

Clustering Produk UMKM Berdasarkan Penjualan Online Menggunakan K-Means

Affandi Catur Pamungkas¹, Didan Wakhid Nur Sakti^{2*}, Fiza Trisna Aqilla³, Kurniadi Miza⁴, Aprilisa Arum Sari⁵

¹Teknik Informatika
Universitas Duta Bangsa
1220103141@mhs.udb.ac.id

²Teknik Informatika
Universitas Duta Bangsa
2*220103157@mhs.udb.ac.id

³ Teknik Informatika
Universitas Duta Bangsa
3220103152@mhs.udb.ac.id

⁴Teknik Informatika
Universitas Duta Bangsa
4220103148@mhs.udb.ac.id

⁵Teknik Informatika
Universitas Duta Bangsa
5aprilisa_arumsari@udb.ac.id

Abstrak— UMKM memiliki peran penting dalam perekonomian Indonesia, terutama dalam meningkatkan lapangan kerja dan Produk Domestik Bruto (PDB). Dengan berkembangnya digitalisasi, banyak pelaku UMKM memasarkan produknya secara online. Namun, banyaknya data transaksi yang dihasilkan memerlukan analisis untuk mengoptimalkan strategi penjualan. Penelitian ini bertujuan untuk mengelompokkan produk UMKM berdasarkan karakteristik penjualan online menggunakan algoritma K-Means. Dataset yang digunakan diperoleh dari E-Commerce Shopee dan mencakup 30 produk dengan fitur harga, jumlah terjual, dan rating. Data diolah melalui tahap normalisasi Min-Max dan proses clustering dengan jumlah cluster sebanyak tiga. Hasil penelitian menunjukkan bahwa algoritma K-Means mampu mengelompokkan produk ke dalam tiga kluster: produk dengan penjualan tinggi, sedang, dan rendah (outlier). Kluster ini memberikan gambaran performa penjualan dan membantu UMKM dalam merumuskan strategi yang lebih tepat, seperti fokus pada produk unggulan atau evaluasi produk berpenjualan rendah. Untuk menilai kualitas pengelompokan, dilakukan evaluasi menggunakan metode Silhouette Coefficient yang menghasilkan nilai rata-rata sebesar 0,443. Nilai ini menunjukkan bahwa hasil clustering sudah cukup baik dan mampu memisahkan data dengan cukup jelas antar kelompok. Kesimpulan dari penelitian ini menunjukkan bahwa metode clustering dengan K-Means dapat digunakan sebagai alat analisis efektif untuk mendukung pengambilan keputusan bisnis UMKM secara digital. Dengan pendekatan ini, UMKM diharapkan dapat meningkatkan daya saing dan keberlanjutan usahanya di era ekonomi digital.

Kata kunci— UMKM, K-Means, Clustering, Penjualan Online, Data Mining

Abstract— Micro, Small, and Medium Enterprises (MSMEs) play a vital role in Indonesia's economy, particularly in employment and Gross Domestic Product (GDP) contribution. With the rise of digitalization, many MSME actors are selling their products online. However, the increasing volume of transaction data requires proper analysis to optimize their sales strategies. This study aims to cluster MSME products based on online sales characteristics using the K-Means algorithm. The dataset used was obtained from Shopee E-Commerce, consisting of 30 products with features including price, number of units sold, and ratings. The data was processed through Min-Max normalization and clustered into three groups. The results show that the K-Means algorithm can effectively group products into three clusters: high, medium, and low (outlier) sales performance. These clusters provide insights into product performance and help MSMEs develop more targeted strategies, such as focusing on high-performing products or evaluating underperforming ones. To assess the clustering performance, a Silhouette Coefficient evaluation was conducted, resulting in an average score of 0.443. This value indicates that the clustering has been performed reasonably well and that data points are adequately separated between groups. The study concludes that clustering using the K-Means algorithm is an effective analytical tool for supporting MSME business decision-making in the digital era. Through this approach, MSMEs are expected to enhance competitiveness and sustainability in a digital economy.

Keywords— MSME, K-Means, Clustering, Online Sales, Data Mining

I. PENDAHULUAN

Usaha Mikro, Kecil, dan Menengah (UMKM) memiliki peran strategis dalam mendorong pertumbuhan ekonomi Indonesia, terutama melalui kontribusinya terhadap Produk Domestik Bruto (PDB) dan penciptaan lapangan kerja.

