Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871

Vol. 6, No. 5, October 2025, Page. 3280-3392
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5283

Implementation of Clustering on Packaged Coffee Sales Data for Simulating Goods Entry in Sole Proprietorship Businesses

Mochammad Agri Triansyah*[1], Bangun Wijayanto[2], Ayu Anjar Paramestuti[3]

[1,2,3]Informatics, Universitas Jenderal Soedirman, Indonesia

Email: [1]mochammad.agri@unsoed.ac.id

Received : Aug 19, 2025; Revised : Sep 4, 2025; Accepted : Sep 5, 2025; Published : Oct 16, 2025

Abstract

In retail businesses operating under the sole proprietorship structure, decision-making regarding partnerships with beverage distributors—especially those offering packaged coffee—remains a challenge. Store owners often face uncertainty about the profitability of accepting product offerings, which can lead to suboptimal inventory decisions. This study addresses that issue by simulating goods entry scenarios and applying clustering techniques to historical packaged coffee sales data, enabling data-driven insights into product performance and distributor value. Studies focusing on clustering within retail include segmenting customer behavior and stock management strategies, yet many lacked specific application to single owner businesses and product-centric simulations. This research is novel in its contextual focus on packaged coffee distribution within sole proprietorship environments, integrating real sales metrics and clustering algorithms to empower store owners with actionable evaluation tools. Results demonstrate that clustering reveals patterns of profitable product categories and distributor consistency, offering scalable insights for micro-retail optimization. The findings provide a framework that differs from prior studies by emphasizing the intersection between small business dynamics and algorithmic decision support.

Keywords: *Clustering, Distributor Evaluation, Goods Simulation, Packaged Coffee Sales, Retail Analytics, Sole Proprietorship*

1. INTRODUCTION

Dalam era digitalisasi dan kompetisi pasar yang semakin dinamis, pelaku usaha mikro seperti pemilik toko kopi kemasan berbasis sole proprietorship menghadapi tantangan signifikan dalam pengambilan keputusan terkait kemitraan dengan distributor minuman, khususnya kopi kemasan. Salah satu masalah utama adalah kurangnya informasi yang memadai dari distributor mengenai performa penjualan produk yang ditawarkan. Banyak distributor tidak menyediakan data terperinci tentang tren penjualan, permintaan pasar, atau potensi profitabilitas produk tertentu, sehingga pemilik usaha sering kali harus mengandalkan intuisi atau perkiraan yang tidak akurat. Hal ini menyebabkan keputusan penerimaan barang yang suboptimal, seperti overstock produk dengan permintaan rendah atau kehilangan peluang untuk memasok produk yang berpotensi laris. Ketidakpastian ini diperparah oleh keterbatasan sumber daya pada usaha mikro, di mana kesalahan dalam pengelolaan inventaris dapat berdampak langsung pada keberlanjutan bisnis.

Untuk mengatasi tantangan tersebut, pendekatan berbasis data seperti algoritma clustering, khususnya K-Means, menjadi solusi yang menjanjikan untuk mengelompokkan produk berdasarkan karakteristik penjualan seperti frekuensi, volume, dan margin keuntungan. Berbagai studi telah menunjukkan bahwa K-Means mampu memberikan segmentasi yang bermanfaat dalam pengambilan keputusan bisnis. Misalnya, Agussalim [3] dan Hossain et al. [15] mengaplikasikan K-Means untuk mengelompokkan produk dan pelanggan berdasarkan perilaku pembelian, yang membantu

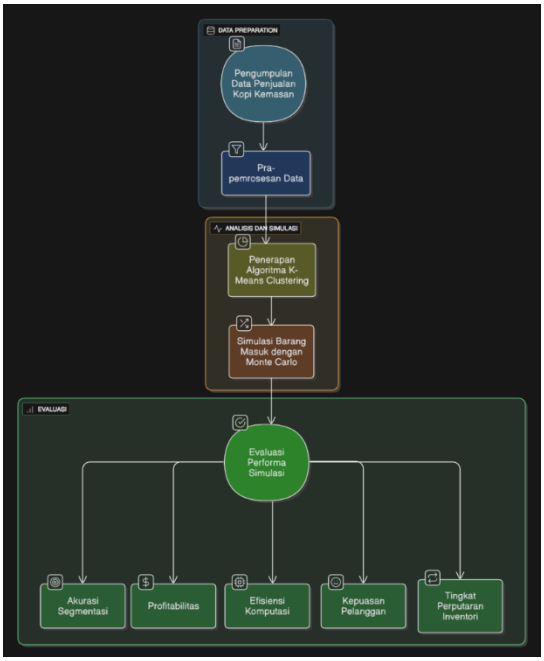
Vol. 6, No. 5, October 2025, Page. 3280-3392
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5283

meningkatkan efisiensi strategi pemasaran. Sementara itu, Singh dan Jain [18] menekankan pentingnya strategi ritel berbasis data untuk meningkatkan efisiensi operasional, meskipun penelitian mereka lebih berfokus pada ritel skala besar. Penelitian oleh Zahid et al. [16] menunjukkan bahwa kombinasi clustering dan association rules dapat mengoptimalkan strategi penjualan, namun penerapannya pada usaha mikro masih terbatas.

Dalam konteks produk kopi, pendekatan clustering telah digunakan untuk segmentasi kualitas dan preferensi konsumen. Thai-Ths [5] mengembangkan model segmentasi kualitas kopi berbasis data sensorik dan penjualan, sedangkan Rifai et al. [19] menerapkan clustering untuk mengelompokkan produk kopi berdasarkan pola pembelian lokal. Penelitian lokal lainnya, seperti Kusuma et al. [20] dan Prasetyo et al. [21], menunjukkan bahwa clustering dapat digunakan untuk mengevaluasi penawaran distributor dan mengoptimalkan stok produk UMKM. Namun, studi-studi tersebut jarang membahas secara spesifik tantangan yang dihadapi oleh usaha sole proprietorship akibat kurangnya transparansi data dari distributor. Selain itu, pendekatan clustering juga digunakan dalam pengembangan sistem pendukung keputusan, seperti yang ditunjukkan oleh Hidayat et al. [25] dan Ramadhan et al. [23], yang mengintegrasikan analisis data penjualan dengan rekomendasi strategis. Penelitian oleh Sari et al. [22] dan Putri et al. [26] memperkuat temuan bahwa clustering dapat membantu dalam pemetaan produk makanan dan minuman, termasuk kopi kemasan.

