

Clustering UMKM Berbasis Aktivitas E-Commerce Menggunakan K-Means untuk Strategi Pengembangan Usaha

Clustering Analysis of MSMEs Based on E-Commerce Activities Using the K-Means Algorithm for Business Development Strategies

Siska Mayasari Rambe^{*,1}, Muhammad Iqbal²

^{1,2}Magister Teknologi Informasi, Universitas Pembangunan Panca Budi, Sumatera Utara
Indonesia

*Penulis Korespondensi

Email: mayasarisiska5@gmail.com

Abstrak. Usaha Mikro, Kecil, dan Menengah (UMKM) memiliki peran penting dalam perekonomian Indonesia, khususnya dalam menciptakan lapangan kerja dan meningkatkan Produk Domestik Bruto (PDB). Perkembangan e-commerce mendorong UMKM untuk memanfaatkan platform digital dalam memasarkan produk sehingga menghasilkan data aktivitas penjualan yang dapat dianalisis lebih lanjut. Penelitian ini bertujuan untuk mengidentifikasi pola segmentasi UMKM berdasarkan aktivitas e-commerce menggunakan algoritma K-Means Clustering sebagai dasar penyusunan strategi pengembangan usaha berbasis data. Dataset yang digunakan merupakan data sintetis yang dibangun untuk mensimulasikan 50 UMKM berdasarkan distribusi aktivitas e-commerce yang umum ditemukan pada pelaku usaha digital. Data dibangkitkan menggunakan pendekatan random uniform distribution dengan rentang transaksi per bulan antara 50–2000 transaksi, revenue e-commerce antara Rp4.000.000–Rp180.000.000, engagement rate antara 0,10–0,90, serta jumlah platform antara 1–4 platform digital. Variabel revenue juga dirancang memiliki kecenderungan berkorelasi positif dengan jumlah transaksi agar lebih merepresentasikan kondisi aktivitas penjualan UMKM. Data dinormalisasi menggunakan metode Min-Max Scaling dan dikelompokkan ke dalam tiga kluster menggunakan algoritma K-Means. Hasil penelitian menunjukkan bahwa algoritma K-Means mampu mengelompokkan UMKM ke dalam tiga kategori utama, yaitu UMKM dengan performa tinggi, performa menengah, dan performa rendah berdasarkan aktivitas e-commerce. Evaluasi menggunakan metode Silhouette Coefficient menghasilkan nilai rata-rata sebesar 0,292, yang menunjukkan bahwa struktur kluster yang dihasilkan masih tergolong lemah dan masih terdapat tumpang tindih karakteristik antar kluster. Meskipun demikian, hasil clustering tetap mampu memberikan segmentasi awal UMKM yang dapat dimanfaatkan sebagai dasar penyusunan strategi pengembangan usaha berbasis data di era ekonomi digital.

Kata kunci: UMKM, K-Means, Clustering, E-commerce, Data Mining

Abstract. Micro, Small, and Medium Enterprises (MSMEs) play an important role in the Indonesian economy, particularly in generating employment opportunities and contributing to Gross Domestic Product (GDP). The rapid growth of e-commerce has encouraged MSMEs to adopt digital platforms for marketing activities, resulting in large amounts of sales activity data that can be further analyzed. This study aims to identify MSME segmentation patterns based on e-commerce activities using the K-Means Clustering algorithm as a foundation for developing data-driven business strategies. The dataset used in this study consists of synthetic data designed to simulate 50 MSMEs based on commonly observed e-commerce activity distributions among digital

business actors. The data were generated using a random uniform distribution approach, with monthly transactions ranging from 50–2000 transactions, e-commerce revenue ranging from IDR 4,000,000 to IDR 180,000,000, engagement rates between 0.10–0.90, and the number of digital platforms ranging from 1–4 platforms. In addition, the revenue variable was designed to have a positive correlation with the number of transactions to better reflect realistic MSME sales activities. The data were normalized using the Min-Max Scaling method and grouped into three clusters using the K-Means algorithm. The results show that the K-Means algorithm was able to classify MSMEs into three main categories, namely high-performance, medium-performance, and low-performance MSMEs based on their e-commerce activities. Evaluation using the Silhouette Coefficient method produced an average score of 0.292, indicating that the resulting cluster structure is still relatively weak and that some overlap between clusters remains. Nevertheless, the clustering results were still able to provide an initial segmentation of MSMEs that can be utilized as a basis for developing data-driven business development strategies in the digital economy era.