Kemajuan teknologi digital turut mendorong UMKM untuk memanfaatkan platform penjualan *online*, sehingga mereka dapat menjangkau pasar yang lebih luas dan meningkatkan potensi pertumbuhan usaha. Meski demikian, aktivitas digital yang menghasilkan beragam data penjualan belum sepenuhnya dimanfaatkan

secara optimal. Banyak UMKM masih mengalami kendala dalam menganalisis pola perilaku konsumen serta dalam mengelompokkan produk secara efektif [9]. Oleh sebab itu, dibutuhkan strategi pemasaran yang lebih tepat sasaran serta pengelolaan inventaris yang efisien guna membantu UMKM dalam mengoptimalkan pemanfaatan data dan mendukung pengambilan keputusan yang lebih baik [1].

UMKM sering menghadapi tantangan dalam mengelola stok secara optimal, menentukan strategi promosi yang tepat, serta mengenali produk-produk unggulan yang paling diminati pasar apabila mereka belum memiliki pemahaman yang mendalam terhadap karakteristik penjualan masing-masing produk. Seiring meningkatnya ragam produk dan volume transaksi, pendekatan manual untuk mengidentifikasi pola penjualan menjadi semakin tidak efisien. Oleh karena itu, diperlukan penerapan teknik analisis data yang dapat secara otomatis mengelompokkan produk berdasarkan kesamaan karakteristik penjualannya [10].

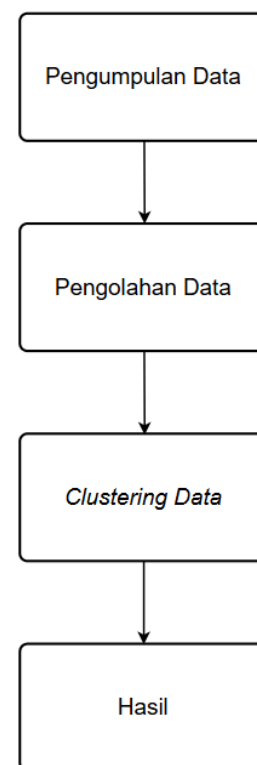
Penelitian ini bertujuan untuk menerapkan algoritma *K-Means* dalam mengelompokkan produk-produk UMKM berdasarkan data penjualan online. Algoritma *K-Means* dipilih karena kemampuannya dalam menangani data berukuran besar serta mengelompokkan data berdasarkan kemiripan fitur, khususnya pada aspek metrik penjualan. Dengan terbentuknya kluster produk, UMKM diharapkan dapat memperoleh pemahaman yang lebih mendalam terhadap performa tiap kelompok produk, mengidentifikasi produk dengan potensi penjualan tinggi, serta merumuskan strategi bisnis yang lebih adaptif dan kompetitif di era digital. Hasil dari penelitian ini diharapkan dapat menjadi acuan praktis bagi UMKM dalam meningkatkan efektivitas penjualan secara daring [2][3].

II. METODOLOGI PENELITIAN

Penelitian ini merupakan studi kuantitatif eksploratif yang bertujuan untuk mengelompokkan produk-produk UMKM

Berdasarkan karakteristik penjualan secara online, penelitian ini menggunakan pendekatan data mining dengan metode pembelajaran tanpa supervisi (*unsupervised learning*), khususnya melalui penerapan algoritma *K-Means Clustering*. Pendekatan ini memungkinkan pengelompokan produk secara otomatis berdasarkan kesamaan pola penjualannya, tanpa memerlukan label data sebelumnya, sehingga cocok untuk membantu UMKM dalam menganalisis dan memahami segmentasi produk mereka secara lebih efektif [4].

Penelitian ini mengadopsi pendekatan analisis kluster dengan tujuan untuk mengelompokkan sejumlah UMKM yang beroperasi di platform E-Commerce Shopee. Data primer dikumpulkan secara langsung dari para pelaku UMKM yang berjualan di Shopee, kemudian diolah menjadi informasi yang bernilai guna memberikan wawasan yang relevan dan mendukung pengambilan keputusan bisnis yang lebih tepat sasaran [5]. Penelitian ini terdiri dari empat tahap, dijelaskan pada gambar 1 sebagai berikut:



Gambar 1. Tahapan Penelitian.

