

## **ANALISIS DATA PENJUALAN PADA UMKM KONVEKSI MENGUNAKAN METODE K-MEANS CLUSTERING DENGAN MENERAPKAN CRISP-DM**

### **DATA ANALYSIS OF MSME CONVECTION USING K-MEANS CLUSTERING METHOD WITH CRISP-DM APPROACH**

**Ali Imron<sup>1</sup>, Erna Daniati<sup>2\*</sup>, Rina Firliana<sup>3</sup>, Muhammad Fauzan Aditiya Mufid<sup>4</sup>**

\*E-mail: [ernadaniati@unpkediri.ac.id](mailto:ernadaniati@unpkediri.ac.id)

<sup>1, 2, 3, 4</sup> Sistem Informasi, Teknik dan Ilmu komputer, Universitas Nusantara PGRI Kediri

#### **Abstrak**

UMKM konveksi memiliki data penjualan yang cukup besar, namun seringkali belum dimanfaatkan secara optimal untuk mendukung strategi bisnis. Penelitian ini bertujuan untuk menganalisis data penjualan menggunakan metode *K-Means Clustering* dengan pendekatan *Cross Industry Standard Process for Data Mining* (CRISP-DM). Data yang digunakan mencakup 360 catatan penjualan produk selama periode 2021–2023 dengan variabel jumlah terjual, harga satuan, dan omzet. Proses analisis dilakukan melalui tahapan CRISP-DM meliputi pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan penyajian hasil. Hasil evaluasi dengan nilai *Silhouette Score* menunjukkan jumlah cluster optimal sebanyak dua. Cluster pertama merepresentasikan produk dengan jumlah terjual tinggi, harga satuan relatif tinggi, dan omzet besar. Sementara cluster kedua berisi produk dengan jumlah terjual rendah, harga satuan lebih rendah, dan omzet kecil. Temuan ini dapat menjadi dasar bagi UMKM dalam merumuskan strategi pemasaran dan pengelolaan stok yang lebih efektif.

**Kata kunci:** *clustering, CRISP-DM, k-means, penjualan produk, UMKM konveksi*

#### **Abstract**

*MSME convection businesses often generate large amounts of sales data that are not fully utilized to support business strategies. This study aims to analyze sales data using the K-Means Clustering method with the Cross Industry Standard Process for Data Mining (CRISP-DM) approach. The dataset consists of 360 sales records from 2021–2023, covering the variables of units sold, unit price, and revenue. The analysis followed the CRISP-DM stages: business understanding, data understanding, data preparation, modeling, evaluation, and presentation. Evaluation results using the Silhouette Score indicate that the optimal number of clusters is two. The first cluster represents products with high sales volume, relatively higher unit prices, and large revenue, while the second cluster consists of products with lower sales, lower prices, and smaller revenue. These findings can be used as a basis for MSMEs to develop more effective marketing strategies and inventory management.*

**Keywords:** *clustering, CRISP-DM, k-means, MSME convection, sales analysis*

## **1. PENDAHULUAN**

Usaha Mikro, Kecil, dan Menengah (UMKM) merupakan salah satu pilar utama perekonomian nasional. Berdasarkan Undang-Undang Nomor 20 Tahun 2008, UMKM dikategorikan menurut jumlah aset dan omzet tahunan[1]. Data Kementerian Koperasi dan UKM (2023) mencatat bahwa UMKM menyumbang lebih dari 60% Produk Domestik Bruto (PDB) Indonesia dan menyerap

sekitar 97% tenaga kerja[2]. Sektor konveksi termasuk salah satu bidang UMKM yang memiliki kontribusi besar, karena permintaan sandang relatif stabil serta mampu menyerap banyak tenaga kerja. Namun, di balik potensi tersebut, banyak UMKM konveksi masih mengelola data penjualan secara sederhana, misalnya hanya mencatat dalam buku atau spreadsheet tanpa analisis lebih lanjut[3].

Pengelolaan data yang terbatas menyebabkan UMKM kesulitan memahami pola penjualan produk. Padahal, menurut teori *data-driven decision making*, keputusan yang berbasis data dapat meningkatkan efektivitas strategi bisnis dan mengurangi ketidakpastian[4]. Data penjualan bukan sekadar catatan transaksi, tetapi dapat menjadi sumber informasi strategis untuk memahami perilaku konsumen, tren permintaan, serta kontribusi tiap produk terhadap omzet. Oleh karena itu, diperlukan metode analisis yang dapat mengekstraksi pola dari data sehingga UMKM mampu menyusun strategi berbasis informasi yang lebih akurat.