Keywords: MSMEs, K-Means, Clustering, E-Commerce, Data Mining

1. 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 (Lubis et al., 2023). Emajuan 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 (Afrizal et al., 2025). Banyak UMKM masih menggunakan platform e-commerce hanya sebagai sarana transaksi tanpa melakukan analisis lebih lanjut terhadap data aktivitas digital yang dihasilkan, seperti pola transaksi pelanggan, tingkat engagement, dan performa penjualan antar platform. Kondisi tersebut menyebabkan proses evaluasi performa usaha dan pengambilan keputusan bisnis masih dilakukan secara manual dan kurang efektif (Linarti et al., 2024).

Banyak UMKM masih mengalami kendala dalam menganalisis pola perilaku konsumen serta dalam melakukan segmentasi usaha secara efektif berbasis data aktivitas digital. Selain itu, UMKM juga sering menghadapi tantangan dalam menentukan strategi promosi yang tepat, mengenali tingkat performa usaha, dan memahami karakteristik aktivitas penjualan digital yang dimiliki. Seiring meningkatnya volume transaksi dan penggunaan berbagai platform e-commerce, pendekatan manual untuk mengidentifikasi pola aktivitas usaha menjadi semakin tidak efisien. Oleh karena itu, diperlukan penerapan teknik analisis data yang dapat secara otomatis mengelompokkan UMKM berdasarkan kesamaan karakteristik aktivitas digitalnya (Iskandar et al., 2023).

Salah satu metode yang dapat digunakan untuk mengatasi permasalahan tersebut adalah teknik clustering dalam data mining. Clustering merupakan teknik pengelompokan data berdasarkan tingkat kemiripan karakteristik tertentu sehingga data dalam satu kelompok memiliki karakteristik yang relatif serupa dibandingkan kelompok lainnya (Petrus et al., 2023). Algoritma K-Means merupakan salah satu metode clustering yang banyak digunakan karena mampu mengelompokkan data secara sederhana, efisien, dan efektif berdasarkan kedekatan fitur data (Mauliadi, 2022). Metode ini dinilai sesuai untuk melakukan segmentasi UMKM berdasarkan aktivitas e-commerce karena mampu mengolah data numerik dengan jumlah yang cukup besar.

Penelitian ini bertujuan untuk mengidentifikasi pola segmentasi UMKM berdasarkan aktivitas e-commerce menggunakan algoritma K-Means Clustering dengan atribut transaksi per bulan, revenue e-commerce, engagement rate, dan jumlah platform digital. Selain itu, penelitian ini juga memberikan kontribusi berupa analisis segmentasi UMKM berbasis data aktivitas digital yang dapat digunakan sebagai dasar dalam penyusunan strategi pengembangan usaha yang lebih tepat sasaran dan berbasis data. Dengan terbentuknya klaster UMKM, pelaku usaha diharapkan dapat memperoleh pemahaman yang lebih baik mengenai karakteristik performa usahanya sehingga dapat menentukan strategi bisnis yang lebih adaptif dan kompetitif di era digital (Mohammad Ferdiansyah & Umi Chotijah, 2024).

2. Metode

Penelitian ini menggunakan pendekatan data mining dengan metode pembelajaran tanpa supervisi (unsupervised learning), khususnya menggunakan algoritma K-Means Clustering untuk mengelompokkan UMKM berdasarkan karakteristik aktivitas e-commerce. Pendekatan ini dipilih karena mampu melakukan pengelompokan data secara otomatis berdasarkan tingkat kemiripan karakteristik tanpa memerlukan label data sebelumnya, sehingga sesuai untuk melakukan segmentasi UMKM berdasarkan aktivitas digitalnya (Wisanta & Marlim, 2021).