Adapun penjelasan tahapan dalam penelitian.

1. Pengumpulan data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diambil dari *Dataset E-Commerce Shopee* yang tersedia di platform *Kaggle*. Setiap entri dalam dataset tersebut merepresentasikan satu produk yang dijual oleh UMKM sehingga memungkinkan analisis karakteristik penjualan pada tingkat produk secara lebih terperinci, dengan fitur-fitur sebagai berikut:

- a) Nama Produk
- b) Jumlah Terjual
- c) Harga
- d) Jumlah Rating

Minimal data yang digunakan dalam penelitian ini adalah 30 produk agar clustering lebih *representative*

2. Pengolahan data

Tahapan pengolahan data dalam penelitian ini dimulai dengan menyiapkan data penjualan produk UMKM yang akan dianalisis. Selanjutnya, dilakukan penentuan jumlah kluster yang akan digunakan dalam proses pengelompokan, yaitu sebanyak dua kluster. Setelah itu, dilakukan perhitungan terhadap masing-masing kriteria atau fitur yang digunakan sebagai dasar dalam proses klusterisasi, agar produk dapat dikelompokkan secara tepat berdasarkan karakteristik penjualannya [2].

3. Clustering Data

Clustering merupakan salah satu teknik utama dalam data mining yang digunakan untuk mengelompokkan data ke dalam beberapa subset atau kluster berdasarkan kemiripan karakteristik atau pola tertentu. Tujuan utama dari teknik ini adalah untuk mengidentifikasi struktur tersembunyi dalam suatu kumpulan data, sehingga dapat memberikan wawasan yang lebih mendalam mengenai pola, segmentasi, atau kategori yang muncul dari data tersebut [6]. Algoritma *K-Means* adalah salah satu algoritma *clustering* yang paling umum dan sederhana digunakan dalam analisis data. Algoritma ini bekerja dengan membagi data ke dalam

sejumlah kluster tertentu berdasarkan kemiripan fitur, sehingga setiap data dalam satu kluster memiliki karakteristik yang relatif serupa. Prosesnya melibatkan penentuan pusat kluster (*centroid*), pengelompokan data berdasarkan kedekatannya dengan *centroid*, dan pembaruan posisi *centroid* hingga diperoleh hasil klusterisasi yang optimal [7]. Algoritma ini bekerja dengan mengelompokkan titik-titik data ke dalam k kluster, di mana setiap titik data akan dimasukkan ke dalam kluster yang memiliki pusat (*centroid*) terdekat. Proses ini dilakukan dengan menghitung jarak antara setiap titik data dan *centroid* yang ada, kemudian memperbarui posisi *centroid* berdasarkan rata-rata posisi data dalam kluster tersebut. Langkah ini diulang secara iteratif hingga posisi *centroid* tidak lagi berubah secara signifikan, sehingga terbentuk kluster yang stabil dan representatif terhadap struktur data [4][8].

Langkah-langkah utama dalam algoritma *K-Means* adalah sebagai berikut:

- a. Inisialisasi: Pilih secara acak k titik awal sebagai pusat kelompok (*centroid*). Titik-titik ini bisa dipilih dari data atau secara acak.
- b. Pembentukan Kelompok: Setiap titik data dikelompokkan ke dalam kelompok yang memiliki pusat terdekat (*centroid*) dengannya. Ini dapat dihitung menggunakan jarak *Euclidean* atau metrik jarak lainnya dengan menggunakan persamaan yang ada pada gambar berikut ini.

$$d(x, c) = \sqrt{\sum_{i=1}^n (x_i - c_i)^2}$$

Gambar 2. Rumus Jarak *Euclident*

- c. Pembaruan Pusat: Hitung ulang pusat setiap kelompok dengan mengambil rata-rata dari semua titik data yang termasuk dalam kelompok tersebut.
- d. Iterasi: Langkah 2 dan 3 diulangi hingga tidak ada perubahan yang signifikan dalam posisi pusat

kelompok atau titik-titik data tidak berpindah kelompok.