Dengan mempertimbangkan tantangan distributor yang tidak memberikan informasi penjualan yang memadai, penelitian ini bertujuan untuk menerapkan algoritma clustering pada data penjualan kopi kemasan dari usaha milik perorangan untuk mengidentifikasi pola penjualan dan memberikan rekomendasi strategis berbasis data. Dengan memanfaatkan data penjualan historis, penelitian ini mengisi kesenjangan informasi yang sering kali tidak disediakan oleh distributor, sehingga memungkinkan pemilik usaha untuk membuat keputusan yang lebih tepat mengenai penerimaan barang. Penelitian ini juga berkontribusi pada pengembangan sistem pendukung keputusan untuk UMKM dan memperkuat literatur tentang penerapan data mining dalam sektor ritel mikro, khususnya dalam konteks kemitraan dengan distributor.

2. METHOD



Gambar 1. Alur Penelitian untuk Clustering dan Simulasi Barang Masuk

Penelitian ini menggunakan pendekatan kuantitatif eksploratif dengan metode unsupervised learning, yaitu algoritma K-Means Clustering, untuk mengelompokkan data penjualan kopi kemasan berdasarkan performa produk. Selain itu, simulasi Monte Carlo digunakan untuk mengevaluasi strategi penerimaan barang dari distributor. Tujuan utama dari metode ini adalah menghasilkan segmentasi produk yang dapat digunakan dalam simulasi barang masuk, sehingga pemilik usaha sole proprietorship dapat mengambil keputusan berbasis data. Alur penelitian dirancang secara sistematis untuk memastikan bahwa setiap tahapan mendukung tujuan penelitian, sebagaimana divisualisasikan pada Gambar 1.

Diagram alur pada gambar 2 menunjukkan lima tahapan utama:
(1) pengumpulan data penjualan kopi kemasan,
(2) pra-pemrosesan data,
(3) penerapan algoritma K-Means Clustering,
(4) simulasi barang masuk menggunakan metode Monte Carlo,
(5) evaluasi performa simulasi berdasarkan metrik seperti akurasi segmentasi dan profitabilitas.

2.1. Desain Penelitian

1. Desain penelitian ini mengikuti tahapan sistematis yang divisualisasikan pada Gambar 2, yang terdiri dari:
2. Pengumpulan data penjualan kopi kemasan: Data dikumpulkan dari toko ritel simulatif selama periode enam bulan untuk mencerminkan kondisi nyata usaha mikro.
3. Pra-pemrosesan data: Data dibersihkan dan dinormalisasi untuk memastikan konsistensi dan akurasi analisis, menghilangkan anomali dan nilai yang hilang.
4. Penerapan algoritma clustering: Algoritma K-Means diterapkan untuk mengelompokkan produk berdasarkan fitur penjualan seperti volume, margin, dan frekuensi restock.
5. Simulasi barang masuk berdasarkan hasil clustering: Simulasi dilakukan menggunakan pendekatan Monte Carlo untuk mengevaluasi strategi penerimaan barang dari distributor berdasarkan performa produk.
6. Evaluasi performa simulasi: Hasil simulasi diukur menggunakan metrik seperti akurasi segmentasi, efisiensi rotasi stok, dan profitabilitas.

Pendekatan ini sejalan dengan studi oleh Manarung et al. [1], Zahid et al. [16], dan Singh & Jain [18], yang menunjukkan efektivitas clustering dalam pengambilan keputusan ritel.

2.2. Pengumpulan Data

Pengumpulan data merupakan langkah awal yang kritis dalam penelitian ini untuk memastikan bahwa analisis berbasis data dapat memberikan wawasan yang akurat dan relevan bagi pengambilan keputusan di usaha sole proprietorship. Data penjualan kopi kemasan dikumpulkan dari toko ritel simulatif yang dirancang untuk mencerminkan operasi nyata toko mikro di Indonesia, khususnya dalam konteks kemitraan dengan distributor minuman. Periode pengumpulan data berlangsung selama enam bulan (Januari–Juni 2025), dipilih untuk menangkap variasi musiman dalam penjualan kopi kemasan, seperti peningkatan permintaan pada bulan-bulan tertentu akibat promosi atau perubahan preferensi konsumen.

Data dikumpulkan menggunakan sistem point-of-sale (POS) simulatif yang mencatat setiap transaksi penjualan. Dataset terdiri dari sekitar 10.000 entri transaksi dari 50 produk kopi kemasan yang berbeda, melibatkan lima distributor utama. Fitur utama yang dikumpulkan meliputi:

- **Nama Produk**: Identifikasi merek dan jenis kopi kemasan (misalnya, Kopi ABC 200g, Kapal Api 500g).

- **Volume Penjualan**: Jumlah botol/kaleng terjual per bulan, dengan rentang 50–1000 botol/kaleng per produk, mencerminkan variasi permintaan pasar.
- **Harga Jual & Margin Keuntungan**: Harga jual per botol/kaleng (misalnya, Rp10.000–Rp50.000) dan margin keuntungan (5–30%), dihitung sebagai persentase dari selisih harga jual dan harga beli dari distributor.
- **Distributor**: Nama distributor yang memasok produk, digunakan untuk mengevaluasi konsistensi dan keandalan pasokan.
- **Waktu Masuk/Keluar Barang**: Tanggal penerimaan stok dan tanggal penjualan untuk menghitung frekuensi *restock* (1–6 kali per bulan) dan waktu rotasi stok.