Salah satu pendekatan analisis data yang banyak digunakan adalah clustering. Clustering merupakan teknik *unsupervised learning* yang bertujuan mengelompokkan data ke dalam cluster berdasarkan tingkat kemiripan karakteristik. Melalui penerapan clustering, pola-pola tersembunyi dalam data dapat terungkap, seperti membedakan produk penyumbang omzet utama dengan produk berperforma rendah[5]. Di antara berbagai metode clustering, K-Means merupakan salah satu algoritma yang paling populer karena kesederhanaannya, kecepatan proses, serta kemampuannya menangani data berskala besar. Algoritma ini bekerja dengan menentukan pusat cluster (*centroid*) dan mengelompokkan data berdasarkan jarak terdekat terhadap centroid tersebut[6]. Meski demikian, K-Means juga memiliki kelemahan, seperti perlunya menentukan jumlah cluster di awal dan sensitivitas terhadap data outlier[7].

Penelitian ini menggunakan kerangka *Cross Industry Standard Process for Data Mining* (CRISP-DM). CRISP-DM adalah model proses standar internasional yang banyak digunakan dalam data mining karena mencakup keseluruhan alur analisis mulai dari pemahaman bisnis hingga implementasi hasil. Terdapat enam tahapan utama dalam CRISP-DM, yaitu *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment* [8]. Dengan pendekatan ini, analisis data tidak hanya berhenti pada tahap pemodelan, tetapi juga menghasilkan wawasan yang aplikatif untuk kebutuhan bisnis.

Penelitian terdahulu telah menerapkan metode clustering pada berbagai sektor, misalnya untuk analisis pola pembelian di UMKM ritel[9] maupun segmentasi pelanggan e-commerce [10]. Namun, sebagian besar studi tersebut belum mengintegrasikan pendekatan CRISP-DM secara menyeluruh, sehingga proses analisis kurang sistematis dan sulit direplikasi. Oleh karena itu, penelitian ini menawarkan kontribusi baru dengan mengombinasikan metode K-Means Clustering dan pendekatan CRISP-DM dalam konteks UMKM konveksi.

Berdasarkan uraian tersebut, tujuan penelitian ini adalah untuk menganalisis data penjualan UMKM konveksi dengan menerapkan metode K-Means Clustering dan kerangka CRISP-DM. Penelitian ini diharapkan dapat menghasilkan segmentasi produk yang lebih terstruktur, sehingga UMKM dapat merumuskan strategi pemasaran, pengelolaan stok, dan pengambilan keputusan bisnis yang lebih efektif berbasis data.

## 2. METODOLOGI

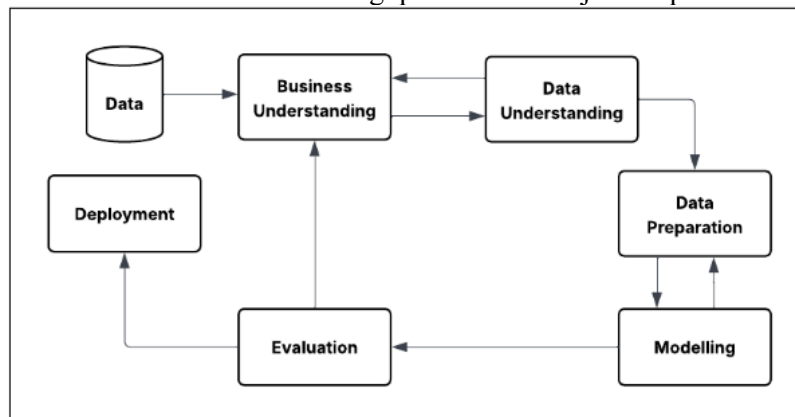
Penelitian ini menggunakan data penjualan dari salah satu UMKM konveksi selama periode 2021 hingga 2023. Dataset terdiri dari 360 catatan transaksi dengan enam atribut, yaitu Tahun, Bulan, Produk, Jumlah Terjual, Harga Satuan, dan Omzet. Analisis penelitian difokuskan pada tiga variabel inti, yaitu Jumlah Terjual, Harga Satuan, dan Omzet, karena dianggap paling representatif dalam menilai kinerja penjualan. Jumlah Terjual mencerminkan tingkat permintaan konsumen, Harga Satuan berkaitan dengan strategi harga yang memengaruhi daya saing produk, dan Omzet menunjukkan kontribusi finansial yang diperoleh. Dengan mempertimbangkan ketiga

variabel ini, performa produk dapat diamati dari sisi volume penjualan, nilai ekonomi, serta strategi harga. Ringkasan spesifikasi dataset ditunjukkan pada Tabel 1.