- e. Penyelesaian: Proses dihentikan ketika kriteria penghentian telah terpenuhi, misalnya, ketika tidak ada lagi perubahan dalam posisi pusat kelompok atau ketika jumlah iterasi maksimum telah tercapai.

$$R_k = \frac{1}{N_k} (X_{1k} + X_{2k} + \dots + X_{nk})$$

Gambar 3. Rumus Pembaruan *Centroid*

4. Hasil

Model yang dihasilkan dari algoritma K-Means dapat dimanfaatkan untuk memprediksi keanggotaan klaster dari data baru, serta menentukan posisi pusat (*centroid*) dari tiap kelompok yang terbentuk. Untuk menilai kualitas model tersebut, dapat digunakan metrik evaluasi seperti *within-cluster sum of squares* (WCSS) dan *silhouette score*, yang berfungsi mengukur sejauh mana hasil pengelompokan mencerminkan struktur data yang sebenarnya [4].

5. Evaluasi

Tahap evaluasi bertujuan untuk menilai sejauh mana kualitas dari penelitian yang telah dilakukan, yang mencerminkan interpretasi terhadap hasil pemodelan yang diterapkan. Dalam penelitian ini, evaluasi hasil *clustering* dilakukan dengan membandingkan kinerja model menggunakan metode *Silhouette Coefficient* [10].

Rumus *Silhouette Coefficient*:

$$s(i) = \frac{b(i) - a(i)}{\max \{a(i), b(i)\}}$$

Keterangan:

s(i) : nilai *silhouette* untuk data ke-i

a(i) : rata-rata jarak antara data ke-i dengan semua data dalam satu cluster (*intra-cluster distance*)

b(i) : rata-rata jarak antara data ke-i dengan semua data pada cluster terdekat lainnya (*nearest-cluster distance*)

Nilai rata-rata *silhouette* dari seluruh data digunakan sebagai penilaian kualitas *clustering* secara keseluruhan.

III. HASIL DAN PEMBAHASAN

Data yang diolah dalam penelitian ini merupakan sampel yang diambil dari Dataset *E-Commerce* Shopee dari kaggle. Dataset ini terdiri dari atribut nama produk, stok, jumlah terjual, harga, jumlah rating. Data sampel yang akan diuji cobakan terdiri dari 30 data *E-Commerce* shopee. Contoh data yang digunakan dapat dilihat pada tabel 1.

Tabel 1. Data Sampel

| Nama | Harga | Jumlah Terjual | Rating |
|---------------------------------|--------------|----------------|--------|
| Tempe orek kering | Rp 22.000,00 | 5300,00 | 4,80 |
| Paket Goceng | Rp 61.000,00 | 10000,00 | 4,80 |
| MACJOS cemilan makaroni | Rp 9.000,00 | 4900,00 | 4,70 |
| Paray food | Rp 7.000,00 | 9400,00 | 4,60 |
| Kulit krispi ayam | Rp 15.000,00 | 10000,00 | 4,70 |
| Keripik pisang lumer | Rp 11.000,00 | 754,00 | 5,00 |
| Seumpia udang | Rp 25.500,00 | 37,00 | 4,90 |
| Snack taiko krezz | Rp 55.500,00 | 10000,00 | 4,60 |
| Keripik pisang coklat | Rp 9.000,00 | 617,00 | 4,80 |
| Keripik korean | Rp 54.000,00 | 526,00 | 4,90 |
| Saji ayam telur puyuh | Rp 6.500,00 | 6600,00 | 4,80 |
| QUINN of spicy ayam | Rp 41.500,00 | 9400,00 | 4,80 |
| Stroberi beku kering | Rp 79.000,00 | 131,00 | 4,80 |
| Keripik brownis | Rp 12.000,00 | 10000,00 | 4,50 |
| Basreng mentah | Rp 6.000,00 | 10000,00 | 4,80 |
| 1kg basreng | Rp 40.000,00 | 10000,00 | 4,70 |
| keripik kulit ayam kering tempe | Rp 12.000,00 | 10000,00 | 4,80 |
| sambal baby cumi | Rp 23.500,00 | 5300,00 | 4,70 |
| basreng kiloan | Rp 22.500,00 | 10000,00 | 4,60 |
| roti saku gajah gemoy | Rp 32.500,00 | 10000,00 | 4,90 |
| roti saku gajah gemoy | Rp 30.000,00 | 50,00 | 4,80 |
| baynana keripik pisang | Rp 19.500,00 | 5300,00 | 4,80 |
| oelek cumi cabe | Rp 30.000,00 | 10000,00 | 4,60 |
| basreng sultan pedas | Rp 17.500,00 | 1700,00 | 4,80 |
| mie gulung cikruh | Rp 79.500,00 | 1200,00 | 4,70 |