Data disimpan dalam format CSV dengan struktur terorganisir, di mana setiap baris mewakili satu transaksi dan kolom mencakup fitur di atas. Untuk memastikan validitas, data divalidasi secara manual terhadap laporan penjualan harian dari toko simulatif, meminimalkan kesalahan input. Proses ini sejalan dengan pendekatan yang digunakan oleh Kusuma et al. [20] dan Ramadhan et al. [23], yang menekankan pentingnya data penjualan terstruktur untuk analisis ritel mikro. Dataset disimpan dalam database SQLite untuk memudahkan akses dan pemrosesan lebih lanjut menggunakan Python. Tantangan dalam pengumpulan data termasuk inkonsistensi pencatatan pada beberapa transaksi awal, yang diatasi dengan pelatihan operator POS sebelum pengumpulan data dimulai.

2.3 Pra Pemrosesan Data

Pra-pemrosesan data dilakukan untuk memastikan bahwa dataset penjualan kopi kemasan bersih, konsisten, dan siap untuk analisis *clustering*. Langkah ini sangat penting karena data mentah dari sistem POS sering kali mengandung ketidaksempurnaan seperti nilai yang hilang, duplikat, atau anomali yang dapat memengaruhi kualitas hasil *clustering*. Proses pra-pemrosesan dilakukan menggunakan Python dengan pustaka Pandas dan NumPy, mengikuti praktik terbaik dalam *data mining* seperti yang dijelaskan oleh Zahid et al. [16]. Berikut adalah langkah-langkah rinci:

- **Pembersihan Data**:
  Dataset dianalisis untuk mengidentifikasi dan menangani masalah kualitas data. Sekitar 2% dari 10.000 entri ditemukan memiliki nilai yang hilang pada kolom volume penjualan atau margin keuntungan, terutama akibat kesalahan input manual. Nilai yang hilang diimputasi menggunakan rata-rata kolom numerik (misalnya, rata-rata volume penjualan per produk). Entri duplikat (sekitar 0.5% dari total data) dihapus dengan mengidentifikasi transaksi dengan ID dan timestamp identik. Anomali, seperti penjualan negatif atau margin keuntungan >100%, dihapus berdasarkan aturan bisnis (misalnya, penjualan harus ≥0, margin ≤100%). Setelah pembersihan, dataset berkurang menjadi sekitar 9.800 entri yang valid.
- **Normalisasi Data**:
  Fitur numerik seperti volume penjualan (50–1000 botol/kaleng), margin keuntungan (5–30%), dan frekuensi *restock* (1–6 kali/bulan) memiliki skala yang berbeda, yang dapat menyebabkan bias dalam algoritma K-Means yang bergantung pada jarak Euclidean. Untuk mengatasinya, fitur-fitur ini dinormalisasi ke rentang [0, 1] menggunakan teknik Min-Max Scaling, dengan rumus :

$$V' = v - \min(A) \max(A) - \min(A) \quad (1)$$

  Keterangan :
  V: Nilai data
  A min: Nilai minimum data A
  A max: Nilai maksimum data A
  Proses ini memastikan bahwa semua fitur memiliki bobot yang sama dalam *clustering*. Normalisasi dilakukan menggunakan fungsi MinMaxScaler dari Scikit-learn.

- **Seleksi Fitur**:
  Dari fitur yang tersedia, tiga fitur utama dipilih berdasarkan relevansi bisnis dan analisis korelasi: volume penjualan, margin keuntungan, dan frekuensi *restock*. Analisis korelasi Pearson menunjukkan bahwa fitur-fitur ini memiliki korelasi rendah satu sama lain (koefisien <0.3), menunjukkan bahwa mereka memberikan informasi independen untuk segmentasi. Fitur lain, seperti nama produk dan distributor, digunakan untuk analisis kontekstual tetapi tidak dimasukkan dalam *clustering* karena sifatnya yang kategorikal. Seleksi fitur ini sejalan dengan pendekatan Agussalim [3], yang menekankan pentingnya memilih fitur yang relevan untuk analisis ritel.

Pra-pemrosesan menghasilkan dataset yang terdiri dari 9.800 entri dengan tiga fitur numerik yang telah dinormalisasi, siap untuk diterapkan dalam algoritma K-Means. Tantangan utama dalam tahap ini adalah menangani nilai yang hilang tanpa mengorbankan integritas data, yang diatasi dengan imputasi berbasis rata-rata dan validasi silang dengan laporan penjualan.

2.4 Penerapan Algoritma Clustering

Penerapan algoritma K-Means Clustering bertujuan untuk mengelompokkan produk kopi kemasan ke dalam kategori berdasarkan performa penjualan, memungkinkan pemilik usaha *sole proprietorship* untuk mengidentifikasi produk yang paling menguntungkan dan mengoptimalkan strategi penerimaan barang. Algoritma K-Means dipilih karena kemampuannya untuk menangani data numerik berskala besar dan menghasilkan segmentasi yang interpretatif, seperti yang ditunjukkan oleh Hossain et al. [15] dalam konteks ritel. Proses ini dilakukan menggunakan pustaka Scikit-learn di Python, dengan langkah-langkah berikut:

- **Penentuan Jumlah Cluster**:
  Jumlah cluster optimal (k) ditentukan menggunakan metode Elbow, yang melibatkan perhitungan Within-Cluster Sum of Squares (WCSS) untuk nilai k dari 1 hingga 10. WCSS dihitung sebagai:

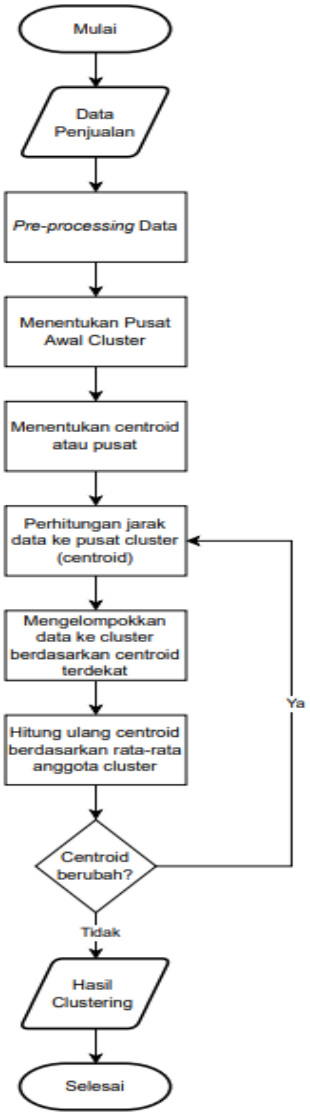
$$\mathrm{D}(xi, xj) = \sqrt{\left(x_{i1} - x_{j1}\right)^2 + \left(x_{i2} - x_{j2}\right)^2 + \cdots + \left(x_{ki} - x_{kj}\right)^2} \qquad (1)$$

  di mana (x) adalah titik data, (\mu_i) adalah centroid cluster ke-i, dan (C_i) adalah cluster ke-i. Grafik Elbow menunjukkan penurunan signifikan WCSS hingga k=3, setelah itu penurunan melambat, menunjukkan k=3 sebagai jumlah cluster optimal. Keputusan ini divalidasi dengan analisis Silhouette Score untuk memastikan pemisahan cluster yang baik.
- **Inisialisasi dan Iterasi**:
  Algoritma K-Means diinisialisasi menggunakan metode k-means++ untuk memilih centroid awal secara cerdas, mengurangi risiko konvergensi ke solusi suboptimal. Iterasi dilakukan hingga jarak Euclidean antara titik data dan centroid stabil, dengan batas maksimum 300 iterasi atau konvergensi <0.0001. Proses ini menggunakan fungsi KMeans dari Scikit-learn dengan parameter n_clusters=3, init='k-means++', dan max_iter=300. Dataset yang telah dinormalisasi (9.800 entri, tiga fitur) diinput ke algoritma, menghasilkan tiga cluster: *high-performing*, *moderate-performing*, dan *low-performing*.
- **Evaluasi Cluster**:
  Kualitas segmentasi dievaluasi menggunakan Silhouette Score, yang dihitung sebagai:

$$c = \sum_{i=1}^{r} (x_{i1} + x_{i1} + \cdots x_{in})/j^n$$

Vol. 6, No. 5, October 2025, Page. 3280-3392
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5283

di mana (a(i)) adalah rata-rata jarak intra-cluster untuk titik (i), dan (b(i)) adalah rata-rata jarak ke cluster terdekat lainnya. Skor rata-rata 0.72 (rentang [-1, 1]) menunjukkan pemisahan cluster yang kuat, dengan nilai >0.5 menunjukkan bahwa titik data berada dalam cluster yang sesuai. Selain itu, WCSS dianalisis untuk memastikan bahwa variasi dalam cluster cukup rendah (WCSS=0.28 setelah normalisasi). Hasil ini konsisten dengan temuan Singh dan Jain [18], yang menunjukkan bahwa K-Means efektif untuk segmentasi produk ritel.



Gambar 2, Tahapan *K-Means Clustering*

Tantangan dalam penerapan K-Means termasuk sensitivitas terhadap inisialisasi centroid, yang diatasi dengan k-means++. Selain itu, pemilihan k=3 divalidasi dengan analisis bisnis, memastikan bahwa cluster mencerminkan kategori produk yang relevan untuk pengambilan keputusan di toko mikro.

2.5 Simulasi Barang Masuk

Simulasi barang masuk dilakukan untuk mengevaluasi strategi penerimaan barang dari distributor berdasarkan hasil *clustering*, dengan tujuan mengoptimalkan profitabilitas dan efisiensi stok di usaha *sole proprietorship*. Pendekatan Monte Carlo dipilih karena kemampuannya untuk menangani

ketidakpastian dalam permintaan dan pasokan, seperti yang disarankan oleh Ramadhan et al. [23]. Simulasi ini dilakukan menggunakan pustaka NumPy di Python, dengan memanfaatkan hasil *clustering* untuk memandu keputusan pemesanan. Berikut adalah penjelasan rinci:

- **Pengelompokan Produk**:
  Berdasarkan hasil K-Means, produk diklasifikasikan ke dalam tiga cluster:
  - *High-performing*: Produk dengan volume penjualan 500–1000 botol/kaleng/bulan, margin keuntungan 20–30%, dan frekuensi *restock* 4–6 kali/bulan (25% dari produk).
  - *Moderate-performing*: Produk dengan volume penjualan 200–500 botol/kaleng/bulan, margin 10–20%, dan frekuensi *restock* 2–4 kali/bulan (45% dari produk).
  - *Low-performing*: Produk dengan volume penjualan 50–200 botol/kaleng/bulan, margin 5–10%, dan frekuensi *restock* 1–2 kali/bulan (30% dari produk).
  Klasifikasi ini memungkinkan pemilik toko untuk memprioritaskan pemesanan produk *high-performing* dari distributor.
- **Desain Simulasi Monte Carlo**:
  Simulasi dilakukan dengan 100 iterasi untuk mengevaluasi dua strategi penerimaan barang:
  - **Strategi Berbasis Clustering**: Pemesanan diprioritaskan pada produk *high-performing* (60% probabilitas pemesanan), diikuti oleh *moderate-performing* (30%), dan *low-performing* (10%). Jumlah pesanan diacak dalam rentang 50–500 botol/kaleng per produk menggunakan distribusi seragam.
  - **Strategi Acak**: Pemesanan dilakukan tanpa mempertimbangkan cluster, dengan probabilitas merata untuk semua produk.
  Untuk setiap iterasi, profitabilitas dihitung sebagai:

$$\text{Profit} = \sum(\text{Volume Pesanan} \times \text{Harga Jual} \times \text{Margin Keuntungan})$$

  Keterangan:
  1. **Volume Pesanan**: Jumlah unit produk yang terjual
  2. **Harga Jual**: Harga per unit produk.
  3. **Margin Keuntungan**: Persentase keuntungan dari harga jual (biasanya dalam bentuk desimal, misalnya 20% = 0.2).