**Tabel 1. Spesifikasi Dataset Penelitian**

| Atribut        | Type Data | Keterangan                  |
|----------------|-----------|-----------------------------|
| Tahun          | Integer   | Tahun penjualan (2021–2023) |
| Bulan          | Integer   | Bulan penjualan (1–12)      |
| Produk         | String    | Nama produk konveksi        |
| Jumlah Terjual | Integer   | Banyaknya produk terjual    |
| Harga Satuan   | Integer   | Harga per unit (Rp)         |
| Omzet          | Integer   | Total pendapatan (Rp)       |

Penelitian ini menggunakan *Cross Industry Standard Process for Data Mining* (CRISP-DM) sebagai kerangka kerja yang terstruktur dan standar untuk membangun sistem data mining secara sistematis[11]. Pendekatan ini dipilih karena dinilai sangat cocok untuk membangun model klasifikasi berbasis machine learning, khususnya dalam mendiagnosis penyakit liver[12]. CRISP-DM terdiri atas enam tahapan utama, yang memberikan panduan lengkap mulai dari pemahaman bisnis hingga evaluasi model. Alur metodologi penelitian ditunjukkan pada Gambar 1.



**Gambar 1. Tahapan Penelitian dengan Pendekatan CRISP-DM**

Pada tahap *business understanding*, ditetapkan tujuan utama penelitian yaitu melakukan pengelompokan produk UMKM konveksi berdasarkan kinerja penjualannya. Hasil pengelompokan diharapkan dapat dimanfaatkan oleh pelaku UMKM untuk menyusun strategi pemasaran yang lebih tepat sasaran serta mendukung pengelolaan persediaan agar lebih efisien. Dengan demikian, proses pengambilan keputusan tidak lagi hanya didasarkan pada intuisi, tetapi juga pada data yang objektif.

Tahap *data understanding* dilakukan analisis awal terhadap dataset untuk mendapatkan gambaran menyeluruh mengenai karakteristik data. Proses eksplorasi dilakukan dengan menghitung ukuran deskriptif seperti nilai rata-rata, nilai minimum, dan nilai maksimum dari tiap variabel. Selain itu, distribusi data juga diperiksa melalui visualisasi sederhana guna mendeteksi kecenderungan maupun potensi anomali. Hasil eksplorasi ini sangat penting karena akan menjadi dasar dalam proses pemodelan yang lebih lanjut.

Tahap *data preparation* meliputi dataset dibersihkan untuk mengatasi adanya nilai yang hilang (*missing value*) maupun ketidakkonsistenan yang mungkin muncul. Kemudian dilakukan pemilihan variabel inti yang relevan, yaitu *Jumlah Terjual*, *Harga Satuan*, dan *Omzet*. Agar variabel tersebut dapat dianalisis secara seimbang, dilakukan normalisasi menggunakan *StandardScaler*. Proses normalisasi ini diperlukan karena omzet memiliki skala nilai yang jauh

lebih besar dibandingkan dengan jumlah terjual atau harga satuan, sehingga tanpa penyesuaian dapat mendominasi hasil pengelompokan[13].

Tahap berikutnya adalah *modeling*, yang dilakukan dengan menggunakan algoritma K-Means Clustering. Algoritma ini dipilih karena mampu bekerja secara efisien dalam menangani dataset berukuran relatif besar dan menghasilkan pengelompokan yang mudah diinterpretasikan[14]. Proses pemodelan dijalankan dengan mencoba beberapa variasi jumlah cluster, mulai dari  $k = 2$  hingga  $k = 10$ , untuk memperoleh struktur pengelompokan yang paling sesuai dengan data. Parameter *random\_state* ditetapkan agar hasil dapat direplikasi dengan konsisten, sementara *n\_init* digunakan untuk meminimalkan pengaruh dari pemilihan centroid awal. Secara matematis, algoritma K-Means bertujuan meminimalkan jarak antara data dengan pusat cluster (*centroid*)[15], sebagaimana ditunjukkan pada Persamaan 1.