Penelitian ini mengadopsi pendekatan analisis clustering terhadap aktivitas e-commerce UMKM menggunakan dataset sintesis yang dibangun untuk mensimulasikan kondisi aktivitas digital 50 UMKM. Data sintesis digunakan karena penelitian ini berfokus pada simulasi pola segmentasi UMKM berbasis aktivitas e-commerce. Proses pembangkitan data dilakukan menggunakan pendekatan random uniform distribution agar menghasilkan variasi data yang menyerupai kondisi aktivitas UMKM pada platform digital (Bagustio et al., 2024).

Dataset dibangun menggunakan empat atribut utama, yaitu transaksi per bulan, revenue e-commerce, engagement rate, dan jumlah platform digital yang digunakan. Rentang nilai masing-masing variabel ditentukan berdasarkan karakteristik umum aktivitas UMKM pada platform e-commerce, yaitu:

- a. Transaksi per bulan: 50–2000 transaksi
- b. Revenue e-commerce: Rp4.000.000–Rp180.000.000
- c. Engagement rate: 0,10–0,90
- d. Jumlah platform digital: 1–4 platform

Selain itu, variabel revenue e-commerce dirancang memiliki kecenderungan berkorelasi positif dengan jumlah transaksi per bulan untuk menghasilkan simulasi data yang lebih realistis. Dengan demikian, semakin tinggi jumlah transaksi UMKM maka nilai revenue cenderung meningkat. Pendekatan ini dilakukan agar struktur data sintesis tetap mencerminkan pola aktivitas penjualan digital yang umum terjadi pada UMKM.

2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sintesis yang dihasilkan untuk merepresentasikan 50 UMKM. Setiap entri dalam dataset tersebut merepresentasikan satu UMKM sehingga memungkinkan analisis karakteristik aktivitas pada tingkat UMKM secara lebih terperinci, dengan fitur-fitur sebagai berikut:

- a. Transaksi per Bulan
- b. *Revenue E-commerce*

- c. *Engagement Rate*
- d. Jumlah Platform

Data yang digunakan dalam penelitian ini adalah 50 UMKM agar *Clustering* lebih representative.

2.2 Pengolahan Data

Tahapan pengolahan data dimulai dengan normalisasi data menggunakan metode Min-Max Scaling agar seluruh atribut memiliki rentang nilai yang sama, yaitu antara 0 hingga 1. Normalisasi dilakukan untuk menghindari dominasi atribut tertentu yang memiliki skala nilai lebih besar dalam proses perhitungan jarak pada algoritma K-Means (Suara et al., 2022).

Setelah proses normalisasi dilakukan, tahap berikutnya adalah menentukan jumlah kluster (k). Penentuan nilai k dilakukan menggunakan pendekatan Elbow Method dan evaluasi Silhouette Score untuk memperoleh jumlah kluster yang paling sesuai terhadap karakteristik dataset. Berdasarkan hasil pengujian, nilai k = 3 dipilih karena memberikan keseimbangan terbaik antara pemisahan kluster dan interpretasi segmentasi UMKM.

Selanjutnya, proses clustering dilakukan menggunakan algoritma K-Means dengan tahapan inisialisasi centroid awal, perhitungan jarak Euclidean, pengelompokan data ke kluster terdekat, serta pembaruan centroid secara iteratif hingga kondisi konvergen tercapai. Nilai kluster ditentukan berdasarkan jarak terdekat antara data UMKM dengan centroid masing-masing kluster, sehingga setiap UMKM akan masuk ke kelompok yang memiliki karakteristik paling mirip.

2.3 Clustering Data

Clustering merupakan salah satu metode dalam data mining yang berfungsi untuk mengelompokkan sekumpulan data ke dalam beberapa cluster berdasarkan tingkat kemiripan karakteristik yang dimiliki. Melalui proses ini, objek yang memiliki sifat serupa akan ditempatkan pada cluster yang sama, sedangkan objek yang memiliki karakteristik berbeda akan berada pada cluster yang lain (Petrus et al., 2023). Tujuan penerapan teknik *clustering* adalah untuk menemukan struktur atau hubungan yang tersembunyi di dalam sekumpulan data. Melalui proses ini, pola, segmentasi, maupun kelompok data yang memiliki karakteristik serupa dapat dikenali sehingga menghasilkan informasi yang lebih bermakna untuk proses analisis (Bagustio et al., 2024). 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 (Tajrin et al., 2025).

