VOL. 03. NO. 06 NOVEMBER 2026

JOURNAL OF COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE (JCSAI)

ANALISIS SEGMENTASI PRODUK STOK TOMORO COFFEE MENGGUNAKAN ALGORITMA K-MEANS

Muhammad Rizki[1], Nining Rahaningsih[2], Irfan Ali[3], Ade Rizki Rinaldi[4]

Program Studi Teknik Informatika[1]
Program Studi Komputerisasi Akuntansi[2]
Program Studi Rekayasa Perangkat Lunak[34]

STMIK IKMI Cirebon
https://ikmi.ac.id/page/18/?lang=de
mhmmdrzk6@gmail.com

(*) Corresponding Author : mhmmdrzk6@gmail.com
Published : 20 Juli 2026

Abstract—*The rapid expansion of Tomoro Coffee requires more efficient inventory management, as variations in product turnover often lead to overstock or stockout conditions. These imbalances increase operational costs, risk of product spoilage, and reduced customer satisfaction. Therefore, an analytical approach is needed to group stock items based on sales behavior and inventory characteristics. This study aims to apply the K-Means Clustering algorithm to segment stock products at Tomoro Coffee Ciremai. The methodology follows the Knowledge Discovery in Databases (KDD) framework, including data selection, cleaning, transformation, determination of the optimal number of clusters using the Elbow Method and Silhouette Score, and evaluation of clustering performance. The dataset consists of historical stock and sales records collected during the observation period. The results indicate that the optimal number of clusters is K = 3, forming three primary product segments: (1) Fast Moving, characterized by high sales and rapid turnover; (2) Slow Moving, with moderate sales movement; and (3) Overstock/Dead Stock, referring to low-selling items with excessive stock levels. This segmentation provides actionable insights for inventory control, including restocking priorities, purchasing adjustments, and minimizing waste. The study concludes that K-Means is effective for stock product segmentation and supports more accurate and efficient inventory management. Future research is recommended to incorporate additional variables such as profit margin, seasonal demand patterns, and predictive forecasting methods.*

Keywords: *K-Means, clustering, inventory management, product segmentation, Tomoro Coffee.*

Abstrak—Pertumbuhan pesat Tomoro Coffee menuntut pengelolaan persediaan yang lebih efisien, terutama karena variasi perputaran produk sering menyebabkan penumpukan stok (*overstock*) maupun kekurangan stok (*stockout*) pada cabang tertentu. Ketidakseimbangan ini berdampak pada biaya operasional, risiko kerugian bahan baku, serta ketidakpuasan pelanggan. Oleh karena itu, diperlukan pendekatan analitik yang mampu mengelompokkan produk berdasarkan pola penjualan dan karakteristik inventori. Penelitian ini bertujuan menerapkan algoritma K-Means Clustering untuk melakukan segmentasi produk stok di Tomoro Coffee Ciremai. Metodologi yang digunakan mengadopsi tahapan Knowledge Discovery in Databases (KDD), meliputi seleksi data, pembersihan data, transformasi data, penentuan jumlah klaster optimal menggunakan Elbow Method dan Silhouette Score, serta evaluasi kesesuaian hasil. Dataset terdiri atas data stok dan penjualan historis selama periode observasi. Hasil penelitian menunjukkan bahwa nilai optimal klaster adalah K = 3, yang menghasilkan tiga segmen produk utama: (1) Fast Moving, dengan tingkat penjualan tinggi dan rotasi cepat; (2) Slow Moving, dengan pergerakan moderat; dan (3) Overstock/Dead Stock, yaitu produk dengan penjualan rendah namun stok tinggi. Segmentasi ini memberikan dasar strategis bagi pengambilan keputusan, termasuk pengaturan restock, kontrol pembelian, dan pengurangan risiko pemborosan. Penelitian menyimpulkan bahwa algoritma K-Means efektif dalam melakukan segmentasi produk stok dan dapat mendukung manajemen inventori yang lebih akurat dan efisien. Studi lanjutan disarankan untuk menambahkan variabel margin keuntungan, tren musiman, serta integrasi metode peramalan.

Kata Kunci : *K-Means*, *clustering*, persediaan, segmentasi produk, Tomoro *Coffee*.