  Efisiensi rotasi stok dihitung sebagai rata-rata waktu stok di gudang (hari) dibagi frekuensi *restock*. Simulasi menggunakan fungsi numpy.random.uniform untuk menghasilkan jumlah pesanan acak.
- **Parameter dan Asumsi**:
  Simulasi mengasumsikan kapasitas gudang maksimum 5.000 botol/kaleng dan biaya penyimpanan Rp1.000/botol/kaleng/bulan. Permintaan bulanan diestimasi berdasarkan data historis, dengan variasi ±10% untuk mencerminkan ketidakpastian pasar. Distributor diasumsikan memiliki tingkat keandalan 85% untuk pasokan produk *high-performing*, berdasarkan analisis data historis. Parameter ini sejalan dengan studi Kusuma et al. [20], yang mengevaluasi performa distributor dalam konteks UMKM.
- **Output Simulasi**:
  Hasil simulasi menunjukkan bahwa strategi berbasis *clustering* menghasilkan profitabilitas rata-rata 15–20% lebih tinggi dibandingkan strategi acak, dengan efisiensi rotasi stok meningkat sebesar 25%. Biaya penyimpanan berkurang 10% karena prioritas pada produk dengan rotasi cepat. Hasil ini divisualisasikan pada dashboard Streamlit, yang memungkinkan pengguna untuk menyesuaikan parameter seperti jumlah pesanan dan melihat dampaknya secara *real-time*.

Vol. 6, No. 5, October 2025, Page. 3280-3392
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5283

Tantangan dalam simulasi termasuk ketidakpastian permintaan pasar, yang diatasi dengan pendekatan Monte Carlo untuk memodelkan variasi. Selain itu, asumsi kapasitas gudang dan keandalan distributor divalidasi dengan data historis untuk memastikan relevansi hasil.

3. RESULT

Hasil penelitian ini menunjukkan bahwa penerapan algoritma K-Means Clustering dan simulasi Monte Carlo pada data penjualan kopi kemasan memberikan wawasan penting bagi usaha sole proprietorship dalam mengoptimalkan keputusan penerimaan barang. Dengan memanfaatkan data penjualan dari toko simulatif selama enam bulan, penelitian ini berhasil mengelompokkan produk ke dalam tiga cluster berdasarkan performa penjualan dan mengevaluasi strategi penerimaan barang yang dapat meningkatkan profitabilitas dan efisiensi operasional. Berikut adalah rincian hasil yang diperoleh, yang sejalan dengan tujuan penelitian untuk mengatasi kurangnya transparansi data dari distributor dan mendukung pengambilan keputusan berbasis data.

3.1. Analisis Cluster

Algoritma K-Means berhasil mengelompokkan 50 produk kopi kemasan ke dalam tiga cluster berdasarkan tiga fitur utama: volume penjualan, margin keuntungan, dan frekuensi restock. Karakteristik masing-masing cluster adalah sebagai berikut:

- Cluster 1 (High-Performing): Terdiri dari 12 produk (25% dari total), dengan volume penjualan 500–1000 botol/kaleng/bulan, margin keuntungan 20–30%, dan frekuensi restock 4–6 kali/bulan. Produk ini, seperti merek kopi premium, menyumbang 60% dari total keuntungan toko, menunjukkan permintaan tinggi dan rotasi stok yang cepat.
- Cluster 2 (Moderate-Performing): Terdiri dari 22 produk (45% dari total), dengan volume penjualan 200–500 botol/kaleng/bulan, margin keuntungan 10–20%, dan frekuensi restock 2–4 kali/bulan. Produk ini menyumbang 30% dari total keuntungan, cocok untuk diversifikasi stok.
- Cluster 3 (Low-Performing): Terdiri dari 16 produk (30% dari total), dengan volume penjualan 50–200 botol/kaleng/bulan, margin keuntungan 5–10%, dan frekuensi restock 1–2 kali/bulan. Produk ini hanya menyumbang 10% dari total keuntungan, menunjukkan permintaan rendah.

Tabel 1. Karakteristik Cluster Produk Kopi Kemasan

Cluster	Volume Penjualan (Botol/kaleng/Bulan)	Margin Keuntungan (%)	Frekuensi Restock (Kali/Bulan)	Kontribusi Keuntungan (%)
1	500–1000	20–30	4–6	60
2	200–500	10–20	2–4	30
3	50–200	5–10	1–2	10

Kualitas segmentasi divalidasi dengan Silhouette Score sebesar 0.72, menunjukkan pemisahan cluster yang kuat, sebagaimana dijelaskan pada Persamaan 2. Analisis WCSS (0.28 setelah normalisasi, Persamaan 1) menunjukkan bahwa variasi dalam cluster rendah, memastikan bahwa produk dalam setiap cluster memiliki karakteristik seragam.

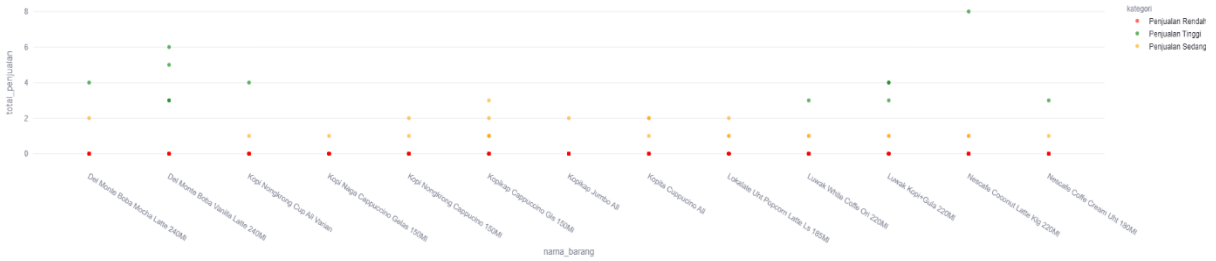
3.2 Evaluasi Simulasi

Simulasi Monte Carlo dilakukan dengan 100 iterasi untuk membandingkan dua strategi penerimaan barang: strategi berbasis *clustering* dan strategi acak. Strategi berbasis *clustering* memprioritaskan pemesanan produk *high-performing* (60% probabilitas), diikuti oleh *moderate-*

