yang menjadi bukti ilmiah bahwa hasil cluster yang diperoleh bersifat stabil dan dapat dipercaya, bukan hasil kebetulan dari satu kali proses komputasi.

3.5 Visualisasi dan Analisis Karakteristik Cluster

Visualisasi hasil clustering pada ruang variabel Frequency, Monetary, dan Reach ditampilkan pada Gambar 5.



Gambar 5. Visualisasi K-Means Clustering Produk Keramik

Berdasarkan Gambar 5, terlihat pemisahan yang cukup jelas antar keempat cluster pada ketiga kombinasi pasangan variabel (Frequency-Monetary, Frequency-Reach, Monetary-Reach), tanpa tumpang tindih (overlap) yang signifikan antar kelompok. Profil statistik masing-masing cluster dirinci pada Tabel 3.

Untuk memperoleh interpretasi karakteristik setiap cluster, dilakukan tahap akhir berupa profiling, yaitu menghitung nilai rata-rata (mean) dari ketiga variabel asli (sebelum normalisasi) untuk setiap anggota pada masing-masing cluster. Profiling ini penting dilakukan karena nilai centroid yang dihasilkan langsung dari algoritma K-Means berada dalam skala ternormalisasi [0,1] yang tidak mudah diinterpretasikan secara bisnis. Dengan mengembalikan



perhitungan ke skala asli (jumlah transaksi, jumlah unit, jumlah pelanggan), manajemen dapat memahami secara konkret seberapa besar perbedaan performa antar cluster, sebagaimana dirangkum pada Tabel 3.

Tabel 3. Ringkasan Statistik Cluster Produk Keramik

Cluster	N Produk	%	Avg Frequency	Avg Monetary (Qty)	Avg Reach	Klasifikasi
1	190	49,9%	1,9	42,9	1,6	Lambat Terjual
2	34	8,9%	30,8	1.448,6	18,5	Produk Unggulan
3	93	24,4%	21,4	809,3	14,7	Potensial
4	64	16,8%	11,0	478,8	8,0	Permintaan Terbatas

Berdasarkan Tabel 3, dapat dilakukan interpretasi mendalam terhadap pola penjualan masing-masing cluster sebagai berikut.

a. Cluster 1: Produk Lambat Terjual

Cluster 1 merupakan kelompok terbesar dengan 190 produk (49,9% dari total produk valid), namun memiliki rata-rata Frequency hanya 1,9 transaksi, Monetary 42,9 unit, dan Reach 1,6 pelanggan. Karakteristik ini menunjukkan bahwa hampir separuh dari total produk keramik CV. Makmur Bersama termasuk kategori slow-moving, yaitu hanya dibeli oleh satu hingga dua pelanggan sepanjang periode pengamatan dengan volume yang sangat kecil. Temuan ini memberikan sinyal penting bagi manajemen untuk melakukan evaluasi terhadap produk-produk dalam cluster ini, baik melalui strategi diskon, bundling dengan produk laris, maupun pengurangan frekuensi pemesanan ulang ke supplier guna menghindari penumpukan modal kerja yang tidak produktif. Besarnya proporsi cluster ini hampir mencapai separuh dari total produk yang dianalisis menjadi temuan yang patut menjadi perhatian utama perusahaan, karena mengindikasikan bahwa sebagian besar variasi produk yang disediakan justru tidak memberikan kontribusi signifikan terhadap volume penjualan maupun jangkauan pelanggan. Dari perspektif manajemen persediaan, kondisi ini berpotensi menyebabkan alokasi modal kerja yang tidak efisien apabila tidak ditindaklanjuti, mengingat ruang gudang dan dana yang tertanam pada produk-produk kategori ini dapat dialokasikan ulang untuk memperkuat ketersediaan produk pada cluster lain yang permintaannya lebih tinggi.

