

## Product Sales Analysis based on sales level using the K-Means Clustering Method

Aisha Bethary Kinasih<sup>1</sup>, Paminto Agung Christianto<sup>2</sup>, Nurul Amalia<sup>3</sup>

<sup>1</sup> Sistem Informasi, Fakultas Teknologi Informasi, Institut Widya Pratama, Kota Pekalongan, 51146, Indonesia

<sup>2,3</sup> Teknik Informatika, Fakultas Teknologi Informasi, Institut Widya Pratama, Kota Pekalongan, 51146, Indonesia

### Informasi Artikel

Diterima : 09 September 2025  
Revisi : 09 Januari 2026  
Publikasi : 31 Maret 2026

### Kata Kunci:

UMKM  
Data Mining  
K-Means Clustering  
Elbow Method  
Davies Bouldin Index

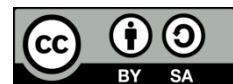
### ABSTRAK

Usaha Mikro, Kecil, dan Menengah (UMKM) memiliki peran yang sangat strategis dalam mendorong pertumbuhan ekonomi Indonesia. Meskipun demikian, sebagian besar pelaku usaha masih belum memanfaatkan teknologi digital secara optimal. Salah satunya adalah Toko Nabila Daster yang memiliki data penjualan sebanyak 475 transaksi periode Januari–Juni 2025, tetapi belum dilakukan analisis untuk mengetahui produk dengan tingkat penjualan tinggi, sedang atau rendah. Kondisi ini dapat menyebabkan penumpukan stok serta kurang tepatnya strategi promosi. Tujuan penelitian ini adalah melakukan pengelompokan produk berdasarkan tingkat penjualan dengan menerapkan metode *K-Means Clustering*. Jumlah kluster optimal ditentukan melalui pendekatan *Elbow Method*, sedangkan kualitas pengelompokan dievaluasi dengan *Davies Bouldin Index (DBI)*. Hasil analisis menunjukkan terbentuknya kluster produk yang membedakan kategori penjualan terlaris, kurang laris, dan tidak laris. Diharapkan dapat menjadi dasar dalam pengambilan keputusan bisnis, khususnya dalam strategi promosi serta pengelolaan stok agar lebih efektif.

### ABSTRACT

*Micro, Small, and Medium Enterprises (MSMEs) play a highly strategic role in driving Indonesia's economic growth. Nevertheless, most business actors have not yet utilized digital technology to its full potential. One such example is Toko Nabila Daster, which recorded 475 sales transactions during the period of January–June 2025, but has not conducted an analysis to identify products with high, medium, or low sales levels. This situation may result in stock accumulation and ineffective promotional strategies. The objective of this study is to group products based on their sales levels using the K-Means Clustering method. The optimal number of clusters is determined through the Elbow Method, while the quality of clustering is assessed using the Davies-Bouldin Index (DBI). The results of the analysis indicate the formation of product clusters that distinguish best-selling, moderately selling, and low-selling categories. These findings are expected to serve as a foundation for business decision-making, particularly in designing promotional strategies and managing inventory more efficiently.*

This is an open-access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



### \*Penulis Koresponden

Email: [aishabethary592@gmail.com](mailto:aishabethary592@gmail.com)

Cara sitasi IEEE::

A. B. Kinasih, P. A. Christianto, & N. Amalia, "uct Sales Analysis based on sales level using the K-Means Clustering Method," *Journal of Artificial Intelligence and Software Engineering (J-AISE)*, vol. 6, no. 1, pp. 1-7, Maret 2026. doi: 10.30811/jaise.v6i1.7710

## 1. PENDAHULUAN

Usaha Mikro, Kecil, dan Menengah (UMKM) berperan strategis dalam menopang perekonomian nasional. Data dari Kementerian Koperasi dan UKM, menunjukkan bahwa jumlah UMKM di Indonesia kini mencapai lebih dari 64 juta unit dan menyumbang 61,07% terhadap PDB atau setara dengan Rp8.573 triliun [1]. Dengan adanya perkembangan teknologi digital, UMKM memiliki kesempatan lebih besar dalam memperluas jangkauan pasar melalui pemanfaatan *e-commerce* dan media sosial. Namun, data menunjukkan bahwa sebagian besar pelaku UMKM masih terbatas dalam mengadopsi platform digital, baik karena minimnya literasi teknologi maupun preferensi terhadap sistem penjualan konvensional [2]. Kondisi ini menunjukkan masih terdapat kesenjangan dalam pemanfaatan teknologi digital untuk mendukung strategi bisnis.