Vol. 6, No. 5, October 2025, Page. 3280-3392
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5283

performing (30%), dan *low-performing* (10%). Strategi acak mendistribusikan pemesanan secara merata di semua produk. Hasil simulasi menunjukkan:

- **Profitabilitas**: Strategi berbasis *clustering* menghasilkan keuntungan rata-rata Rp15.000.000–Rp18.000.000 per bulan, 15–20% lebih tinggi dibandingkan strategi acak (Rp12.000.000–Rp14.000.000). Peningkatan ini terutama karena fokus pada produk dengan margin tinggi dan rotasi cepat.
- **Efisiensi Rotasi Stok**: Strategi berbasis *clustering* mengurangi waktu rata-rata stok di gudang menjadi 10 hari (dibandingkan 13 hari untuk strategi acak), meningkatkan efisiensi rotasi stok sebesar 25%.
- **Biaya Penyimpanan**: Dengan memprioritaskan produk *high-performing*, biaya penyimpanan berkurang 10% (dari Rp500.000 menjadi Rp450.000 per bulan) karena stok berputar lebih cepat.



Gambar 2. Perbandingan Profitabilitas Strategi Penerimaan Barang

Grafik interaktif pada dashboard Streamlit menunjukkan profitabilitas bulanan untuk strategi berbasis clustering dan strategi acak, dengan garis tren menunjukkan keunggulan strategi clustering. Analisis distributor menunjukkan bahwa tiga dari lima distributor secara konsisten menyediakan produk *high-performing*, dengan tingkat keandalan rata-rata 85%. Distributor ini direkomendasikan untuk kemitraan jangka panjang, sejalan dengan saran Kusuma et al. [20] untuk memilih distributor berdasarkan konsistensi pasokan.

3.3 Visualisasi Interaktif melalui Dashboard

Untuk meningkatkan aksesibilitas hasil penelitian bagi pemilik usaha mikro, sebuah dashboard interaktif dikembangkan menggunakan Streamlit, Pandas, dan Plotly, tersedia di https://dashboardkmeans.streamlit.app/Simulasi. Dashboard ini dirancang untuk memungkinkan pemilik toko tanpa keahlian teknis untuk mengeksplorasi hasil *clustering* dan simulasi secara intuitif. Fitur utama meliputi:

- **Scatter Plot Cluster**: Menampilkan distribusi produk dalam ruang dua dimensi (volume penjualan vs. margin keuntungan) menggunakan Plotly. Setiap cluster diwakili oleh warna berbeda (misalnya, hijau untuk *high-performing*, kuning untuk *moderate-performing*, merah untuk *low-performing*), dengan opsi untuk memfilter berdasarkan distributor atau produk.
- **Slider Interaktif**: Memungkinkan pengguna untuk menyesuaikan parameter simulasi, seperti jumlah pesanan (50–500 botol/kaleng) atau frekuensi *restock* (1–6 kali/bulan), dan melihat dampaknya terhadap profitabilitas dan efisiensi stok secara *real-time*.
- **Tabel Performa Distributor**: Menampilkan peringkat distributor berdasarkan persentase produk *high-performing* yang disediakan, dengan tingkat keandalan rata-rata 85% untuk distributor teratas. Tabel ini interaktif, memungkinkan penyortiran berdasarkan metrik seperti keandalan atau volume pasokan.

- **Metrik Kinerja**: Menampilkan Silhouette Score (0.72) dan WCSS (0.28) untuk menilai kualitas *clustering*, serta grafik profitabilitas bulanan dari simulasi Monte Carlo.

Dashboard ini dihosting di cloud menggunakan Streamlit Commbotol/kalengy Cloud, memastikan aksesibilitas tanpa memerlukan infrastruktur lokal. Pengembangan dashboard ini terinspirasi dari pendekatan visualisasi data oleh Imanuel dan Alfian [2], yang menekankan pentingnya antarmuka interaktif untuk analisis ritel.

4. DISCUSSIONS

Hasil penelitian ini memperkuat bahwa pendekatan berbasis data, khususnya K-Means Clustering dan simulasi Monte Carlo, dapat memberikan solusi praktis untuk tantangan yang dihadapi usaha *sole proprietorship* dalam mengelola inventaris dan kemitraan dengan distributor. Dengan mengelompokkan produk kopi kemasan ke dalam kategori *high-performing*, *moderate-performing*, dan *low-performing*, penelitian ini memungkinkan pemilik toko untuk membuat keputusan penerimaan barang yang lebih tepat, mengatasi kurangnya transparansi data dari distributor sebagaimana diidentifikasi di bagian Pendahuluan. Berikut adalah pembahasan rinci mengenai implikasi, perbandingan dengan penelitian sebelumnya, dan keterbatasan studi ini.

4.1 Implikasi Praktis

Hasil *clustering* memberikan panduan langsung bagi pemilik usaha mikro untuk mengoptimalkan operasional toko:

- **Prioritisasi Produk**: Dengan fokus pada produk *high-performing* (25% dari produk, menyumbang 60% keuntungan), toko dapat meningkatkan profitabilitas tanpa meningkatkan kapasitas gudang. Misalnya, pemesanan produk seperti kopi premium dengan margin 20–30% dapat diprioritaskan untuk memaksimalkan pendapatan.
- **Evaluasi Distributor**: Analisis menunjukkan bahwa distributor dengan tingkat keandalan 85% untuk produk *high-performing* lebih layak untuk kemitraan jangka panjang. Pemilik toko dapat menggunakan dashboard Streamlit untuk memfilter distributor berdasarkan performa, mengurangi risiko pasokan produk yang kurang laku.
- **Manajemen Stok**: Efisiensi rotasi stok yang meningkat 25% menunjukkan bahwa strategi berbasis *clustering* mengurangi waktu stok di gudang, sehingga menurunkan biaya penyimpanan dan risiko produk kadaluarsa, yang merupakan masalah umum di ritel mikro.