b. Cluster 2: Produk Unggulan

Cluster 2 merupakan kelompok terkecil dengan 34 produk (8,9%), namun menunjukkan performa tertinggi pada seluruh variabel, dengan rata-rata Frequency 30,8 transaksi, Monetary 1.448,6 unit, dan Reach 18,5 pelanggan. Produk dalam cluster ini merupakan produk best seller yang secara konsisten diminati banyak pelanggan dengan frekuensi pembelian yang tinggi. Manajemen disarankan untuk memprioritaskan ketersediaan stok pada cluster ini guna mencegah risiko kehabisan stok (stockout) yang berpotensi menghilangkan peluang penjualan. Meskipun jumlah anggotanya hanya 8,9% dari total produk yang dianalisis, kontribusi cluster ini terhadap total volume transaksi perusahaan kemungkinan jauh lebih besar dibandingkan proporsi jumlah produknya, mengingat rata-rata kuantitas terjual pada cluster ini (1.448,6 unit) lebih dari 30 kali lipat dibandingkan Cluster 1. Pola ini sejalan dengan prinsip umum dalam manajemen persediaan yang sering disebut sebagai analisis Pareto (aturan 80/20), di mana sebagian kecil produk justru menyumbang sebagian besar volume penjualan. Implikasinya, kebijakan pengadaan barang perusahaan sebaiknya memprioritaskan ketersediaan stok pada cluster ini secara konsisten, termasuk mempertimbangkan kerja sama jangka panjang dengan supplier untuk menjamin pasokan yang stabil.

c. Cluster 3: Produk Potensial

Cluster 3 beranggotakan 93 produk (24,4%) dengan rata-rata Frequency 21,4 transaksi, Monetary 809,3 unit, dan Reach 14,7 pelanggan. Performa cluster ini berada pada posisi menengah ke atas, mendekati karakteristik Cluster 2 namun belum sekuat produk unggulan. Produk-produk pada cluster ini berpotensi untuk ditingkatkan penjualannya melalui strategi promosi tambahan agar dapat bergeser ke kategori produk unggulan. Jika dibandingkan dengan Cluster 2, rata-rata nilai Frequency pada Cluster 3 setara dengan sekitar 69% dari Cluster 2, sementara Monetary dan Reach masing-masing setara dengan sekitar 56% dan 79%. Kesenjangan yang relatif kecil pada variabel Reach dibandingkan Monetary mengindikasikan bahwa produk-produk pada cluster ini sebenarnya sudah memiliki basis pelanggan yang cukup luas, namun belum dibeli dalam kuantitas besar oleh masing-masing pelanggan tersebut. Temuan ini membuka peluang strategi pemasaran yang lebih spesifik, misalnya melalui penawaran pembelian dalam jumlah lebih besar (cross-selling atau up-selling) kepada pelanggan yang sudah ada, dibandingkan strategi yang berfokus pada akuisisi pelanggan baru.

d. Cluster 4: Produk dengan Permintaan Terbatas

Cluster 4 terdiri dari 64 produk (16,8%) dengan rata-rata Frequency 11,0 transaksi, Monetary 478,8 unit, dan Reach 8,0 pelanggan. Cluster ini berada pada posisi menengah ke bawah, menunjukkan produk dengan permintaan yang masih ada namun terbatas pada segmen pelanggan tertentu. Manajemen perlu memantau pergerakan cluster ini secara berkala, karena produk dalam kategori ini berisiko bergeser ke Cluster 1 (lambat terjual) apabila tidak mendapat perhatian, namun juga berpotensi naik ke Cluster 3 apabila diberikan dukungan promosi yang tepat. Posisi Cluster 4 yang berada di antara Cluster 1 (lambat terjual) dan Cluster 3 (potensial) menjadikan kelompok ini sebagai segmen yang paling membutuhkan pemantauan berkelanjutan (monitoring), karena arah pergerakannya di masa depan baik menurun ke kategori slow-moving maupun naik ke kategori potensial sangat dipengaruhi oleh keputusan manajerial yang diambil saat ini. Sebagai langkah praktis, perusahaan dapat menjadikan hasil segmentasi ini sebagai dasar pemantauan berkala, misalnya dengan



menjalankan kembali proses clustering secara periodik (misalnya setiap kuartal) untuk mendeteksi pergeseran posisi suatu produk antar cluster dari waktu ke waktu, sehingga keputusan strategis dapat diambil secara lebih responsif terhadap perubahan pola permintaan pasar.