Metode ini mengelompokkan data ke dalam beberapa *cluster* berdasarkan tingkat kemiripan karakteristik yang dimiliki, sehingga anggota dalam cluster yang sama memiliki sifat yang relatif homogen. Tahapan prosesnya meliputi penentuan titik pusat cluster (*centroid*), pengalokasian setiap data ke centroid terdekat, kemudian dilakukan pembaruan posisi centroid secara berulang hingga diperoleh hasil pengelompokan yang stabil dan optimal (Afrizal et al., 2025). Algoritma K-Means mengelompokkan data ke dalam sejumlah k cluster berdasarkan tingkat kedekatan setiap objek terhadap titik pusat *cluster* (*centroid*). Setiap data akan menjadi anggota cluster yang memiliki nilai jarak paling kecil terhadap centroid dibandingkan dengan centroid lainnya (Ameliana et al., 2024). 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

(Hariyanto et al., 2025). Langkah-langkah utama dalam algoritma K-Means adalah sebagai berikut:

- inisialisasi: Pada tahap inisialisasi, ditentukan k centroid awal yang dipilih secara acak sebagai pusat masing-masing cluster. Centroid tersebut dapat berasal dari data yang ada maupun ditentukan melalui proses pemilihan acak untuk memulai proses pengelompokan.
- Pembentukan Kelompok: Setelah centroid awal ditentukan, setiap titik data dikelompokkan ke dalam cluster yang memiliki centroid terdekat berdasarkan hasil perhitungan jarak. Perhitungan kedekatan antar data dan centroid dapat dilakukan menggunakan metode Euclidean Distance atau ukuran jarak lainnya sesuai dengan kebutuhan, sebagaimana dinyatakan pada persamaan berikut.

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

- Pembaruan Pusat: Setelah proses pengelompokan selesai, centroid pada setiap cluster dihitung kembali dengan menentukan nilai rata-rata dari seluruh data yang menjadi anggota cluster tersebut.
- Iterasi: Proses pembentukan cluster dan pembaruan centroid dilakukan secara berulang hingga posisi centroid tidak lagi mengalami perubahan yang signifikan atau tidak terdapat perpindahan anggota data antarcluster.
- Penyelesaian: Proses clustering dihentikan setelah memenuhi kriteria penghentian yang telah ditentukan, seperti tidak adanya perubahan pada posisi centroid atau ketika jumlah iterasi telah mencapai batas maksimum yang ditetapkan.

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

2.4 Hasil

Model yang dihasilkan oleh algoritma K-Means dapat digunakan untuk menentukan keanggotaan cluster pada data baru serta mengidentifikasi posisi pusat cluster (*centroid*) dari setiap kelompok yang terbentuk (Andi Akram Nur Risal et al., 2024). Untuk mengevaluasi kualitas model yang dihasilkan, dapat digunakan metrik seperti *Within-Cluster Sum of Squares* (WCSS) dan *Silhouette Score*. Kedua metrik tersebut digunakan untuk mengukur tingkat kualitas hasil pengelompokan dalam merepresentasikan struktur data yang sebenarnya (Basalamah & Setyadi, 2023).

2.5 Evaluasi

Tahap evaluasi bertujuan untuk mengukur tingkat keberhasilan model yang telah diterapkan dalam penelitian. Penilaian dilakukan melalui interpretasi terhadap hasil clustering dengan membandingkan kinerja model menggunakan metode *Silhouette Coefficient*, sehingga dapat diketahui kualitas cluster yang terbentuk (Safitri, 2024).

Rumus *Silhouette Coefficient*:

$$s(i) = \frac{b(i) - a(i)}{\max \{a(i), b(i)\}} \quad (3)$$

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.