$$J = \sum_{i=1}^k \sum_{x \in C_i} ||x - \mu_i||^2 \quad \dots\dots\dots(1)$$

Tahap *evaluation* dilakukan untuk menilai kualitas hasil pengelompokan. Evaluasi dilakukan dengan menggunakan metrik Silhouette Score, yaitu ukuran yang membandingkan kedekatan data terhadap cluster tempat ia berada dengan cluster lainnya. Nilai silhouette berkisar antara -1 hingga 1, di mana semakin mendekati 1 menunjukkan bahwa pemisahan antar cluster semakin baik, sedangkan nilai mendekati 0 mengindikasikan adanya tumpang tindih antar cluster[16].

Tahap terakhir adalah *deployment*, Pada tahap ini, hasil analisis disajikan dalam bentuk tabel ringkasan serta berbagai visualisasi. Visualisasi yang digunakan mencakup scatterplot dua dimensi untuk memperlihatkan distribusi data antar cluster, plot tiga dimensi untuk menggambarkan interaksi antarvariabel, serta heatmap untuk menampilkan perbedaan nilai rata-rata pada setiap cluster. Seluruh visualisasi tersebut akan ditampilkan pada bagian Hasil dan Pembahasan sebagai pendukung interpretasi pola yang terbentuk.

Metodologi yang diterapkan dalam penelitian ini telah dirancang untuk memberikan hasil analisis yang sistematis, namun tetap memiliki beberapa keterbatasan. Batasan utama penelitian ini adalah jumlah variabel yang relatif terbatas serta penggunaan hanya satu jenis metrik evaluasi. Oleh karena itu, hasil yang diperoleh perlu ditafsirkan secara hati-hati dengan mempertimbangkan keterbatasan tersebut.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Hasil

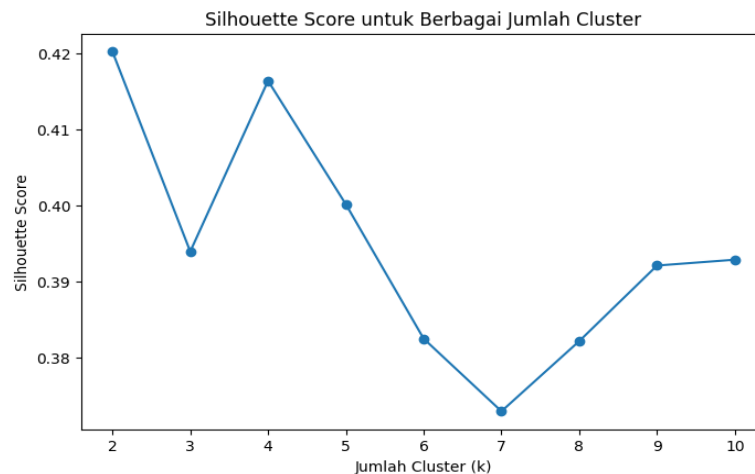
Dataset penelitian terdiri atas 360 catatan penjualan UMKM konveksi selama periode 2021–2023, dengan variabel utama *Jumlah Terjual*, *Harga Satuan*, dan *Omzet*. Statistik deskriptif dari ketiga variabel tersebut ditampilkan pada Tabel 2.

**Tabel 2. Statistik Deskriptif Variabel Penelitian**

| Variabel       | Minimum   | Maksimum    | Rata-rata  | Standar Deviasi |
|----------------|-----------|-------------|------------|-----------------|
| Jumlah Terjual | 22        | 994         | 514,36     | 280,08          |
| Harga Satuan   | 50.106    | 249.551     | 150.979,02 | 59.447,19       |
| Omzet          | 1.852.470 | 240.096.100 | 78.812.390 | 55.741.820      |

Berdasarkan Tabel 2, terlihat bahwa omzet produk bervariasi cukup besar, ditunjukkan dengan nilai deviasi standar yang tinggi. Hal ini menunjukkan adanya produk-produk unggulan yang berkontribusi jauh lebih besar dibandingkan produk lainnya.

Untuk menentukan jumlah cluster yang paling sesuai, dilakukan perhitungan Silhouette Score pada rentang  $k = 2$  hingga  $k = 10$ . Hasil pengujian ditunjukkan pada Gambar 2.