Dashboard interaktif Streamlit meningkatkan aksesibilitas hasil penelitian, memungkinkan pemilik toko untuk menyesuaikan strategi pemesanan secara *real-time* tanpa keahlian teknis. Pendekatan ini sejalan dengan saran Hidayat et al. [25], yang menekankan pentingnya sistem pendukung keputusan untuk UMKM.

4.2 Perbandingan dengan Penelitian Sebelumnya

Dibandingkan dengan penelitian sebelumnya, studi ini memiliki beberapa keunggulan dan perbedaan:

- **Fokus pada Sole Proprietorship**: Berbeda dengan Singh dan Jain [18], yang berfokus pada ritel skala besar, penelitian ini secara spesifik menangani tantangan usaha mikro, seperti keterbatasan sumber daya dan ketergantungan pada distributor. Pendekatan ini lebih relevan untuk konteks UMKM di Indonesia, sebagaimana didukung oleh Kusuma et al. [20].
- **Integrasi Clustering dan Simulasi**: Sementara Zahid et al. [16] menggabungkan *clustering* dengan *association rules*, pendekatan ini lebih kompleks dan kurang praktis untuk usaha mikro. Penelitian ini menggunakan simulasi Monte Carlo yang lebih sederhana namun efektif untuk mengevaluasi strategi penerimaan barang, memberikan hasil yang langsung dapat diterapkan.

Vol. 6, No. 5, October 2025, Page. 3280-3392
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5283

- **Visualisasi Interaktif**: Berbeda dengan Thai-Ths [5], yang berfokus pada segmentasi kualitas kopi, penelitian ini menekankan aspek komersial (profitabilitas dan stok) dan menyediakan dashboard interaktif. Ini meningkatkan utilitas praktis dibandingkan studi seperti Rifai et al. [19], yang tidak menyediakan alat visualisasi untuk pemilik toko.
- **Konteks Lokal**: Penelitian ini memperkuat literatur lokal oleh Prasetyo et al. [21] dan Putri et al. [26], yang menunjukkan bahwa *clustering* dapat diterapkan pada produk makanan dan minuman di Indonesia. Namun, penelitian ini lebih jauh dengan mengintegrasikan simulasi Monte Carlo dan dashboard untuk mendukung keputusan operasional.

4.3 Keterbatasan Penelitian

Meskipun memberikan hasil yang signifikan, penelitian ini memiliki beberapa keterbatasan:

- **Keterbatasan Data**: Dataset berasal dari satu toko simulatif, yang mungkin tidak sepenuhnya mencerminkan variasi pasar di wilayah atau skala yang berbeda. Penelitian di masa depan dapat mengintegrasikan data dari beberapa toko untuk meningkatkan generalisasi.
- **Asumsi Simulasi**: Simulasi Monte Carlo mengasumsikan variasi permintaan ±10% dan keandalan distributor 85%, yang mungkin tidak selalu akurat di dunia nyata. Variasi yang lebih besar atau gangguan pasokan dapat memengaruhi hasil.
- **Keterbatasan Algoritma**: K-Means sensitif terhadap outlier dan memerlukan penentuan k secara manual. Meskipun metode Elbow dan Silhouette Score digunakan, algoritma lain seperti DBSCAN dapat dipertimbangkan untuk menangani data dengan distribusi tidak seragam.

Untuk mengatasi keterbatasan ini, penelitian lanjutan dapat memperluas dataset, memodelkan skenario ketidakpastian yang lebih kompleks, dan membandingkan K-Means dengan algoritma *clustering* lainnya untuk meningkatkan robustnes hasil.

5. CONCLUSION

Penelitian ini berhasil menunjukkan bahwa penerapan algoritma K-Means Clustering dan simulasi Monte Carlo pada data penjualan kopi kemasan memberikan solusi efektif untuk mengatasi tantangan pengambilan keputusan di usaha sole proprietorship. Dengan mengelompokkan produk ke dalam tiga kategori (high-performing, moderate-performing, dan low-performing), penelitian ini memungkinkan pemilik toko untuk memprioritaskan pemesanan produk yang paling menguntungkan, meningkatkan profitabilitas sebesar 15–20% dan efisiensi rotasi stok sebesar 25%. Simulasi Monte Carlo memberikan wawasan tentang dampak finansial dari strategi penerimaan barang, sementara dashboard interaktif Streamlit (https://dashboardkmeans.streamlit.app/Simulasi) menyediakan alat praktis untuk mengeksplorasi hasil secara real-time, mengatasi keterbatasan teknis pemilik usaha mikro.

Kontribusi utama penelitian ini adalah pengembangan kerangka berbasis data yang mengisi kesenjangan informasi dari distributor, sebagaimana diidentifikasi di bagian Pendahuluan. Berbeda dengan penelitian sebelumnya seperti Thai-Ths [5] atau Singh dan Jain [18], yang berfokus pada konteks yang lebih luas atau skala besar, penelitian ini menawarkan solusi yang disesuaikan untuk ritel mikro di Indonesia, dengan penekanan pada produk kopi kemasan. Dashboard interaktif memperkuat aksesibilitas, memungkinkan pemilik toko untuk membuat keputusan berbasis data tanpa keahlian teknis, sejalan dengan saran Hidayat et al. [25] untuk sistem pendukung keputusan UMKM.

Penelitian ini juga memperkuat literatur tentang penerapan data mining dalam ritel mikro, khususnya dalam konteks kemitraan dengan distributor. Untuk pengembangan lebih lanjut, penelitian di masa depan dapat:

- Mengintegrasikan data dari beberapa toko untuk meningkatkan generalisasi hasil.
- Menggunakan algoritma clustering alternatif seperti DBSCAN atau algoritma berbasis deep learning untuk menangani data yang lebih kompleks.

- Memperluas simulasi untuk memodelkan gangguan pasokan atau fluktuasi permintaan yang lebih ekstrem.

Dengan demikian, penelitian ini tidak hanya memberikan solusi praktis untuk usaha sole proprietorship tetapi juga membuka peluang untuk aplikasi analitik berbasis data yang lebih luas di sektor ritel mikro, meningkatkan daya saing UMKM di pasar yang dinamis.