| | | | |
|-------------------------|--------------|----------|------|
| kerupuk seblak campur | Rp 23.000,00 | 10000,00 | 4,80 |
| kue nagasaki | Rp 35.700,00 | 10000,00 | 4,80 |
| cokelat cookies | Rp 21.000,00 | 2400,00 | 4,80 |
| si umang mie lidi | Rp 14.400,00 | 10000,00 | 4,90 |
| chipzdele keripik tempe | Rp 16.500,00 | 293,00 | 4,80 |

Setelah data produk UMKM yang mencakup nama produk, harga, jumlah penjualan, dan rating berhasil dikumpulkan, tahap berikutnya adalah proses pengolahan data. Langkah awal dalam proses ini adalah pembersihan data, di mana seluruh format yang masih mengandung simbol seperti “Rp” maupun tanda titik sebagai pemisah ribuan dikonversi ke dalam format numerik standar. Hasil data yang telah dibersihkan ditampilkan pada Tabel 2..

Tabel 2. Data Numerik

| Nama | Harga | Jumlah Terjual | Rating |
|-------------------------|----------|----------------|--------|
| Tempe orek kering | 22000,00 | 5300 | 4,8 |
| Paket Goceng | 61000,00 | 10000 | 4,8 |
| MACJOS cemilan makaroni | 9000,00 | 4900 | 4,7 |
| Parayu food | 7000,00 | 9400 | 4,6 |
| Kulit krispi ayam | 15000,00 | 10000 | 4,7 |
| Keripik pisang lumer | 11000,00 | 754 | 5 |
| Seumpia udang | 25500,00 | 37 | 4,9 |
| Snack taiko krezz | 55500,00 | 10000 | 4,6 |
| Keripik pisang coklat | 9000,00 | 617 | 4,8 |
| Keripik korean | 54000,00 | 526 | 4,9 |
| Saji ayam telur puyuh | 6500,00 | 6600 | 4,8 |
| QUINN of spicy ayam | 41500,00 | 9400 | 4,8 |
| Stroberi beku kering | 79000,00 | 131 | 4,8 |
| Keripik brownis | 12000,00 | 10000 | 4,5 |
| Basreng mentah | 6000,00 | 10000 | 4,8 |
| 1kg basreng | 40000,00 | 10000 | 4,7 |
| keripik kulit ayam | 12000,00 | 10000 | 4,8 |
| kering tempe | 23500,00 | 5300 | 4,7 |
| sambal baby cumi | 22500,00 | 10000 | 4,6 |

| | | | |
|-------------------------|----------|-------|-----|
| basreng kiloan | 32500,00 | 10000 | 4,9 |
| roti saku gajah gemoy | 30000,00 | 50 | 4,8 |
| baynana keripik pisang | 19500,00 | 5300 | 4,8 |
| oelek cumi cabe | 30000,00 | 10000 | 4,6 |
| basreng sultan pedas | 17500,00 | 1700 | 4,8 |
| mie gulung cikruh | 79500,00 | 1200 | 4,7 |
| kerupuk seblak campur | 23000,00 | 10000 | 4,8 |
| kue nagasaki | 35700,00 | 10000 | 4,8 |
| cokelat cookies | 21000,00 | 2400 | 4,8 |
| si umang mie lidi | 14400,00 | 10000 | 4,9 |
| chipzdele keripik tempe | 16500,00 | 293 | 4,8 |

Dataset yang digunakan dalam penelitian ini mencakup 30 produk UMKM, dengan masing-masing data memiliki atribut berupa Harga, Jumlah Terjual, dan Rating. Sebelum dilakukan proses clustering, data terlebih dahulu dinormalisasi menggunakan metode *Min-Max Scaling*. Teknik ini berfungsi untuk menyeragamkan skala antar fitur ke dalam rentang 0 hingga 1, sehingga tidak ada satu fitur pun yang mendominasi proses klusterisasi akibat perbedaan skala nilai.