3.6 Pembahasan

Hasil penelitian ini secara umum sejalan dengan temuan pada sejumlah penelitian sebelumnya, namun juga memperlihatkan beberapa perbedaan yang bersumber dari konteks data dan pendekatan metodologis yang digunakan.

Dibandingkan dengan penelitian Purnama et al. [1] yang menerapkan K-Means pada data penjualan pakaian, kedua penelitian sama-sama menghasilkan segmentasi yang dapat membedakan produk laris dari produk stagnan. Namun, penelitian [1] tidak menyertakan uji stabilitas maupun evaluasi DBI secara eksplisit, sehingga konsistensi hasil clusteringnya tidak dapat diverifikasi secara kuantitatif. Penelitian ini melengkapi gap tersebut melalui uji stabilitas lima iterasi dengan standar deviasi DBI sebesar 0,0034 yang membuktikan keandalan hasil.

Penelitian Hadi dan Diana [11] yang menerapkan K-Means pada data penjualan bahan bangunan memiliki kemiripan domain yang paling dekat dengan penelitian ini. Meskipun demikian, penelitian tersebut menggunakan variabel stok tunggal sebagai dasar clustering, sedangkan penelitian ini mengekstraksi tiga variabel multidimensi (FMR) dari data transaksi historis. Perbedaan pendekatan ini menghasilkan segmentasi yang lebih kaya secara informasi, karena mampu membedakan produk yang terjual dalam volume besar ke satu pelanggan dengan produk yang terjual dalam volume sedang tetapi kepada banyak pelanggan dua kondisi yang membutuhkan strategi manajemen berbeda namun tidak dapat dibedakan jika hanya mengandalkan satu variabel.

Penelitian Yuliana et al. [3] yang juga menerapkan K-Means pada konteks produk keramik menggunakan dua variabel, yaitu data pemilihan keramik dan pelanggan potensial. Penelitian ini memperluas cakupan tersebut dengan menambahkan dimensi Frequency dan Monetary di samping Reach, sehingga menghasilkan profil cluster yang lebih komprehensif. Nilai DBI sebesar 0,8954 yang diperoleh pada penelitian ini juga berada pada kategori baik (di bawah 1,0), yang mengindikasikan bahwa penambahan dimensi variabel tidak menurunkan kualitas separasi antar cluster.

Jika dibandingkan dengan penelitian yang mengadopsi kerangka RFM penuh seperti Butsianto dan Mayangwulan [9] serta Jupriyanto dan Nurlela [5], penelitian ini memilih untuk menggantikan dimensi Recency dengan Reach. Keputusan ini bukan merupakan penyederhanaan, melainkan adaptasi yang disengaja karena data transaksi yang tersedia tidak memiliki variasi temporal yang cukup untuk menghitung Recency secara bermakna. Pendekatan adaptasi serupa juga telah dilakukan oleh Wicaksana et al. [10] yang memodifikasi komponen RFM menyesuaikan ketersediaan data, sehingga kerangka FMR yang diusulkan dalam penelitian ini memiliki landasan metodologis yang dapat dipertanggungjawabkan.

Dari sisi ukuran data, penelitian ini memproses 6.328 transaksi yang menghasilkan 417 produk unik sebelum tahap outlier removal, lebih besar dibandingkan sebagian besar penelitian terapan K-Means pada konteks distribusi yang umumnya bekerja pada ratusan hingga dua ribu transaksi [4], [7]. Besarnya skala data ini memperkuat validitas generalisasi temuan yang dihasilkan, sekaligus menunjukkan bahwa pendekatan yang diusulkan dapat diterapkan pada volume data transaksi yang lebih besar tanpa kehilangan efisiensi komputasi, yang merupakan salah satu keunggulan yang melekat pada algoritma K-Means [7].

4. KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa implementasi algoritma K-Means yang divalidasi dengan Davies-Bouldin Index (DBI) terbukti efektif dalam mengelompokkan produk keramik berdasarkan pola historis penjualan menggunakan pendekatan rekayasa fitur Frequency, Monetary, dan Reach (FMR). Hasil Elbow Method menetapkan $k=4$ sebagai jumlah cluster optimal, dengan nilai DBI sebesar 0,8954 yang dikategorikan baik dan diperkuat melalui uji stabilitas pada 5 iterasi yang menunjukkan konsistensi tinggi (standar deviasi DBI sebesar 0,0034). Dari 417 produk yang dianalisis, sebanyak 36 produk dikategorikan sebagai outlier dan dikeluarkan dari proses utama, sehingga clustering dilakukan terhadap 381 produk yang menghasilkan empat segmen, yaitu Cluster 1 (190 produk) sebagai produk lambat terjual, Cluster 2 (34 produk) sebagai produk unggulan dengan rata-rata frekuensi transaksi tertinggi, Cluster 3 (93 produk) sebagai produk potensial, dan Cluster 4 (64 produk) sebagai produk dengan permintaan terbatas. Penelitian ini memberikan kontribusi praktis bagi CV. Makmur Bersama dalam menentukan prioritas restocking pada produk unggulan, strategi promosi pada produk potensial dan permintaan terbatas, serta evaluasi pengurangan order pada produk lambat terjual guna mengoptimalkan efisiensi modal kerja dan pengelolaan gudang. Penelitian selanjutnya disarankan untuk menambahkan variabel harga atau margin keuntungan apabila data tersedia, serta mempertimbangkan dimensi temporal (Recency) yang lebih akurat untuk memperkaya kerangka segmentasi yang digunakan.

REFERENCES

- [1] C. Purnama, W. Witanti, and P. Nurul Sabrina, "Klasterisasi Penjualan Pakaian untuk Meningkatkan Strategi Penjualan Barang Menggunakan K-Means," *Journal of Information Technology*, vol. 4, no. 1, pp. 35–38, Mar. 2022, doi: 10.47292/joint.v4i1.79.
- [2] D. Handoko, H. S. Tambunan, and J. T. Hardinata, "Analisis Penjualan Produk Paket Kuota Internet Dengan Metode K-Nearest Neighbor," *Jurasik (Jurnal Riset Sistem Informasi dan Teknik Informatika)*, vol. 6, no. 1, p. 111, Feb. 2021, doi: 10.30645/jurasik.v6i1.275.