**Gambar 2. Silhouette Score untuk berbagai jumlah cluster**

Dari Gambar 2, terlihat bahwa nilai tertinggi diperoleh pada  $k = 2$  dengan skor 0,4202. Dengan demikian, jumlah cluster yang digunakan dalam analisis adalah dua cluster. Rincian setiap cluster ditampilkan pada tabel 3 berikut.

**Tabel 3. Silhouette Score setiap Cluster**

| Cluster | Silhouette Score |
|---------|------------------|
| 2       | 0.4202           |
| 3       | 0.3939           |
| 4       | 0.4164           |
| 5       | 0.4001           |
| 6       | 0.3824           |
| 7       | 0.3729           |
| 8       | 0.3821           |
| 9       | 0.3921           |
| 10      | 0.3928           |

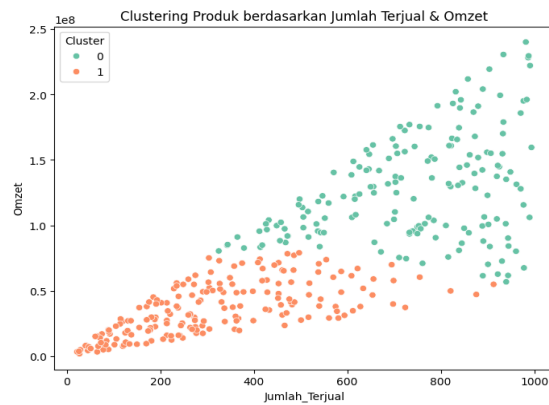
Setelah proses clustering, diperoleh ringkasan karakteristik tiap cluster sebagaimana ditampilkan pada Tabel 4.

**Tabel 4. Rata-rata Variabel pada Tiap Cluster**

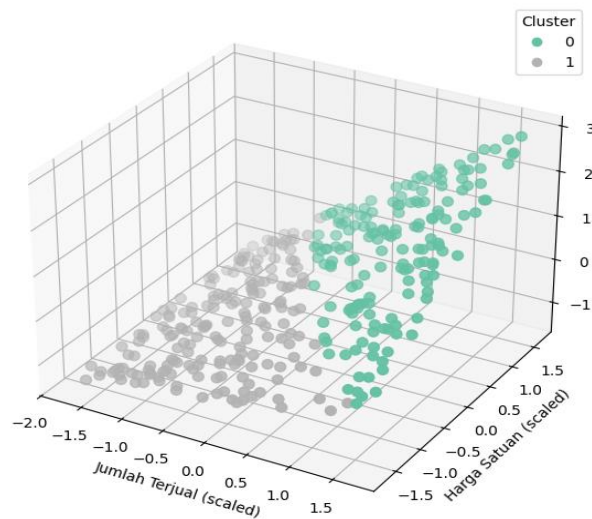
| Cluster | Jumlah Terjual (Rata-rata) | Harga Satuan (Rata-rata) | Omzet (Rata-rata) |
|---------|----------------------------|--------------------------|-------------------|
| 0       | 742,07                     | 178.190,04               | 128.102.800       |
| 1       | 315,11                     | 127.169,38               | 35.683.260        |

Dari Tabel 4 terlihat bahwa Cluster 0 memiliki rata-rata jumlah terjual lebih tinggi, harga satuan relatif mahal, serta omzet yang jauh lebih besar. Sebaliknya, Cluster 1 memiliki jumlah terjual lebih rendah, harga lebih murah, dan omzet yang juga lebih kecil.

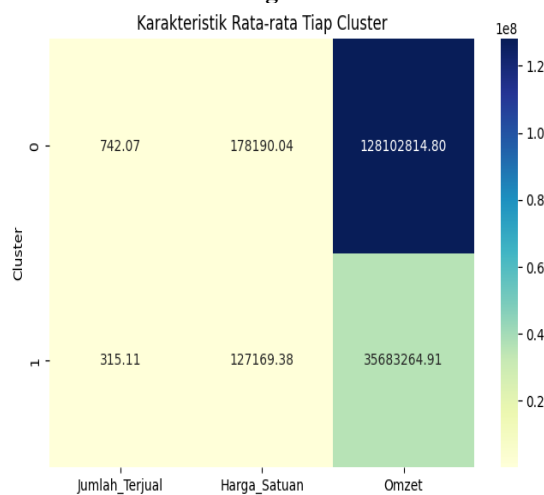
Visualisasi hasil clustering ditampilkan pada beberapa bentuk gambar. Scatterplot pada Gambar 3 memperlihatkan distribusi data berdasarkan jumlah terjual dan omzet. Selanjutnya, visualisasi tiga dimensi pada Gambar 4 menunjukkan keterkaitan antarvariabel yang membentuk cluster. Perbandingan rata-rata tiap variabel antar cluster divisualisasikan melalui heatmap pada Gambar 5.