Rumus normalisasi Min-Max:

$$x' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Contoh normalisasi pada fitur Harga:

- a) Minimum: 6.000
- b) Maksimum: 79.500
- c) Harga Tempe orek kering (22.000) dinormalisasi:

$$X' = \frac{22000 - 6000}{79500 - 6000} = \frac{16000}{73500} \approx 0.2177$$

Proses L-Means Clustering

Clustering dilakukan dengan jumlah cluster k = 3, mewakili tiga kategori performa produk:

- a) Cluster 0 → Penjualan tinggi
- b) Cluster 1 → Penjualan sedang
- c) Cluster 2 → Penjualan rendah atau outlier

Setelah semua dinormalisasi:

Centroid₀ = rata-rata dari 3 vektor

Langkah-langkah:

- a) Normalisasi fitur
- b) Inisialisasi 3 centroid secara acak
- c) Hitung jarak setiap data ke semua centroid menggunakan rumus Euclidean:

$$c_0 = \left(\frac{0.746 + 0.014 + 0.122}{3}, \frac{1 + 0.939 + 1}{3}, \frac{0.5 + 0.25 + 0.25}{3} \right) \approx (0.294, 0.98, 0.333)$$

$$d(x, c) = \sqrt{(x_1 - c_1)^2 + (x_2 - c_2)^2 + (x_3 - c_3)^2}$$

- d) Tentukan cluster berdasarkan jarak terdekat
- e) Update centroid :

$$C_k = \frac{1}{N_k} \sum_{i=1}^{N_k} x_i$$

- f) Ulangi langkah 3-5 hingga konvergen

Contoh Perhitungan Manual Update Centroid (Cluster 0). Tiga produk awal yang termasuk dalam Cluster 0 :

- a) Paket Goceng: (61000, 10000, 4.8)
- b) Parayu food: (7000, 9400, 4.7)
- c) Kulit krispi ayam: (15000, 10000, 4.7)

Langkah 1: Normalisasi ketiga data

- a) Kisaran Harga = 6000–79500
- b) Kisaran Jumlah Terjual = 37–10000
- c) Kisaran Rating = 4.6–5.0

Contoh (Paket Goceng):

- a) Harga: $(61000 - 6000) / (79500 - 6000) \approx 0.746$
- b) Terjual: $(10000 - 37) / (10000 - 37) = 1$
- c) Rating: $(4.8 - 4.6) / (5.0 - 4.6) = 0.5$

Tabel 3. Hasil pengelompokan (output K-Means)

| Cluster | Jumlah Produk | Karakteristik |
|---------|---------------|------------------------------------|
| 0 | 15 | Penjualan tinggi, harga bervariasi |
| 1 | 12 | Penjualan sedang, harga menengah |
| 2 | 3 | Penjualan sangat rendah (outlier) |

Contoh produk tiap cluster:

1. Cluster 0:
 - a) Paket Goceng (Rp61.000, 10.000 terjual)
 - b) Parayu Food (Rp7.000, 9.400 terjual)
 - c) Kulit Krispi Ayam (Rp15.000, 10.000 terjual)
2. Cluster 1:
 - a) Tempe Orek Kering (Rp22.000, 5.300 terjual)
 - b) MACJOS Cemilan M (Rp9.000, 4.900 terjual)
 - c) Keripik Pisang Lumer (Rp11.000, 754 terjual)
3. Cluster 2 (outlier):
 - a) Keripik Korean (Rp54.000, 526 terjual)
 - b) Stroberi Beku Kering (Rp79.000, 131 terjual)
 - c) Mie Gulung Cikruh (Rp79.500, 1.200 terjual)

Hasil *clustering* memberikan gambaran segmentasi produk berdasarkan performa :

- Klaster 0 merepresentasikan produk-produk dengan tingkat daya saing yang tinggi dan penjualan yang konsisten, meskipun memiliki variasi dalam hal harga.
- Klaster 1 terdiri dari produk-produk yang memiliki potensi untuk dikembangkan, ditandai dengan rating yang baik meskipun jumlah penjualannya masih tergolong sedang.
- Klaster 2 merupakan kategori outlier, yang mencakup produk-produk dengan harga relatif tinggi atau tingkat penjualan yang sangat rendah. Produk dalam klaster ini dapat menjadi prioritas untuk evaluasi lebih lanjut, seperti dari segi strategi promosi, penyesuaian harga, kemasan, maupun optimasi kata kunci SEO.