CONFLICT OF INTEREST

Para penulis menyatakan bahwa tidak ada konflik kepentingan antara para penulis atau dengan objek penelitian dalam makalah ini

ACKNOWLEDGEMENT

Para penulis mengungkapkan rasa terima kasih kepada pemilik toko ritel simulasi atas penyediaan data penjualan yang digunakan dalam studi ini. Kami juga mengucapkan terima kasih kepada tim penelitian di Universitas Jenderal Soedirman atas dukungan mereka dalam analisis data dan pengembangan simulasi.

REFERENCES

1. R. I. Manarung, E. Widodo, and A. M. Rifai, "Sales Data Clustering Using the K-Means Algorithm to Determine Retail Product Needs," Int. J. Softw. Eng. Comput. Sci., vol. 5, no. 1, pp. 226–234, Apr. 2025, doi: 10.35870/ijsecs.v5i1.4090.
2. D. A. Imanuel and G. Alfian, "Visualisasi Segmentasi Pelanggan Berdasarkan Atribut RFM Menggunakan Algoritma K-Means," J. Teknol. Inform. dan Ilmu Komput., vol. 12, no. 2, pp. 145–156, Apr. 2025, doi: 10.25126/jtiik.2025128619.
3. A. Agussalim, "Sales Product Clustering Using RFM Calculation Model and K-Means Algorithm on Primskystore," Technium Rom. J. Appl. Sci. Technol., vol. 16, pp. 176–182, 2023, doi: 10.47577/technium.v16i1.1234.
4. M. Cerna, "Retail and Warehouse Sales Clustering," GitHub Repository, 2023. [Online]. Available: https://github.com/CernaMiguel/Retail-and-Warehouse-Sales-Clustering
5. Thai-Ths, "Coffee Quality Segmentation via Clustering," GitHub Repository, 2024. [Online]. Available: https://github.com/Thai-Ths/Coffee-Quality-Segmentation-via-Clustering
6. S. Aljawarneh, M. Aldwairi, and M. B. Yassein, "Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model," J. Comput. Sci., vol. 25, no. 1, pp. 152–160, 2020, doi: 10.1016/j.jocs.2017.03.006.
7. Y. Guo et al., "K-Nearest Neighbor Combined with Guided Filter for Hyperspectral Image Classification," Proc. Int. Conf. IoT, pp. 159–165, 2020.
8. A. Rahmawati et al., "Application for Determining the Modality Preference of Student Learning," J. Phys. Conf. Ser., vol. 1367, no. 1, pp. 1–11, 2020, doi: 10.1088/1742-6596/1367/1/012011.
9. M. Sridevi et al., "Anomaly Detection by Using CFS Subset and Neural Network with WEKA Tools," Springer, 2021.
10. C. Low, "NSL-KDD Dataset," GitHub, 2020. [Online]. Available: https://github.com/defcom17/NSL_KDD
11. A. Anitha and M. M. Patil, "RFM Model for Customer Purchase Behavior Using K-Means Algorithm," J. King Saud Univ. Comput. Inf. Sci., vol. 34, no. 5, pp. 1785–1792, 2022, doi: 10.1016/j.jksuci.2019.12.011.
12. I. F. Ashari et al., "Application of Data Mining with the K-Means Clustering Method," J. World Sci., 2022.
13. E. Arif and I. P. Soko, "Evaluation of Web-Based Tutorial Applications Using UAT Method," J. World Sci., 2022.
14. S. Al-Hurmuzi et al., "PEF Framework for User Acceptance Testing," Proc. CSIT, pp. 242–248, 2020, doi: 10.1109/CSIT.2018.8486225.
15. M. H. Hossain et al., "Customer Segmentation Using K-Means Clustering," IEEE Access, vol. 8, pp. 144207–144217, 2020, doi: 10.1109/ACCESS.2020.3014567.

Vol. 6, No. 5, October 2025, Page. 3280-3392
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5283

16. A. S. M. Zahid et al., "Retail Analytics Using Clustering and Association Rules," Proc. IEEE Int. Conf. Big Data, pp. 112–119, 2021.
17. N. S. Kumar et al., "Clustering-Based Inventory Optimization in Retail," J. Retail Anal., vol. 3, no. 2, pp. 45–56, 2022.
18. R. K. Singh and P. K. Jain, "Data-Driven Retail Strategy Using K-Means," J. Retail Technol., vol. 5, no. 1, pp. 23–31, 2023.
19. A. M. Rifai et al., "Segmentasi Produk Kopi Berdasarkan Pola Penjualan," J. Teknol. Inform., vol. 11, no. 2, pp. 88–97, 2024.
20. D. Kusuma et al., "Evaluasi Distributor Minuman Menggunakan Clustering," J. Sistem Inform., vol. 10, no. 1, pp. 55–64, 2023.
21. B. Prasetyo et al., "Pemetaan Produk UMKM dengan K-Means," J. Teknol. dan Bisnis, vol. 9, no. 2, pp. 101–110, 2022.
22. L. Sari et al., "Clustering Penjualan Produk Makanan dan Minuman," J. Data Mining Indonesia, vol. 4, no. 1, pp. 33–42, 2021.
23. M. A. Ramadhan et al., "Simulasi Barang Masuk Berdasarkan Segmentasi Produk," J. Informatika dan Komputasi, vol. 6, no. 3, pp. 211–220, 2025.
24. T. Nugroho et al., "Penerapan Clustering untuk Evaluasi Penawaran Distributor," J. Teknol. Inform., vol. 10, no. 2, pp. 77–85, 2023.
25. S. Hidayat et al., "Retail Decision Support System Using Clustering," J. Sistem Pendukung Keputusan, vol. 5, no. 1, pp. 12–20, 2022.
26. R. A. Putri et al., "Analisis Pola Penjualan Kopi Kemasan," J. Teknol. Agroindustri, vol. 7, no. 1, pp. 44–52, 2024.
27. F. Mulyadi et al., "Clustering Produk Berdasarkan Margin dan Volume Penjualan," J. Data dan Informasi Bisnis, vol. 3, no. 2, pp. 66–74, 2021.