INTRODUCTION

Industri kedai kopi di Indonesia mengalami pertumbuhan yang fenomenal. Pertumbuhan ini, bagaimanapun, menciptakan lanskap pasar yang sangat dinamis dan kompetitif. Sejumlah penelitian terbaru menyoroti bahwa industri kedai kopi di kota-kota besar Indonesia kini menghadapi "persaingan yang ketat karena kejenuhan pasar" (market saturation) [1]. Dalam lingkungan yang sangat kompetitif ini, bisnis tidak bisa lagi hanya mengandalkan ekspansi, tetapi harus fokus pada strategi bisnis yang berkelanjutan [2] dan mencari keunggulan kompetitif melalui efisiensi operasional.

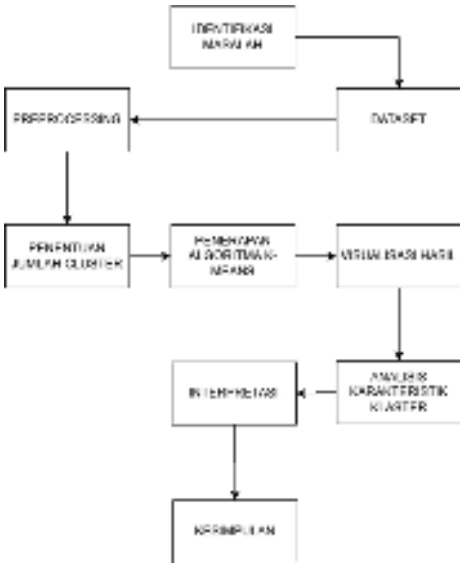
Ketika persaingan di lini depan (pemasaran dan pengalaman pelanggan) semakin sengit, keunggulan kompetitif harus dicari di lini belakang, yaitu melalui efisiensi operasional. Dalam industri ritel *Food and Beverage* (F&B), "manajemen inventori yang efektif sangat krusial" [3]. Manajemen persediaan merupakan "faktor yang sangat penting untuk diterapkan pada perusahaan retail," karena persediaan barang dagangan, baik bahan baku maupun produk jadi, merupakan "komponen utama dalam kegiatan operasional" [4]. Kegagalan dalam mengelola persediaan secara efektif akan memanifestasikan dirinya dalam dua masalah utama yang secara langsung menggerus profitabilitas. Tujuan utama dari optimalisasi persediaan adalah "guna menghindari terjadinya *stock out* (kekurangan stok) maupun *overstock* (kelebihan stok)" [4].

Dalam konteks kedai kopi, *overstock* sangat berbahaya. Ini tidak hanya mengikat modal kerja dalam inventaris yang tidak bergerak, tetapi juga meningkatkan biaya penyimpanan dan risiko kerugian total akibat bahan baku segar (seperti susu dan *pastry*) yang melewati masa kedaluwarsa. Stockout (Kekurangan Stok) merujuk pada kondisi di mana stok barang habis saat dibutuhkan. Dalam pasar yang jenuh [1], dampak *stockout* sangat fatal karena pelanggan dapat dengan mudah beralih ke kedai kopi pesaing. Mengatasi dilema ini menuntut pendekatan yang lebih cerdas dan berbasis data (*data-driven*) (Kahar et al., 2024). *Data mining*, sebagai sebuah metode untuk menemukan pola tersembunyi dari kumpulan data historis yang besar, menawarkan solusi yang menjanjikan. Salah satu teknik *data mining* yang paling banyak digunakan adalah *clustering*, khususnya algoritma *K-Means*.

Penelitian dari empat tahun terakhir telah menunjukkan efektivitas *K-Means* sebagai "salah satu teknik *data mining* yang digunakan untuk merancang strategi persediaan barang yang efektif menggunakan data transaksi penjualan" [6]. Penelitian sebelumnya telah berhasil menerapkan *K-Means* untuk segmentasi pelanggan sebagai strategi "untuk menghindari *overstock*" di *online shop* [7] dan untuk "menentukan persediaan stok barang" di *mini market* [6]. Bahkan dalam industri kedai kopi, *K-Means* telah terbukti berhasil digunakan untuk segmentasi pelanggan guna menyusun strategi pemasaran (Kahar et al., 2024). Meskipun K-Means telah terbukti berhasil di sektor ritel dan untuk segmentasi *pelanggan* kedai kopi, penerapan spesifiknya untuk segmentasi *produk* sebagai landasan strategi *inventori* di kedai kopi Indonesia yang memiliki karakteristik unik seperti bahan baku segar (*perishable*) dan persaingan jenuh masih belum banyak dieksplorasi. Penelitian ini diajukan untuk mengisi celah tersebut dengan menerapkan algoritma *K-Means clustering* untuk segmentasi produk berdasarkan data historis penjualan, guna merumuskan strategi manajemen persediaan yang optimal.