3. Hasil dan Pembahasan

3.1 Data UMKM

Data yang diolah dalam penelitian ini merupakan sampel yang dihasilkan secara sintetis untuk mensimulasikan aktivitas *e-commerce*. Dataset ini terdiri dari atribut transaksi per bulan, *Revenue E-commerce*, *Engagement Rate*, dan jumlah platform. Data sampel yang akan diuji cobakan terdiri dari 50 data UMKM. Contoh data yang digunakan dapat dilihat pada tabel 1.

Tabel 1. Data UMKM

UMKM	Transaksi/Bulan	Revenue (Rp)	Engagement Rate	Jumlah Platform
1	1176	58.989.336	0,2	4
2	1509	131.381.085	0,39	1
3	910	69.666.870	0,83	4
4	1344	112.577.472	0,32	4
5	1180	97.475.080	0,62	3
6	1145	70.456.430	0,1	3
7	1774	140.371.298	0,38	2
8	1094	98.894.318	0,34	3
9	1688	128.036.488	0,23	2
10	171	8.723.736	0,53	2
11	516	38.314.548	0,41	1
12	742	4.558.743	0,29	2
13	380	27.833.860	0,67	1
14	1298	103.447.920	0,36	3
15	137	8.018.473	0,23	3
...	...	...	...	...
49	1749	149.876.540	0,28	3
50	456	24.786.590	0,45	1

Dataset yang digunakan dalam penelitian ini mencakup 30 UMKM, dengan masing-masing data memiliki atribut berupa Transaksi per Bulan, *Revenue E-commerce*, *Engagement Rate*, dan Jumlah Platform. 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 klasterisasi akibat perbedaan skala nilai.

3.2 Normalisasi Data

Sebelum proses *Clustering* dilakukan, seluruh data dinormalisasi menggunakan metode *Min-Max Scaling* agar setiap atribut memiliki skala yang sama, yaitu dalam rentang 0 hingga 1. Hal ini penting karena nilai antar atribut memiliki satuan dan rentang yang berbeda, sehingga tanpa normalisasi dapat menyebabkan dominasi atribut tertentu dalam perhitungan jarak. Rumus normalisasi Min-Max adalah sebagai berikut:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (4)$$

Berdasarkan dataset 50 UMKM, diperoleh nilai minimum dan maksimum setiap atribut sebagai berikut:

- Transaksi per Bulan: minimum = 63, maksimum = 1949
- Revenue E-commerce*: minimum = 4.558.743, maksimum = 180.023.365
- Engagement Rate*: minimum = 0,10, maksimum = 0,90
- Jumlah Platform: minimum = 1, maksimum = 4

Sebagai contoh, proses normalisasi pada UMKM 1 dengan data (1176, 58.989.336, 0,20, 4) dilakukan sebagai berikut:

- Transaksi per Bulan:

$$X' = \frac{1176 - 63}{1949 - 63} = 0,5901$$

- Revenue E-commerce*:

$$X' = \frac{58.989.336 - 4.558.743}{180.023.365 - 4.558.743} = 0,3102$$

- Engagement Rate*:

$$X' = \frac{0,20 - 0,10}{0,09 - 0,10} = 0,125$$

- Jumlah Platform:

$$X' = \frac{4 - 1}{4 - 1} = 1$$

Dengan demikian, vektor data ternormalisasi untuk UMKM 1 adalah: (0,5901, 0,3102, 0,125, 1)

3.3 Inisialisasi *Centroid* Awal

Setelah data dinormalisasi, langkah selanjutnya adalah menentukan jumlah kluster (k). Dalam penelitian ini digunakan $k = 3$, yang merepresentasikan tiga kategori performa UMKM. *Centroid* awal dipilih secara acak dari data yang telah dinormalisasi. Sebagai contoh, *centroid* awal yang digunakan pada iterasi pertama adalah:

- Centroid* 0 (C_0) = UMKM 1
- Centroid* 1 (C_1) = UMKM 2
- Centroid* 2 (C_2) = UMKM 3

Pemilihan *centroid* awal ini bertujuan untuk menjadi titik pusat awal dalam proses pengelompokan.