**Gambar 3. Scatterplot hasil clustering berdasarkan jumlah terjual dan omzet**  
3D Visualization of Clusters



**Gambar 4. Visualisasi tiga dimensi hasil clustering**



**Gambar 5. Heatmap karakteristik rata-rata tiap cluster**

### 3.2 Pembahasan

Hasil clustering menunjukkan bahwa produk UMKM konveksi terbagi ke dalam dua kelompok dengan karakteristik yang berbeda. Cluster 0 terdiri atas produk dengan jumlah terjual tinggi, harga relatif lebih mahal, dan omzet yang jauh lebih besar. Produk dalam kelompok ini dapat dikategorikan sebagai produk premium, karena meskipun jumlahnya tidak mendominasi, kontribusi terhadap pendapatan perusahaan sangat signifikan. Sebaliknya, Cluster 1 berisi produk dengan jumlah terjual rendah, harga lebih murah, serta omzet yang relatif kecil. Produk dalam kelompok ini dapat disebut sebagai produk reguler, yang berperan menjaga variasi penjualan dan aksesibilitas pasar.

Interpretasi tersebut sesuai dengan konsep segmentasi pasar menurut Kotler [5], di mana pemetaan produk ke dalam kelompok berbeda memungkinkan strategi pemasaran yang lebih efektif. Produk premium dapat diposisikan untuk meningkatkan citra merek, memberikan layanan bernilai tambah, dan mempertahankan loyalitas pelanggan. Sementara itu, produk reguler dapat diarahkan untuk strategi penetrasi pasar, seperti bundling, promosi diskon, atau paket hemat guna menjangkau konsumen yang lebih luas.

Temuan ini juga konsisten dengan prinsip Pareto (80/20 rule), yang menyatakan bahwa sebagian kecil produk biasanya menyumbang sebagian besar pendapatan. Dalam konteks penelitian ini, produk premium pada Cluster 0 berfungsi sebagai backbone omzet, sementara produk reguler pada Cluster 1 berkontribusi dalam menjaga volume transaksi. Dengan demikian, kedua kelompok produk memiliki peran yang saling melengkapi dalam mendukung keberlangsungan usaha UMKM.

Jika dibandingkan dengan penelitian terdahulu, hasil ini serupa dengan studi sebelumnya pada UMKM ritel yang menemukan perbedaan mencolok antara produk beromzet tinggi dan rendah. Namun, penelitian ini memiliki keunggulan karena menggunakan kerangka CRISP-DM, sehingga proses analisis berjalan lebih sistematis mulai dari tahap pemahaman bisnis hingga penerapan hasil. Pendekatan ini membuat penelitian lebih mudah direplikasi oleh pelaku UMKM lain yang ingin menerapkan teknik clustering pada data penjualan mereka.

Meskipun hasil yang diperoleh bermanfaat, terdapat beberapa keterbatasan yang perlu dicermati. Nilai Silhouette Score sebesar 0,42 menunjukkan bahwa kualitas cluster masih belum optimal, karena ada kemungkinan beberapa produk memiliki karakteristik yang mendekati lebih dari satu cluster. Selain itu, variabel penelitian masih terbatas pada jumlah terjual, harga, dan omzet, sehingga faktor-faktor eksternal seperti promosi, tren musiman, maupun saluran distribusi belum tercakup. Hal ini membuka peluang penelitian selanjutnya untuk menambahkan variabel lain serta membandingkan hasil dengan algoritma clustering yang berbeda, misalnya DBSCAN atau Hierarchical Clustering.

Dari sisi praktis, hasil penelitian ini memberikan panduan nyata bagi UMKM konveksi. Produk premium dapat difokuskan pada peningkatan kualitas dan nilai merek, serta diarahkan pada segmen konsumen menengah ke atas. Sebaliknya, produk reguler dapat dikelola dengan strategi harga kompetitif, promosi intensif, dan perluasan pasar untuk menjaga arus kas dan mempertahankan basis konsumen. Dengan mengombinasikan kedua strategi ini, UMKM dapat meningkatkan daya saing sekaligus memperkuat ketahanan usaha dalam menghadapi dinamika pasar yang kompetitif.