MATERIALS AND METHODS

Dalam penelitian berbasis data, diperlukan alur kerja yang sistematis agar proses analisis dapat berjalan secara terstruktur dan menghasilkan temuan yang valid. Salah satu pendekatan yang umum digunakan dalam pengelompokan data adalah algoritma K-Means, yang mampu mengidentifikasi pola tersembunyi dalam dataset. Untuk memastikan penerapan metode ini berjalan optimal, setiap tahapan harus dirancang dengan jelas, mulai dari identifikasi masalah hingga penarikan kesimpulan. Diagram alur berikut menggambarkan langkah-langkah utama dalam proses clustering menggunakan K-Means, yang tidak hanya berfokus pada proses komputasi, tetapi juga pada pemahaman hasil analisis secara menyeluruh dan kontekstual.



Gambar 1. Alur Penelitian

Gambar tersebut menunjukkan alur proses penelitian atau analisis data menggunakan metode K-Means clustering yang disusun secara sistematis. Tahap pertama adalah *identifikasi masalah*, yaitu proses menentukan tujuan penelitian dan permasalahan utama yang ingin diselesaikan melalui analisis data. Tahapan ini sangat penting karena menjadi dasar dalam menentukan pendekatan dan metode yang digunakan.

Selanjutnya terdapat tahap *dataset*, yaitu pengumpulan data yang akan dianalisis. Data ini bisa berasal dari berbagai sumber dan harus relevan dengan permasalahan yang telah diidentifikasi. Setelah data diperoleh, dilakukan *preprocessing*, yang meliputi pembersihan data, normalisasi, serta transformasi agar data siap digunakan dalam proses clustering.

Tahap berikutnya adalah *penentuan jumlah cluster*, yang bertujuan untuk menentukan jumlah kelompok terbaik dalam data. Metode seperti Elbow atau Silhouette biasanya digunakan pada tahap ini. Setelah jumlah cluster ditentukan, dilakukan *penerapan algoritma K-Means* untuk mengelompokkan data berdasarkan kemiripan karakteristik.

Hasil dari proses clustering kemudian ditampilkan dalam tahap *visualisasi hasil*, yang membantu dalam memahami distribusi dan pola data secara lebih intuitif melalui grafik atau plot. Selanjutnya dilakukan *analisis karakteristik klaster*, yaitu mengidentifikasi ciri khas dari setiap kelompok yang terbentuk.

Tahap *interpretasi* bertujuan untuk memberikan makna terhadap hasil analisis sehingga dapat digunakan dalam pengambilan keputusan. Terakhir, seluruh proses dirangkum dalam tahap *kesimpulan*, yang berisi temuan utama dan implikasi dari hasil penelitian. Keseluruhan alur ini menunjukkan proses yang terstruktur dari awal hingga akhir dalam analisis clustering.

RESULTS AND DISCUSSION

Tahap ini merupakan inti komputasi penelitian, di mana algoritma K-Means diaplikasikan untuk mempartisi 45 SKU produk ke dalam kelompok-kelompok (*clusters*) yang memiliki karakteristik internal yang homogen namun heterogen antar kelompok.

Salah satu tantangan fundamental dalam K-Means adalah penentuan jumlah cluster (k) yang harus didefinisikan di awal (apriori). Penelitian ini menggunakan pendekatan hibrida yang menggabungkan metode statistik Elbow Method dan pertimbangan domain bisnis (*Domain Knowledge*).