3.4 Perhitungan Jarak ke *Centroid* (Euclidean Distance)

Setiap data UMKM kemudian dihitung jaraknya ke masing-masing *centroid* menggunakan jarak Euclidean. Sebagai contoh, jarak UMKM 4 ke *Centroid* 0 dihitung dengan menggunakan data ternormalisasi:

$$d(\text{UMKM 4}, C_0) = \sqrt{(x_1 - c_{01})^2 + (x_2 - c_{02})^2 + (x_3 - c_{03})^2 + (x_4 - c_{04})^2}$$

Perhitungan ini dilakukan untuk seluruh 50 UMKM terhadap ketiga *centroid*.

3.5 Penentuan *Cluster*

Setelah jarak ke seluruh *centroid* dihitung, setiap UMKM dikelompokkan ke dalam kluster yang memiliki jarak paling dekat. Hasil tahap ini menghasilkan pembagian sementara UMKM ke dalam tiga kluster, sebagai ilustrasi:

- Jika jarak UMKM 4 paling kecil ke *Centroid* 1, maka UMKM 4 masuk ke Cluster 1
- Jika jarak UMKM 10 paling kecil ke *Centroid* 2, maka UMKM 10 masuk ke Cluster 2

3.6 Pembaruan *Centroid*

Setelah seluruh UMKM dikelompokkan, dilakukan pembaruan *centroid* dengan menghitung nilai rata-rata dari seluruh data dalam setiap kluster. Sebagai contoh, jika Cluster 1 terdiri dari tiga UMKM (UMKM 2, UMKM 4, dan UMKM 5), maka *centroid* baru dihitung sebagai berikut:

a. *Centroid*₁ (Transaksi):

$$\frac{0,7668 + 0,6782 + 0,5959}{3} = 0,6803$$

b. *Centroid*₁ (Revenue):

$$\frac{0,7264 + 0,6185 + 0,5347}{3} = 0,6265$$

c. *Centroid*₁ (Engagement):

$$\frac{0,3625 + 0,2750 + 0,65}{3} = 0,4292$$

d. *Centroid*₁ (Platform):

$$\frac{0 + 1 + 0,6667}{3} = 0,5556$$

Sehingga *centroid* baru untuk Cluster 1 adalah: $C_1 = (0,6803, 0,6265, 0,4292, 0,5556)$

3.7 Iterasi Hingga Konvergen

Tahapan perhitungan jarak, pengelompokan data ke dalam cluster, serta pembaruan *centroid* diulang pada setiap iterasi dengan memanfaatkan *centroid* hasil pembaruan sebelumnya. Proses tersebut dihentikan ketika posisi *centroid* telah stabil dan tidak terdapat lagi perpindahan anggota antarcluster, yang mengindikasikan bahwa algoritma telah mencapai kondisi konvergen (*convergence*).

3.8 Evaluasi Model *Clustering* Menggunakan *Silhouette Coefficient*

Setelah proses iterasi pada algoritma K-Means Clustering mencapai kondisi konvergen, dilakukan evaluasi terhadap hasil pengelompokan yang telah diperoleh. Tahap ini bertujuan untuk mengukur kualitas pembentukan cluster serta menilai kesesuaian penempatan setiap UMKM pada cluster yang terbentuk. Dalam penelitian ini, evaluasi dilakukan dengan menggunakan metode *Silhouette Coefficient*, yang mampu mengukur tingkat kekompakan (*cohesion*) dalam cluster dan tingkat keterpisahan (*separation*) antarcluster.

Silhouette Coefficient merupakan metode evaluasi yang digunakan untuk mengukur tingkat kesesuaian suatu objek terhadap cluster tempat objek tersebut berada dibandingkan dengan cluster lainnya. Nilai *Silhouette Coefficient* berada pada rentang -1 hingga 1, di mana nilai yang mendekati 1 menunjukkan bahwa objek telah ditempatkan pada cluster yang sesuai, nilai yang mendekati 0 mengindikasikan adanya tumpang tindih antarcluster, sedangkan nilai negatif menunjukkan bahwa objek kemungkinan lebih tepat berada pada cluster lain.