Sebagai contoh salah satu UMKM yang menghadapi kendala tersebut adalah Toko Nabila Daster, yang berfokus pada penjualan pakaian batik rumahan atau yang sering dikenal dengan istilah daster. Hingga saat ini, toko tersebut masih mengandalkan pola penjualan konvensional dan belum terintegrasi dengan sistem digitalisasi. Meskipun toko ini sudah melakukan pencatatan transaksi secara manual, data tersebut belum diolah secara sistematis untuk mengetahui produk terlaris maupun kurang diminati. Akibatnya, pemilik usaha berpotensi mengalami penumpukan stok, kesulitan dalam menentukan strategi promosi, serta pengelolaan persediaan yang kurang optimal. Oleh karena itu, dibutuhkan analisis berbasis data yang dapat memberikan gambaran lebih mendalam terhadap pola penjualan produk.

Dalam penerapan *data mining*, *K-Means Clustering* terbukti mampu mengelompokkan data secara efektif berdasarkan kesamaan ciri. Algoritma ini banyak digunakan karena efisien dalam menangani dataset besar, sederhana dalam implementasi, dan hasilnya mudah dipahami [3]. Beberapa penelitian sebelumnya juga telah menunjukkan keberhasilan penerapan metode ini. Penelitian pada industri kosmetik, misalnya, mampu mengelompokkan produk *skincare* ke dalam kategori penjualan rendah, sedang, dan tinggi untuk mendukung strategi pemasaran [4]. Penelitian lain pada sektor farmasi menunjukkan bahwa *K-Means* dapat membagi produk obat menjadi kluster penjualan rendah dan tinggi dengan akurasi yang cukup baik [5].

Dalam penelitian ini, jumlah kluster optimal ditentukan melalui *Elbow Method* yaitu teknik yang mengidentifikasi titik tekukan pada grafik untuk memperoleh jumlah kluster yang paling tepat. [6]. Sedangkan kualitas hasil pengelompokan dievaluasi dengan *Davies Bouldin Index*, semakin rendah nilai DBI, semakin baik kualitas kluster yang terbentuk lebih rapat dan terpisah dengan jelas [7]. Kombinasi kedua metode ini diharapkan dapat menghasilkan pembagian kluster yang sesuai terhadap pola penjualan produk.

Kebaruan penelitian ini terletak pada penggunaan algoritma *K-Means Clustering* untuk menganalisis data penjualan pada usaha mikro bidang pakaian rumahan dengan studi kasus pada Toko Nabila Daster. Hasil penelitian diharapkan tidak hanya memberikan kontribusi akademis, tetapi juga kontribusi praktis bagi pemilik UMKM dalam meningkatkan strategi promosi dan pengelolaan stok yang lebih efektif.

## 2. METODE

### 2.1 Tahap Awal

Dalam penelitian ini, data diperoleh melalui beberapa tahapan, di antaranya :

#### 1) Wawancara

Proses wawancara dengan pemilik untuk memperoleh informasi yang relevan dengan penelitian, khususnya terkait data transaksi penjualan. Dari wawancara ini diperoleh informasi mengenai laporan penjualan serta atribut yang dibutuhkan antara lain nama produk, harga awal, harga jual, jumlah produk terjual. Informasi tersebut menjadi dasar dalam penyusunan dataset yang akan digunakan pada tahap pengolahan data.

#### 2) Studi Literatur

Kegiatan studi literatur dilakukan dengan membaca, memahami, mengkaji dan menganalisis sejumlah sumber referensi yang memiliki keterkaitan dengan topik atau permasalahan penelitian yang diteliti sehingga dapat menjadi landasan dalam penelitian.