- [3] Y. Yuliana, M. Richie, and H. Agung, "Implementasi Algoritma K-Means Untuk Pemilihan Keramik Dan Pelanggan Potensial Pada Cv. Jaya Tunggal Keramik," *Jurnal Algoritma, Logika dan Komputasi*, vol. 3, no. 2, Dec. 2020, doi: 10.30813/j-alu.v3i2.2157.
- [4] W. W. Kristianto, "Penerapan Data Mining Pada Penjualan Produk Menggunakan Metode K-Means Clustering (Studi Kasus Toko Sepatu Kakikaki)," *Jurnal Pendidikan Teknologi Informasi (JUKANTI)*, vol. 5, no. 2, pp. 90–98, Nov. 2022, doi: 10.37792/jukanti.v5i2.547.
- [5] J. Jupriyanto and S. Nurela, "Kerangka Pengambilan Keputusan Untuk Pemasaran Presisi Menggunakan Metode Rfm, Algoritma K-Means Dan Decision Tree," *Jurnal Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 227–234, Sep. 2019, doi: 10.33480/pilar.v15i2.618.
- [6] I. S. Mangku Negara, P. Purwono, and I. A. Ashari, "Analisa Cluster Data Transaksi Penjualan Minimarket Selama Pandemi Covid-19 dengan Algoritma K-means," *JOINTECS (Journal of Information Technology and Computer Science)*, vol. 6, no. 3, p. 153, Sep. 2021, doi: 10.31328/jointecs.v6i3.2693.
- [7] S. Aulia, "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, Jun. 2021, doi: 10.46576/djtechno.v1i1.964.
- [8] A. Muni, "Analisis Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Penjualan Sepeda Motor Studi Kasus PT. Alfa Scorpii," *JUTI UNISI*, vol. 4, no. 1, pp. 1–8, Jul. 2020, doi: 10.32520/juti.v4i1.1087.
- [9] S. Butsianto and N. T. Mayangwulan, "Penerapan Data Mining Untuk Prediksi Penjualan Mobil Menggunakan Metode K-Means Clustering," *Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)*, vol. 3, no. 3, pp. 187–201, Dec. 2020, doi: 10.32672/jnkti.v3i3.2428.
- [10] R. S. Wicaksana, D. Heksaputra, T. A. Syah, and F. F. Nur'aini, "Pendekatan K-Means Clustering Metode Elbow Pada Analisis Motivasi Pengunjung Festival Halal JHF#2," *Jurnal Ilmiah Ekonomi Islam*, vol. 9, no. 3, p. 4162, Nov. 2023, doi: 10.29040/jiei.v9i3.10591.
- [11] F. Hadi and Y. Diana, "Pengklusteran Penjualan Bahan Bangunan Menggunakan Algoritma K-Means," *JOISIE (Journal Of Information Systems And Informatics Engineering)*, vol. 4, no. 1, p. 22, Jun. 2020, doi: 10.35145/joisie.v4i1.629.
- [12] F. Indriyani and E. Irfiani, "Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means," *JUITA : Jurnal Informatika*, vol. 7, no. 2, p. 109, Nov. 2019, doi: 10.30595/juita.v7i2.5529.
- [13] F. Nurdiansyah and I. Akbar, "Implementasi Algoritma K-Means untuk Menentukan Persediaan Barang pada Poultry Shop," *Jurnal Teknologi dan Manajemen Informatika*, vol. 7, no. 2, pp. 86–94, Dec. 2021, doi: 10.26905/jtmi.v7i2.6377.
- [14] A. Wahid, A. Nazir, S. K. Gusti, - Yusra, and F. Syafria, "Pengelompokan Keberhasilan Produksi Peternak Ayam Broiler di Riau Berdasarkan Index Performance Menggunakan K-Means," *Techno.Com*, vol. 22, no. 1, pp. 176–185, Feb. 2023, doi: 10.33633/tc.v22i1.7282.
- [15] A. Alvin Anzaz Islami, E. Haerani, Novriyanto, and A. Nazir, "Pengelompokan pembagian zakat dengan menggunakan metode clustering k-means," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 4, no. 1, pp. 154–163, Apr. 2023, doi: 10.37859/coscitech.v4i1.4804.
- [16] A. Ahyuna, M. Lasena, R. Aminuddin, A. Ardimansyah, and Z. Azhar, "Pembentukan Pola Peminjaman Buku Pada Perpustakaan Dengan Menerapkan Metode CART dan Normalisasi Z-Score," *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 1, pp. 314–324, Jun. 2024, doi: 10.47065/bits.v6i1.5238.
- [17] I. D. Iskandar and I. Zufria, "Clustering Pasien Rawat Inap Di RS USU Menggunakan Algoritma K-Means," *Journal of Computer Science and Informatics Engineering*, pp. 54–63, May 2024, doi: 10.55537/cosie.v3i2.768.
- [18] A. W. Aranski, S. Astiti, R. A. Putra, and D. Darmansah, "Pengaplikasian Data Mining Dalam Mengelompokan Data Penerima Bantuan Subsidi Rumah dengan Menggunakan Metode K-Means Clustering," *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 1, pp. 480–489, Jun. 2024, doi: 10.47065/bits.v6i1.5366.
- [19] Khoirul Muhammad Habib, Miftahus Sholihin, and Agus Setia Budi, "Penerapan Algoritma K Means Clustering Hasil Tangkap Ikan Di Pelabuhan Brondong," *Jurnal Publikasi Ilmu Komputer dan Multimedia*, vol. 5, no. 1, pp. 117–126, Jan. 2026, doi: 10.55606/jupikom.v5i1.5352.
- [20] Jiawei. Han, Micheline. Kamber, and Jian. Pei, *Data mining : concepts and techniques*. Elsevier/Morgan Kaufmann, 2012.