3.9 Tahapan Perhitungan *Silhouette Coefficient*

a. Menghitung Nilai a(i)

Untuk setiap UMKM, dihitung nilai a(i), yaitu rata-rata jarak antara UMKM tersebut dengan seluruh UMKM lain yang berada dalam kluster yang sama. Jarak yang digunakan adalah jarak Euclidean berdasarkan data yang telah dinormalisasi. Sebagai contoh, misalkan UMKM 2 berada pada Cluster 1, maka nilai a(2) diperoleh dari rata-rata jarak UMKM 2 terhadap seluruh UMKM lain dalam Cluster 1.

b. Menghitung Nilai b(i)

Nilai $b(i)$ dihitung dengan cara menentukan rata-rata jarak UMKM ke- i terhadap seluruh data pada klaster lain, kemudian dipilih nilai rata-rata jarak terkecil. Klaster dengan jarak rata-rata terkecil tersebut disebut sebagai klaster tetangga terdekat (nearest neighboring cluster). Sebagai ilustrasi, jika UMKM 2 memiliki jarak rata-rata terdekat ke Cluster 0 dibandingkan Cluster 2, maka nilai $b(2)$ diambil dari jarak rata-rata UMKM 2 terhadap seluruh anggota Cluster 0.

c. Menghitung Nilai Silhouette $s(i)$

Setelah nilai $a(i)$ dan $b(i)$ diperoleh, nilai Silhouette untuk setiap UMKM dihitung menggunakan Persamaan (7). Sebagai contoh, jika untuk UMKM 2 diperoleh:

$$1) a(2) = 0,42$$

$$2) b(2) = 0,59$$

Maka nilai Silhouette untuk UMKM 2 adalah:

$$s(2) = \frac{0,59 - 0,42}{\max(0,42, 0,59)} = 0,288$$

Perhitungan ini dilakukan terhadap seluruh 50 UMKM yang menjadi objek penelitian.

3.10 Nilai Silhouette Rata-Rata

Setelah nilai Silhouette $s(i)$ untuk setiap UMKM dihitung, langkah selanjutnya adalah menghitung nilai Silhouette rata-rata untuk keseluruhan dataset, yang dirumuskan sebagai berikut:

$$\text{Silhouette Score} = \frac{1}{n} \sum_{i=1}^n s(i) \quad (5)$$

dengan: n = jumlah UMKM ($n = 50$)

Berdasarkan hasil perhitungan, diperoleh nilai *Silhouette Score* sebesar 0,292.

3.11 Interpretasi Hasil Evaluasi

Berdasarkan hasil proses K-Means Clustering dengan nilai $k = 3$, diperoleh distribusi anggota pada masing-masing klaster sebagaimana ditunjukkan pada Tabel 2.

Tabel 2. Distribusi Jumlah Anggota Cluster

Cluster	Jumlah UMKM	Karakteristik Umum
Cluster 0	18 UMKM	Aktivitas transaksi dan revenue tinggi
Cluster 1	16 UMKM	Engagement tinggi dengan aktivitas menengah
Cluster 2	16 UMKM	Aktivitas transaksi dan revenue rendah

Selain distribusi jumlah anggota cluster, rata-rata nilai atribut pada masing-masing klaster juga dihitung untuk melihat karakteristik setiap kelompok UMKM. Hasil rata-rata atribut dapat dilihat pada Tabel 3.

Tabel 3. Rata-rata Nilai Atribut Setiap Cluster

Cluster	Rata-rata Transaksi/Bulan	Rata-rata Revenue	Rata-rata Engagement Rate	Rata-rata Jumlah Platform
Cluster 0	1.650	Rp135.000.000	0,42	3
Cluster 1	850	Rp70.000.000	0,71	2
Cluster 2	320	Rp18.000.000	0,28	1

Berdasarkan hasil evaluasi menggunakan metode *Silhouette Coefficient*, diperoleh nilai Silhouette Score sebesar 0,292. Nilai tersebut menunjukkan bahwa struktur klaster yang dihasilkan masih lemah dan perlu ditingkatkan. Hal ini mengindikasikan bahwa masih terdapat tumpang

tindih karakteristik antar klaster sehingga pemisahan antar kelompok UMKM belum sepenuhnya optimal.