### 2.2 Tahap Penelitian

#### 1) Tahap Data Selection

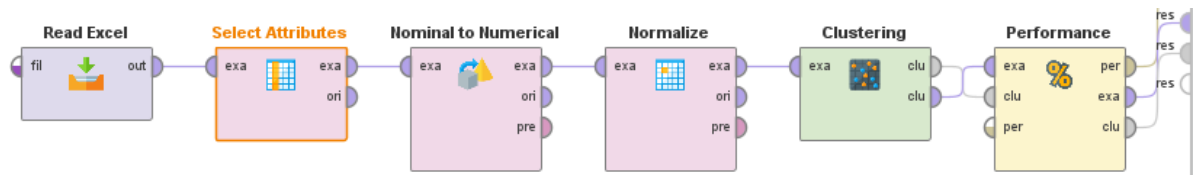
Proses penyeleksian data yang diperoleh dari tahapan pengumpulan data untuk menyesuaikan dengan analisa yang dilakukan. Data yang digunakan merupakan data transaksi penjualan dengan total 37 jenis produk dan 475 record transaksi pada periode Januari – Juni 2025. Adapun atribut yang digunakan sebanyak 6 antara lain: Id produk, harga awal, harga jual, jumlah produk terjual, total penjualan, dan total laba.

Berikut merupakan deskripsi variabel dataset yang digunakan.

Table 1. Deskripsi Variabel Data

No	Atribut	Tipe Data	Deskripsi
1.	Id_produk	Object	Kode unik yang digunakan untuk membedakan setiap produk.
2.	Harga Awal	Float	Harga dasar atau modal dari produk sebelum dijual
3.	Harga Jual	Float	Harga jual produk kepada konsumen.
4.	Jumlah produk terjual	Int	Total jumlah unit produk yang berhasil terjual pada periode tertentu.
5.	Total Penjualan	Float	Nilai keseluruhan penjualan yang diperoleh dari jumlah produk terjual dikalikan harga jual.
6.	Total Laba	Float	Keuntungan bersih dihitung berdasarkan perbedaan antara harga jual dan harga awal, lalu dikalikan dengan jumlah produk terjual.

- 2) Tahap Pre-processing dan Cleaning  
Tahap preprocessing dilakukan dengan memeriksa konsistensi data penjualan, menambahkan atribut laba sebagai variabel turunan yang dihitung dari selisih harga jual dan harga awal dikalikan jumlah produk terjual, serta memastikan tidak terdapat data kosong. Cleaning data pada penelitian ini adalah penghapusan atribut yang tidak dibutuhkan seperti tanggal dan variasi warna produk karena tidak termasuk dalam pengolahan perhitungan penjualan produk.
- 3) Tahap Data Transformation  
Pada tahap ini atribut berupa nama produk yang masih bertipe teks diubah menjadi kode numerik. Transformasi ini diperlukan agar data dapat dikenali dan diolah menggunakan algoritma K-Means yang hanya bisa memproses nilai berupa numerik.
- 4) Tahap Data Mining  
Tahap data mining dilakukan dengan mengolah data menggunakan teknik clustering dengan bantuan tools RapidMiner, yang memanfaatkan algoritma *K-Means* sebagai metode pengelompokannya.