Analisis Statistik: Grafik *Within-Cluster Sum of Squares* (WCSS) diplot terhadap berbagai nilai k (1 hingga 10). WCSS mengukur varians total dalam cluster. Grafik menunjukkan penurunan nilai WCSS yang tajam dari *k=1* ke *k=2*, dan melandai secara signifikan setelah *k=3*. Titik "siku" ini mengindikasikan bahwa penambahan cluster di atas 3 tidak memberikan penurunan varians yang signifikan (*diminishing returns*).

Validasi Bisnis: Angka *k=3* sangat selaras dengan kerangka teori manajemen inventaris klasik, yaitu analisis FSN (*Fast, Slow, Non-moving*). Membagi produk menjadi tiga kategori memungkinkan manajemen untuk menerapkan tiga strategi rantai pasok yang berbeda (Agresif, Moderat, Konservatif).

Berdasarkan konvergensi bukti statistik dan teoritis tersebut, ditetapkan nilai *k=3*. Algoritma dijalankan dengan parameter inisialisasi centroid acak (*random initialization*). Algoritma bekerja secara iteratif melalui dua langkah utama:

1. Assignment Step: Setiap titik data (produk) dialokasikan ke cluster dengan centroid terdekat berdasarkan jarak Euclidean.

$$d(p,q) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2}$$

2. Update Step: Posisi *centroid* diperbarui dengan menghitung nilai rata-rata (*mean*) dari seluruh atribut anggota cluster yang baru terbentuk.

Proses ini berulang hingga mencapai kondisi konvergensi, yaitu ketika posisi *centroid*

tidak lagi bergeser atau pergeserannya berada di bawah ambang batas toleransi ($< 10^{-4}$). Dalam eksperimen ini, algoritma mencapai konvergensi stabil pada Iterasi ke-5. Kecepatan konvergensi ini menunjukkan bahwa struktur data memiliki pola pengelompokan yang alami dan cukup terpisah.

Hasil akhir dari proses mining adalah terbentuknya tiga cluster dengan karakteristik pusat (*centroid*) sebagai berikut:

	K	Silhouette Score
0	2	0.607186
1	3	0.572441
2	4	0.514332
3	5	0.484899
4	6	0.476035
5	7	0.431641
6	8	0.463319
7	9	0.485336
8	10	0.493993

Gambar 2 Hasil Akhir Clustering

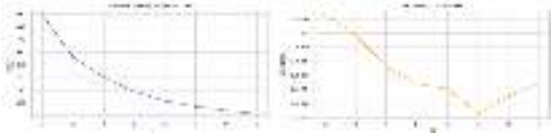
Berdasarkan Gambar 2 menunjukkan pengujian validitas cluster menggunakan metode *Silhouette Score* pada rentang K=2 hingga K=10, ditemukan bahwa jumlah cluster optimal untuk penelitian ini adalah K=2 dengan nilai skor tertinggi sebesar 0,607186. Secara teoretis, nilai *Silhouette Score* yang melebihi angka 0,5 menunjukkan bahwa struktur pengelompokan yang terbentuk memiliki pemisahan yang kuat (*strong structure*) dan tingkat kohesi internal yang tinggi. Meskipun nilai pada K=3 masih tergolong baik (0,572441), terjadi penurunan skor yang cukup konsisten pada penambahan jumlah cluster berikutnya (K=4 hingga K=10), yang mengindikasikan bahwa peningkatan jumlah kelompok justru akan memperburuk batas antar-cluster dan memicu terjadinya tumpang tindih (*overlapping*) data. Oleh karena itu, penelitian ini menetapkan K=2 sebagai konfigurasi terbaik untuk menjamin bahwa setiap objek diklasifikasikan ke dalam kelompok yang paling homogen dan memiliki perbedaan karakteristik yang signifikan antar kelompoknya.

Interpretasi dan Evaluasi

Validasi model clustering dilakukan secara kuantitatif menggunakan metrik internal untuk memastikan objektivitas hasil, menghindari subjektivitas visual semata.