#### **4. KESIMPULAN DAN SARAN**

Penelitian ini telah menerapkan metode *K-Means Clustering* dengan pendekatan CRISP-DM untuk menganalisis data penjualan pada UMKM konveksi. Berdasarkan hasil evaluasi, diperoleh jumlah cluster optimal sebanyak dua. Cluster pertama (Cluster 0) merepresentasikan produk premium dengan jumlah penjualan tinggi, harga satuan relatif mahal, dan omzet yang besar, sedangkan cluster kedua (Cluster 1) merepresentasikan produk reguler dengan jumlah penjualan lebih rendah, harga lebih murah, dan omzet kecil. Temuan ini menunjukkan bahwa metode *K-*

*Means Clustering* mampu mengidentifikasi pola segmentasi produk yang jelas dan dapat digunakan sebagai dasar pengambilan keputusan bisnis. Segmentasi tersebut memberikan gambaran yang lebih terstruktur mengenai perbedaan performa antar produk, sehingga UMKM dapat merancang strategi pemasaran dan pengelolaan stok yang lebih tepat sasaran.

Hasil penelitian ini juga memberikan beberapa implikasi praktis. Produk yang masuk dalam cluster premium sebaiknya dijadikan fokus utama strategi bisnis dengan menjaga kualitas, mempertahankan harga premium, serta memperkuat promosi untuk meningkatkan loyalitas pelanggan. Sementara itu, produk dalam cluster reguler tetap perlu dikelola melalui strategi efisiensi biaya produksi, inovasi desain, serta penetapan harga yang lebih kompetitif agar mampu menjangkau segmen konsumen yang lebih luas. Dengan cara ini, produk reguler dapat berperan sebagai produk pendukung atau *entry-level product* yang membantu memperluas pasar sekaligus mendukung produk premium, misalnya melalui strategi bundling.

Selain kontribusi praktis bagi UMKM, penelitian ini juga membuka peluang untuk pengembangan lebih lanjut. Pada penelitian mendatang, disarankan untuk menambahkan variabel lain seperti faktor musiman, wilayah pemasaran, atau strategi promosi sehingga hasil segmentasi dapat lebih komprehensif. Dengan demikian, penerapan metode *K-Means Clustering* tidak hanya efektif dalam mengelompokkan data penjualan, tetapi juga memberikan dasar yang kuat untuk merumuskan strategi bisnis berbasis data yang dapat meningkatkan daya saing UMKM konveksi.

## 5. DAFTAR RUJUKAN

- [1] R. Indonesia, “Undang-Undang Republik Indonesia Nomor 20 Tahun 2008 tentang Usaha Mikro, Kecil, dan Menengah,” 2008. [Online]. Available: [https://peraturan.bpk.go.id/Details/39653/uu-no-20-tahun-2008?utm\\_source=chatgpt.com](https://peraturan.bpk.go.id/Details/39653/uu-no-20-tahun-2008?utm_source=chatgpt.com)
- [2] S. Ramadani, D. A. Ramadhani, M. Ikrom, and L. M. Harahap, “Peran Strategis UMKM dalam Mendorong Pertumbuhan Ekonomi Berkelanjutan di Indonesia,” *J. Ekon. Bisnis dan Manaj.*, vol. 4, no. 1, pp. 158–166, 2025, doi: 10.58192/ebismen.v4i1.3183.
- [3] E. Dewata, Y. Sari, and H. Jauhari, “Penyusunan Laporan Keuangan Terkomputerisasi Berdasarkan SAK EMKM Pada UMKM Konveksi,” *Interv. Komunitas*, vol. 2, no. 1, pp. 11–16, 2020, doi: 10.32546/ik.v2i1.676.
- [4] O. Abdul-Azeez, A. O. Ihechere, and C. Idemudia, “Enhancing business performance: The role of data-driven analytics in strategic decision-making,” *Int. J. Manag. Entrep. Res.*, vol. 6, no. 7, pp. 2066–2081, 2024, doi: 10.51594/ijmer.v6i7.1257.
- [5] X. Wei, Z. Zhang, H. Huang, and Y. Zhou, “An overview on deep clustering,” *Neurocomputing*, vol. 590, p. 127761, Jul. 2024, doi: 10.1016/j.neucom.2024.127761.
- [6] A. Maulana, A. Ristyan, A. R. Ndun, and A. Ristyan, “Prediksi Volume Sampah Perkotaan Berbasis Data Spasial Menggunakan Random Forest di DKI Jakarta,” *Pros. SEMNAS ...*, vol. 9, pp. 1667–1672, 2025, [Online]. Available: <https://proceeding.unpkediri.ac.id/index.php/inotek/article/view/7513%0Ahttps://proceeding.unpkediri.ac.id/index.php/inotek/article/download/7513/5068>
- [7] E. Daniati and H. Utama, “Clustering K Means for Criteria Weighting With Improvement Result of Alternative Decisions Using SAW and TOPSIS,” in 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), IEEE, Nov. 2019, pp. 73–78. doi: 10.1109/ICITISEE48480.2019.9003858.
- [8] W. Nursahid, B. I. Nugroho, and S. Syefudin, “Optimalisasi Preprocessing Data Menggunakan Pendekatan CRISP-DM untuk Meningkatkan Kualitas Klasifikasi Penyakit Jantung,” *J. Artif. Intell. Digit. Bus.*, vol. 4, no. 3, pp. 3621–3626, 2025, doi: <https://doi.org/10.31004/riggs.v4i3.2514>.
- [9] R. Siagian, P. S. Pahala Sirait, and A. Halima, “E-Commerce Customer Segmentation Using K-Means Algorithm and Length, Recency, Frequency, Monetary Model,” *J.*