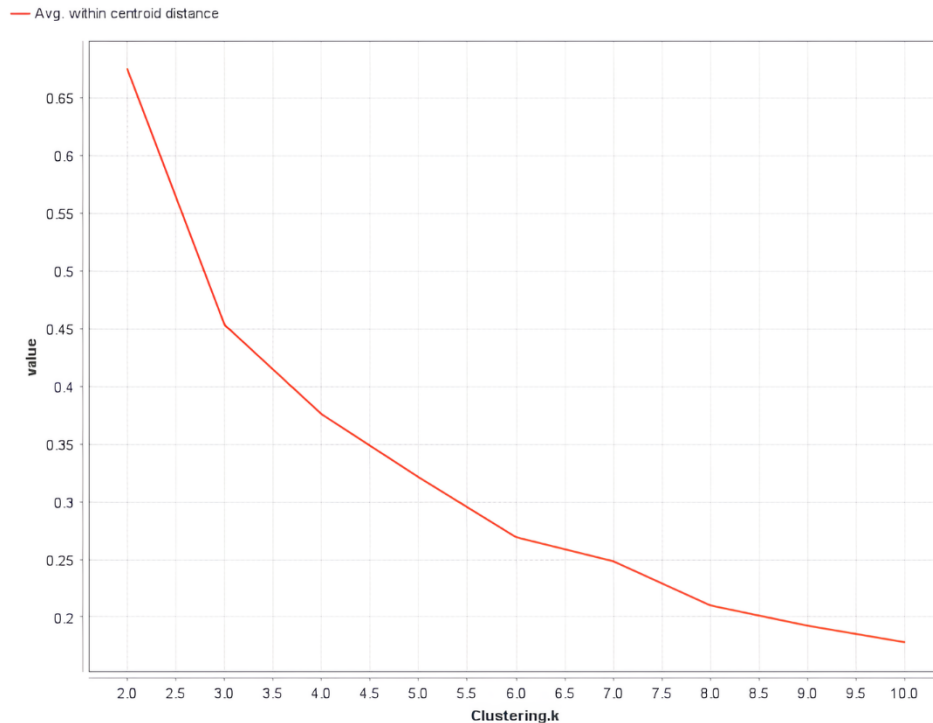
Gambar 1. Alur proses klastering K-Means menggunakan RapidMiner (Sumber : Aplikasi RapidMiner,2025)

Proses analisis dimulai dengan mengimpor data penjualan dari Excel ke dalam RapidMiner agar seluruh atribut dapat diolah secara menyeluruh. Selanjutnya dilakukan pemilihan atribut yang relevan, antara lain jumlah produk terjual, total penjualan, serta laba. Sementara ID produk tetap dipertahankan sebagai penanda. Untuk memastikan konsistensi data, tahap konversi nominal ke numerik diterapkan sehingga seluruh atribut dapat dikenali oleh algoritma. Pada tahap berikutnya dilakukan proses normalisasi agar setiap atribut memiliki skala yang seragam dan dapat dikenali dengan baik oleh algoritma. Setelah normalisasi barulah algoritma K-Means digunakan untuk mengelompokkan produk menjadi beberapa kluster berdasarkan kesamaan atribut atau ciri penjualannya..

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Penentuan kluster optimal menggunakan Metode Elbow dan Davies Bouldin Index ( DBI )

Penentuan jumlah kluster optimal dilakukan dengan Elbow Method. Dari grafik yang diperoleh, terlihat adanya titik tekukan pada  $k = 3$ , sehingga dapat disimpulkan jumlah kluster yang paling sesuai dalam penelitian ini adalah 3.



Gambar 2. Grafik Elbow Method yang menunjukkan titik optimal jumlah kluster (Sumber : Aplikasi RapidMiner,2025)

Table 2. Penentuan nilai k menggunakan Davies Bouldin Index

Iteration	Clustering	Davies Bouldin Index
1	2	0.116
2	3	0.099
3	4	0.109
4	5	0.103
5	6	0.102
6	7	0.119
7	8	0.110
8	9	0.115
9	10	0.122

Nilai yang diperoleh dari perhitungan Davies-Bouldin Index (DBI) memperlihatkan bahwa nilai DBI terkecil terdapat pada  $k = 3$  dengan nilai 0.099 yang menandakan pemisahan antar kluster cukup baik. Baik metode Elbow maupun DBI sama-sama menunjukkan jika jumlah kluster optimal adalah  $k = 3$ , sehingga nilai ini dipilih dalam penelitian.

#### 3.2 Hasil Clustering

Pengelompokan data menggunakan algoritma K-Means yang dijalankan melalui RapidMiner untuk menghasilkan output. Dengan menggunakan  $k = 3$  untuk membentuk kelompok produk dari dataset 475 data. Dari jumlah tersebut, 146 produk termasuk dalam Cluster 0, sebanyak 93 produk berada di Cluster 1, dan 236 produk lainnya masuk ke Cluster 2. Distribusi ini menunjukkan bahwa sebagian besar data terkonsentrasi pada Cluster 2, sedangkan sisanya terbagi ke dalam Cluster 0 dan Cluster 1. Kondisi tersebut menandakan adanya perbedaan pola penjualan yang cukup jelas antar kelompok produk.