Contoh perhitungan *update* centroid (Klaster 1)

- Tahapan ini dilakukan setelah seluruh data berhasil dikelompokkan ke dalam klaster, kemudian dilakukan penghitungan ulang centroid untuk Klaster 1 berdasarkan nilai rata-rata dari seluruh data yang telah dinormalisasi dalam klaster tersebut.
- Kita ambil 3 contoh produk dari Cluster 1 berdasarkan hasil sebelumnya.

Tabel 1. Tabel Matrix Produk berdasarkan Rating

| Produk | Harga (Rp) | Jumlah Terjual | Rating |
|-------------------|------------|----------------|--------|
| MACJOS cemilan m | 9.000 | 4.900 | 4,7 |
| Tempe orek kering | 22.000 | 5.300 | 4,8 |
| Keripik pisang | 11.000 | 754 | 5,0 |

Langkah 1: Normalisasi data menggunakan *Min-Max Scaling*

Nilai Min dan Max dari dataset:

- Harga: min = 6.000, max = 79.500
- Jumlah Terjual: min = 37, max = 10.000
- Rating: min = 4,6, max = 5,0

Normalisasi :

1. Tempe orek kering

- Harga = $\frac{(22.000-6.000)}{79.500-6.000} \approx 0.2177$
- Terjual = $\frac{(5.300-37)}{(10.000-37)} \approx 0.5256$
- Rating = $\frac{(4.8-4.6)}{(5.0-4.6)} \approx 0.5$

1. MACJOS cemilan m

- Harga = $\frac{11.000-6.000}{79.500-6.000} \approx 0.068$
- Terjual = $\frac{(754-37)}{(10.000-37)} \approx 0.0719$
- Rating = $\frac{(5.0-4.8)}{(5.0-4.6)} \approx 1.0$

Langkah 2: Hitung centroid baru (rata-rata)

Centroid baru dihitung dari rata-rata masing-masing fitur :

$$\text{Centroid}_1 (\text{Harga}) = (0.2177 + 0.0408 + 0.068) / 3 \approx 0.1088$$

$$\text{Centroid}_1 (\text{Terjual}) = (0.5256 + 0.4834 + 0.0719) / 3 \approx 0.3603$$

$$\text{Centroid}_1 (\text{Rating}) = (0.5 + 0.25 + 1.0) / 3 \approx 0.5833$$

Jadi, centroid baru untuk Cluster 1:

$$\text{Centroid}_1 = (0.1088, 0.3603, 0.5833)$$

Artinya, pada iterasi berikutnya, titik pusat Cluster 1 akan berpindah ke koordinat ini, dan proses iterasi berlanjut dengan menghitung jarak seluruh data ke centroid terbaru.

Evaluasi model clustering menggunakan metode *Silhouette Coeficient* sebagai berikut :

$$s(i) = \frac{0.70 - 0.6915}{\max\{0.70, 0.6915\}}$$

$$= \frac{0.0085}{0.70} = 0.0121$$

kemudian hitung nilai rata-ratanya: *Silhouette Score*.

$$(\text{rata-rata seluruh } s(i)) = \sum s(i) / n = 0.443$$

IV. KESIMPULAN

Hasil dari penelitian ini memperlihatkan bahwa algoritma K-Means mampu mengelompokkan produk UMKM ke dalam tiga kluster berdasarkan kinerja penjualan secara online, yaitu kluster dengan tingkat penjualan tinggi, menengah, dan rendah. Hasil klusterisasi ini memberikan gambaran yang lebih terstruktur mengenai segmentasi produk berdasarkan variabel harga, jumlah penjualan, dan rating. Produk-produk yang berada dalam kluster penjualan tinggi dapat dijadikan fokus utama untuk dipertahankan serta dikembangkan lebih lanjut melalui strategi pemasaran yang tepat. Di sisi lain, produk yang termasuk dalam kluster menengah dan rendah dapat menjadi objek evaluasi guna menyusun strategi perbaikan, baik dari segi promosi, penyesuaian harga, maupun peningkatan kualitas produk.