Gambar 3 Kode Program Evaluasi

Gambar 3 Kode tersebut melakukan proses evaluasi pemilihan jumlah cluster (*k*) pada algoritma K-Means dengan menghitung beberapa metrik performa untuk setiap nilai *k* dalam rentang 2 hingga 9. Untuk setiap iterasi, model K-Means dibangun menggunakan data yang telah melalui proses standarisasi, kemudian menghasilkan label cluster yang digunakan untuk menghitung empat indikator utama, yaitu Inertia (digunakan pada Elbow Method, semakin kecil semakin baik karena menunjukkan cluster lebih kompak), Silhouette Score (mengukur kualitas pemisahan dan kekompakan cluster, semakin mendekati 1 semakin baik), Davies-Bouldin Index atau DBI (menilai jarak antar cluster, semakin kecil semakin baik), dan Calinski-Harabasz Index atau CHI (menilai pemisahan antar cluster berdasarkan variasi internal dan eksternal, semakin besar semakin baik). Seluruh hasil perhitungan ini disimpan ke dalam sebuah DataFrame untuk kemudian divisualisasikan dalam dua grafik: grafik Elbow, yang menunjukkan perubahan inertia untuk membantu menentukan titik siku sebagai kandidat jumlah cluster optimal, serta grafik Silhouette, yang menggambarkan kualitas pemisahan cluster untuk setiap nilai *k*. Dengan demikian, tahapan ini membantu menentukan nilai *k* terbaik secara lebih objektif berdasarkan kombinasi beberapa metrik evaluasi.



Gambar 4 Visualisasi Elbhow Method vs Silhoutte Score

Gambar 4.11 Grafik *Elbow Method* pada bagian kiri menunjukkan pola penurunan inertia (WCSS) yang sangat signifikan dari K=2 hingga K=4, di mana nilai inertia turun dari sekitar 89 menjadi 40. Setelah titik tersebut, penurunan inertia mulai melambat, terutama pada K=5 hingga K=9, yang hanya menunjukkan pengurangan kecil dari 39 menuju 11. Pola ini menegaskan bahwa titik siku (*elbow point*) berada di sekitar K=4 atau K=5, karena pada titik tersebut penambahan jumlah klaster tidak lagi memberikan pengurangan WCSS yang berarti. Dengan kata lain, penggunaan jumlah klaster lebih dari lima tidak memberikan peningkatan kualitas klaster yang signifikan, sehingga nilai K=4-5 dapat dipertimbangkan sebagai pilihan optimal dari sudut pandang *compactness* klaster.

Sementara itu, grafik *Silhouette Score* pada bagian kanan menggambarkan tingkat pemisahan dan konsistensi setiap klaster. Nilai siluet tertinggi dicapai pada K=2 dengan skor sekitar 0.61, yang menunjukkan klaster yang sangat terpisah dan terdefinisi dengan baik. Namun, penggunaan K=2 sering kali dianggap terlalu sederhana untuk kebutuhan segmentasi produk. Pada K=3, nilai siluet masih relatif tinggi (0.57), menunjukkan struktur klaster yang cukup baik. Setelah itu, nilai siluet menurun pada K=4 hingga K=7, dan mencapai titik terendah pada K=7 dengan nilai sekitar 0.43, menandakan bahwa pada jumlah klaster tersebut, pemisahan antar klaster kurang jelas. Nilai siluet kemudian mulai meningkat kembali pada K=8 dan K=9, namun tidak mencapai kualitas sebagus K=2 atau K=3.

Jika kedua grafik tersebut dievaluasi secara bersamaan, dapat disimpulkan bahwa meskipun K=2 memberikan kualitas pemisahan terbaik, konfigurasi tersebut tidak cukup representatif untuk segmentasi produk dalam konteks bisnis. Jumlah klaster yang lebih seimbang menurut *Elbow Method* dan masih memiliki siluet yang cukup baik adalah K=3 atau K=4. Oleh karena itu, K=3–4 menjadi pilihan jumlah klaster yang paling ideal karena menawarkan keseimbangan antara *compactness* (struktur klaster yang padat) dan *separation* (pemisahan antar klaster yang jelas), sekaligus memberikan segmentasi yang lebih bermakna untuk analisis stok produk Tomoro Coffee.