Table 3. Hasil Clustering

Clustering	Cluster Model
0	146 items
1	93 items
2	236 items

### 3.3 Visualisasi hasil klustering dalam bentuk tabel

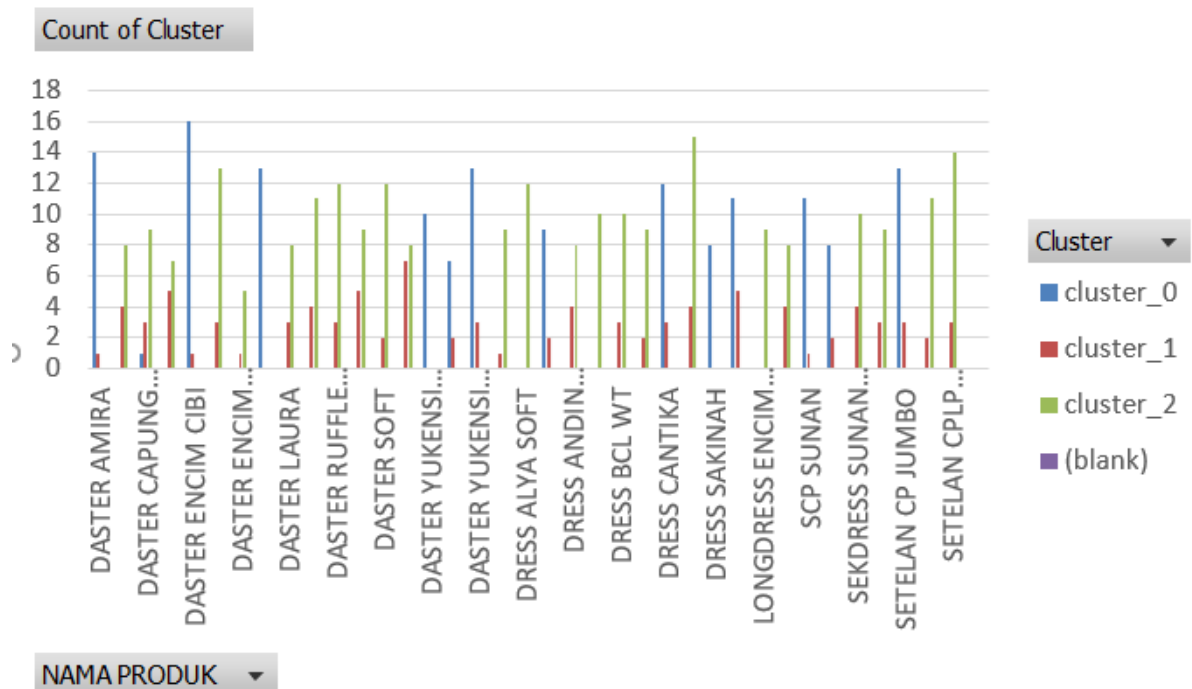
Table 4. Rata – Rata Hasil Klustering

No	Label	Avg Harga Awal	Avg Harga Jual	Avg Produk Terjual	Avg Total Penjualan	Avg Laba	Kategori
1.	Cluster_0	Rp 100.933	Rp 110.933	16 pcs	Rp 1.774.711	Rp 158.630	Kurang laris
2.	Cluster_1	Rp 53.427	Rp 63.427	72 pcs	Rp 4.487.035	Rp 715.161	Terlaris
3.	Cluster_2	Rp 38.123	Rp 48.123	18 pcs	Rp 836.167	Rp 179.492	Tidak Laris
	Total	Rp 60.425	Rp 70.425	28 pcs	Rp 1.839.448	Rp 277.958	

Hasil clustering menunjukkan bahwa Cluster\_1 termasuk kategori terlaris dengan rata-rata penjualan tertinggi 72 pcs. Cluster\_0 masuk kategori kurang laris dengan penjualan menengah 16 pcs, meskipun memiliki harga jual dan laba relatif tinggi. Sedangkan Cluster\_2 dikategorikan tidak laris karena rata-rata penjualannya paling rendah 18 pcs meskipun menghasilkan laba per produk cukup besar.