Temuan dari penelitian ini membawa dampak yang berarti bagi pelaku UMKM dalam merumuskan kebijakan yang lebih terfokus untuk meningkatkan daya saing produk di pasar digital. Penelitian ini turut membuktikan bahwa metode clustering tidak hanya berfungsi dalam analisis data, tetapi juga dapat dimanfaatkan sebagai pendekatan strategis berbasis data dalam mendukung proses pengambilan keputusan bisnis. Oleh karena itu, tujuan utama dari penelitian ini yaitu membantu UMKM dalam

memahami kinerja produk melalui pengelompokan data penjualan berhasil dicapai dengan efektif.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Fakultas Ilmu Komputer Universitas Duta Bangsa Surakarta atas dukungan sarana dan prasarana yang telah diberikan selama proses penelitian ini berlangsung. Penghargaan juga disampaikan kepada dosen pembimbing yang telah memberikan arahan dan masukan yang konstruktif. Selain itu, ucapan terima kasih ditujukan kepada pihak yang telah menyediakan dataset *E-Commerce Shopee* melalui platform Kaggle, yang menjadi sumber utama dalam proses analisis data. Semua kontribusi tersebut sangat berarti dalam mendukung keberhasilan penyusunan jurnal ini.

REFERENSI

- [1] L. Hamim Annisa, D. Rusvinasari, and L. H. Annisa, "Segmentasi Pembelian Produk Menggunakan Algoritma K-Means Berdasarkan Clusterisasi pada pemilihan menu yang ada diUMKM Kuliner", doi: 10.30591/jpit.v9ix.xxx.
- [2] A. Yudistira and R. Andika, "Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering," *J. Artif. Intell. Technol. Inf.*, vol. 1, no. 1, pp. 20–28, Mar. 2023, doi: 10.58602/jaiti.v1i1.22.
- [3] W. Sudrajat, I. Cholid, and J. Petrus, "Wahyu Sudrajat et al, Penerapan Algoritma K-Means Untuk"
- [4] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," vol. 3, pp. 46–56, 2024.
- [5] D. Marcelina, A. Kurnia, and T. Tertaaavini, "Analisis Kluster Kinerja Usaha Kecil dan Menengah Menggunakan Algoritma K-Means Clustering," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 293–301, Nov. 2023, doi: 10.57152/malcom.v3i2.952.
- [6] T. Amalina, D. Bima, A. Pramana, and B. N. Sari, "Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food," *J. Ilm. Wahana Pendidik.*, vol. 8, no. 15, pp. 574–583, 2022, doi: 10.5281/zenodo.7052276.
- [7] E. Febrianty, L. Awalina, and W. I. Rahayu, "Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail Optimizing Marketing Strategies with Customer Segmentation Using K-Means Clustering on Online Retail Transactions," *J. Teknol. dan Inf.*, vol. 13, 2023, doi: 10.34010/jati.v13i2.
- [8] H. Al Rasyid, B. Firmansyah, K. Soebari, D. Satria, and Y. Kartika, "Prosiding Seminar Nasional Teknologi dan Sistem Informasi (SITASI) 2022 IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING UNTUK PENGELOMPOKAN PENJUALAN PRODUK PADA ONLINE SHOP TOKO GIZI IMPLEMENTATION OF K-MEANS CLUSTERING ALGORITHM FOR GROUPING PRODUCT SALES IN ONLINE SHOP TOKO GIZI." [Online]. Available: <http://sitasi.upnjatim.ac.id/%7C242>

- [9] T. Amalina, D. Bima, A. Pramana, and B. N. Sari, "Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food," *J. Ilm. Wahana Pendidik.*, vol. 8, no. 15, pp. 574–583, 2022, doi: 10.5281/zenodo.7052276.
- [10] N. Syifa and R. N. Fahmi, "Implementasi Metode K-Means Clustering dalam Analisis Persebaran UMKM di Jawa Barat," *JOINS (Journal Inf. Syst.*, vol. 6, no. 2, pp. 211–220, 2021, doi: 10.33633/joins.v6i2.5310.