CONCLUSION

Berdasarkan seluruh tahapan penelitian yang telah dilakukan, mulai dari pengumpulan data, pra-pemrosesan, hingga implementasi dan evaluasi algoritma *K-Means Clustering* pada data stok produk Tomoro Coffee gerai Ciremai Raya, dapat ditarik kesimpulan sebagai berikut:

Penerapan algoritma K-Means terbukti efektif dalam mengelompokkan produk stok berdasarkan pola penjualan dan ketersediaan barang. Melalui metode evaluasi Elbow Method dan validasi Silhouette Score, ditemukan bahwa jumlah klaster optimal adalah 3 (*K=3*). Algoritma berhasil mencapai konvergensi yang stabil pada iterasi ke-5, menunjukkan bahwa data stok Tomoro Coffee memiliki struktur pengelompokan alami yang kuat yang dapat dipisahkan secara matematis untuk mendukung pengambilan keputusan.

Penelitian ini berhasil mengidentifikasi tiga segmen produk dengan karakteristik yang distingtif, yang memetakan kondisi operasional toko secara akurat:

Klaster 2 (*Fast Moving*): Mencakup sekitar 13% produk (6 SKU) seperti "Tomoro Aren Latte". Karakteristik utamanya adalah volume penjualan sangat tinggi dan perputaran cepat. Tingginya stok pada klaster ini bukan inefisiensi, melainkan kebutuhan *buffer* untuk mencegah *stockout*.[3]

Klaster 1 (*Slow Moving*): Mencakup sekitar 31% produk (14 SKU) dengan penjualan moderat, seperti varian *matcha* atau *oat latte*. Risiko utama pada klaster ini adalah masa kadaluarsa bahan baku karena perputaran yang tidak secepat produk utama.

Klaster 0 (*Dead Stock/Low Performance*): Merupakan kelompok terbesar (55% SKU) namun dengan kontribusi penjualan terendah. Adanya stok yang mengendap pada kelompok ini mengindikasikan inefisiensi alokasi modal kerja dan pemborosan ruang penyimpanan.

Kualitas hasil klasterisasi sangat dipengaruhi oleh tahap normalisasi data. Penggunaan *Min-Max Scaling* terbukti krusial untuk menyeimbangkan variabel Total Sales (yang memiliki rentang ribuan unit) dengan Average Stock, sehingga produk best-seller yang merupakan outlier positif tidak mendistorsi perhitungan jarak Euclidean dan dapat dikelompokkan dengan tepat.

REFERENCE

[1] A. L. Dwiputri, Syahputra, and M. Pradana, “From Sight to Sip: The Role of Sensory Marketing in Local Coffee Shops,” *International Journal of Applied Economics, Finance and Accounting*, vol. 2, no. 5, 2024, doi: 10.59890/ijaeam.v2i5.2463.

[2] F. M. A. F. Mubarok and Idris, “Pengembangan Rencana Bisnis Berkelanjutan Ditengah Pertumbuhan Industri Kopi: Studi Empiris Kedai Kopi Tekari,” *Jurnal Dinamika Bisnis dan Kewirausahaan*, vol. 1, no. 2, 2024.

[3] N. A. Gicheru and N. P. Ngugi, "E-Inventory management and performance of food and beverages manufacturing firms in Kenya," *The International Journal of Business & Management*, 2023.

[4] F. Ardianto and D. Wardana, "Optimalisasi Manajemen Persediaan dengan EOQ, ROP, dan Safety Stock (Studi Kasus pada Perusahaan 'ASK')," *RISTANSI: Riset Akuntansi*, vol. 6, no. 1, pp. 1–15, 2025.

[5] A. Kahar and (Penulis Lain tidak disebutkan), "Penerapan Data-Driven Marketing Menggunakan K-Means Clustering dan Decision Tree (Judul diinferensi dari konteks)," *IJDB*, 2024.

[6] H. Prastiwi, J. Pricilia, and E. Rasywir, "Implementasi Data Mining Untuk Menentuksn Persediaan Stok Barang Di Mini Market Menggunakan Metode K-Means Clustering," *Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM)*, vol. 2, no. 1, 2022, doi: 10.33998/jakakom.2022.2.1.34.

[7] A. Dewabharata, "Customer Segmentation Using the K-Means Clustering as a Strategy to Avoid Overstock in Online Shop Inventory," in *Proceedings of the 1st International Conference on Contemporary Risk Studies, ICONIC-RS 2022*, 2022. doi: 10.4108/eai.31-3-2022.2320688.