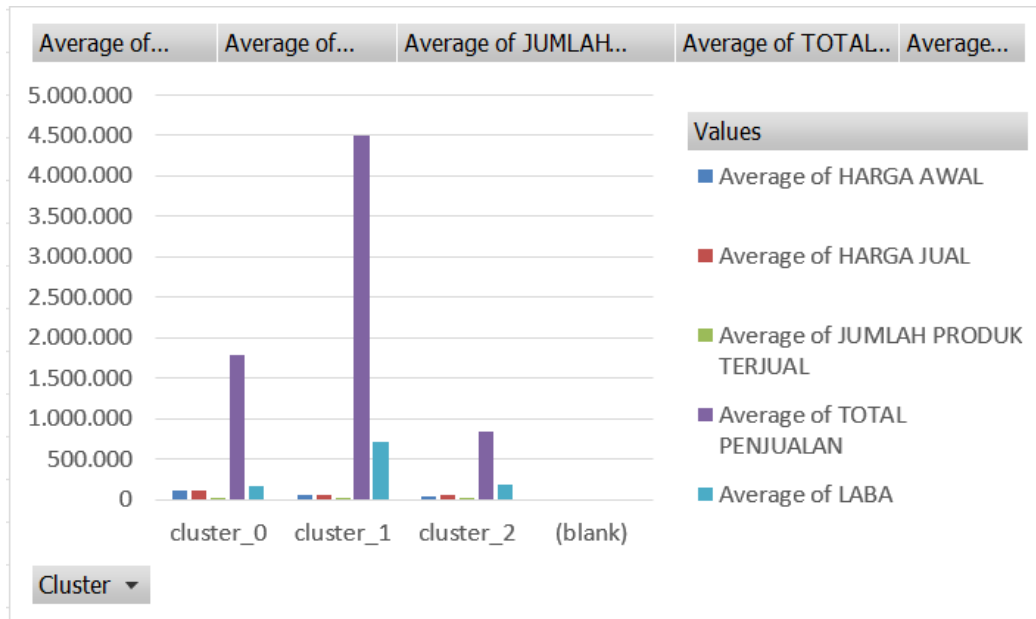
### 3.4 Visualisasi hasil klustering dalam bentuk grafik

Visualisasi hasil klustering memperlihatkan distribusi produk yang dikelompokkan berdasarkan nama produk. Grafik menunjukkan bahwa setiap kluster memiliki jumlah produk yang berbeda-beda, sehingga terlihat variasi distribusi pada masing-masing kelompok.



Gambar 3. Distribusi jumlah produk pada setiap kluster berdasarkan nama produk

Visualisasi berikutnya menampilkan perbandingan rata-rata harga awal, harga jual, jumlah produk terjual, total penjualan, dan laba pada masing-masing kluster. Dari grafik tersebut terlihat adanya perbedaan karakteristik antar kluster berdasarkan variabel-variabel tersebut.



Gambar 4. Perbandingan rata-rata variabel pada setiap kluster

### 3.5 Evaluasi

Evaluasi terhadap kualitas kluster dilakukan berdasarkan metode Davies-Bouldin Index (DBI). Indeks ini menilai seberapa dekat antar objek dalam satu kluster dibandingkan dengan kluster lainnya, semakin kecil nilai DBI maka semakin baik kualitas hasil pengelompokan [8].



Gambar 5. Hasil Perbandingan nilai Davies Bouldin Index (Sumber : Aplikasi RapidMiner,2025)

Pada pengujian dengan jumlah kluster  $k = 2$  diperoleh nilai sebesar 0,581, sementara untuk  $k = 4$  nilai DBI mencapai 0,544. Berdasarkan hasil perhitungan Davies Bouldin Index (DBI) jumlah kluster optimal ditetapkan pada  $k = 3$  karena memiliki nilai DBI terendah dibandingkan dengan jumlah kluster lainnya. Dengan nilai DBI pada  $k = 3$  adalah 0,494.