- INFORMATICS Telecommun. Eng., vol. 5, no. 1, pp. 21–30, Jul. 2021, doi: 10.31289/jite.v5i1.5182.
- [10] I. Yunita, P. R. Ali, M. A. Kartawidjaja, and R. Sukwadi, “Segmentasi Pelanggan Menggunakan K-Means Clustering: Menganalisis Metrik RFM untuk Strategi Pemasaran,” *J. Media Tek. dan Sist. Ind.*, vol. 9, no. 1, p. 58, 2025, doi: 10.35194/jmtsi.v9i1.4452.
- [11] O. D. Kurnia et al., “Analisis Perbandingan Algoritma Naïve Bayes dengan K-Nearest Neighbor (KNN) Pada Dataset Mobile Price Classification Penulis Korespondensi: Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi) 1174,” *Pros. SEMNAS INOTEK (Seminar Nas. Inov. Teknol.*, vol. 8, pp. 2549–7952, 2024, [Online]. Available: <https://proceeding.unpkediri.ac.id/index.php/inotek/article/download/5053/3475/19073>
- [12] C. Azzaria, E. Daniati, and A. Ristyan, “Peningkatan Akurasi Deteksi Liver Disease melalui Hyperparameter Tuning pada Algoritma Random Forest,” *Indones. J. Comput. Sci. Res.*, vol. 4, no. 2, pp. 139–147, 2025, [Online]. Available: <https://subset.id/index.php/IJCSR>
- [13] A. Harmain, P. Paiman, H. Kurniawan, K. Kusri, and D. Maulina, “NORMALISASI DATA UNTUK EFISIENSI K-MEANS PADA PENGELOMPOKAN WILAYAH BERPOTENSI KEBAKARAN HUTAN DAN LAHAN BERDASARKAN SEBARAN TITIK PANAS,” *Tek. Teknol. Inf. dan Multimed.*, vol. 2, no. 2, pp. 83–89, Jan. 2022, doi: 10.46764/teknimedia.v2i2.49.
- [14] N. Salsabila, K. Aulisari, and H. Z. Zahro, “Penerapan Algoritma K-Means Untuk Klasterisasi Produktivitas Tanaman Jahe,” *Infotek J. Inform. dan Teknol.*, vol. 8, no. 1, pp. 228–238, Jan. 2025, doi: 10.29408/jit.v8i1.28195.
- [15] M. Z. Nasution and M. S. Hasibuan, “Pendekatan Initial Centroid Search Untuk Meningkatkan Efisiensi Iterasi Klustering K-Means,” *Techno.Com*, vol. 19, no. 4, pp. 341–352, Nov. 2020, doi: 10.33633/tc.v19i4.3875.
- [16] R. Hidayati, A. Zubair, A. H. Pratama, and L. Indana, “Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering,” *Techno.Com*, vol. 20, no. 2, pp. 186–197, May 2021, doi: 10.33633/tc.v20i2.4556.