## 4. KESIMPULAN

Dari hasil penelitian terlihat bahwa algoritma K-Means Clustering efektif dalam melakukan pengelompokan 475 data penjualan Toko Nabila Daster ke dalam tiga kluster optimal. Melalui analisis nilai Davies-Bouldin Index (DBI) dan visualisasi data, diperoleh hasil bahwa Cluster 0 berisi produk dengan jumlah transaksi yang relatif stabil sehingga dikategorikan sebagai produk dengan penjualan sedang. Cluster 1 terdiri dari produk dengan penjualan tinggi dan kontribusi laba yang signifikan, sehingga dapat digolongkan sebagai produk terlaris. Sementara itu, Cluster 2 mencakup produk dengan volume penjualan paling rendah, yang berpotensi menimbulkan penumpukan stok, sehingga masuk kategori produk tidak laris. Hal yang menjadi pembeda sekaligus kebaruan dalam penelitian ini adalah penerapan algoritma K-Means untuk menganalisis penjualan produk pakaian rumahan pada UMKM yang masih jarang dilakukan dalam penelitian sebelumnya. Temuan ini dapat menjadi dasar dalam perumusan strategi promosi dan pengelolaan persediaan yang lebih terarah. Keterbatasan penelitian ini terletak pada data yang digunakan hanya transaksi penjualan selama enam bulan. Batasan tersebut dapat memengaruhi variasi pola penjualan yang terbentuk. Untuk penelitian berikutnya disarankan agar menggunakan periode waktu yang lebih luas atau menambahkan variabel lain agar hasil pengelompokan lebih akurat. Informasi yang diperoleh diharapkan dapat dimanfaatkan sebagai dasar dalam pengambilan keputusan bisnis, khususnya untuk strategi promosi dan pengelolaan persediaan yang lebih tepat sasaran.

**REFERENSI**

- [1] K. S. P. D. K. C. M. Junaidi, "Umkm Hebat, Perekonomian Nasional Meningkatkan," Ditjen Perbendaharaan Kemenkeu RI. Accessed: Aug. 01, 2025. [Online]. Available: <https://djpb.kemenkeu.go.id/kppn/curup/id/data-publikasi/artikel/2885-umkm-hebat,-perekonomian-nasional-meningkat.html>
- [2] P. Direktorat Statistik Keuangan, Teknologi Informasi, *Statistik E-Commerce 2023*. Badan Pusat Statistik, 2025.
- [3] F. Khalish, N. M. Piranti, and O. Martadireja, "Implementasi Data Mining Menggunakan Teknik Clustering dengan Metode K-Means," vol. 8, pp. 5392–5397, 2025.
- [4] M. Barata, I. S. Ayuni, A. Y. Kartini, and Z. Alawi, "Algoritma K-Means dalam Clustering Produk Skincare untuk Menentukan Strategi Pemasaran," *J. Inform. Polinema*, vol. 10, no. 3, pp. 421–428, 2024, doi: 10.33795/jip.v10i3.5167.
- [5] M. Fajar, N. Rahaningsih, and R. Danar Dana, "Analisis Pola Penjualan Obat Di Apotek an-Naafi Menggunakan Metode K-Means Clustering," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, pp. 486–492, 2024, doi: 10.36040/jati.v8i1.8395.
- [6] M. Cui, "on the Elbow Method," *Clausius Sci. Press*, vol. 1, no. 1, pp. 5–8, 2020, doi: 10.23977/accaf.2020.010102.
- [7] T. T. Alifa, R. Astuti, and F. M. Basysyar, "Implementasi Data Mining Menggunakan Algoritma K-Means Untuk Pengelompokkan," *J. Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 602–607, 2024.
- [8] F. Ros, R. Riad, and S. Guillaume, *PDBI: A partitioning Davies-Bouldin index for clustering evaluation*, vol. 528. 2023. doi: 10.1016/j.neucom.2023.01.043.