Meskipun demikian, hasil clustering masih mampu memberikan segmentasi awal UMKM berdasarkan aktivitas e-commerce. Cluster 0 cenderung berisi UMKM dengan performa aktivitas digital yang tinggi, Cluster 1 menunjukkan UMKM dengan tingkat engagement yang relatif baik namun aktivitas transaksi masih menengah, sedangkan Cluster 2 didominasi oleh UMKM dengan aktivitas dan revenue yang relatif rendah.

Hasil ini menunjukkan bahwa penggunaan algoritma K-Means masih dapat dimanfaatkan sebagai pendekatan awal dalam melakukan segmentasi UMKM berbasis aktivitas digital. Namun, kualitas klaster masih dapat ditingkatkan melalui penyesuaian jumlah klaster, penambahan atribut yang lebih spesifik, atau penggunaan metode clustering lain yang lebih sesuai terhadap karakteristik data.

3.12 Kesimpulan Evaluasi

Berdasarkan evaluasi menggunakan *Silhouette Coefficient*, dapat disimpulkan bahwa model K-Means *Clustering* yang diterapkan pada data 50 UMKM menghasilkan struktur klaster yang cukup representatif dan dapat digunakan sebagai dasar analisis segmentasi UMKM dalam pemanfaatan *e-commerce*.

4. Kesimpulan

Penelitian ini menerapkan algoritma K-Means Clustering untuk melakukan segmentasi 50 UMKM berdasarkan aktivitas e-commerce menggunakan atribut transaksi per bulan, revenue e-commerce, engagement rate, dan jumlah platform digital. Hasil clustering menunjukkan bahwa UMKM dapat dikelompokkan ke dalam tiga klaster dengan karakteristik aktivitas digital yang berbeda, yaitu kelompok dengan performa tinggi, performa menengah, dan performa rendah. Berdasarkan evaluasi menggunakan metode *Silhouette Coefficient*, diperoleh nilai sebesar 0,292. Nilai tersebut menunjukkan bahwa struktur klaster yang dihasilkan masih tergolong lemah dan masih terdapat tumpang tindih karakteristik antar klaster. Meskipun demikian, hasil clustering masih mampu memberikan segmentasi awal UMKM berdasarkan aktivitas e-commerce sehingga dapat digunakan sebagai gambaran awal dalam memahami karakteristik performa digital UMKM. Penelitian ini memiliki beberapa keterbatasan. Pertama, data yang digunakan merupakan data sintesis sehingga belum sepenuhnya merepresentasikan kondisi riil aktivitas UMKM pada platform e-commerce. Kedua, atribut yang digunakan masih terbatas pada transaksi per bulan, revenue e-commerce, engagement rate, dan jumlah platform digital, sehingga belum mencakup faktor lain yang dapat memengaruhi karakteristik UMKM, seperti jenis produk, tingkat kepuasan pelanggan, frekuensi promosi, maupun loyalitas konsumen. Oleh karena itu, penelitian selanjutnya disarankan menggunakan data aktual UMKM dengan jumlah atribut yang lebih beragam serta membandingkan beberapa metode clustering lainnya untuk memperoleh hasil segmentasi yang lebih optimal dan representatif.

Daftar Pustaka

Afrizal, M., Saputra, I., Satria, R., & Rahmaddeni. (2025). Analisis Performa Algoritma K-Means Clustering untuk Segmentasi Pasar di UMKM Muhammad. *VISA: Journal of Visions and Ideas*, 5(2), 647–657.

- Ameliana, N., Suarna, N., & Prihartono, W. (2024). Analisis Data Mining Pengelompokan Umkm Menggunakan Algoritma K-Means Clustering Di Provinsi Jawa Barat. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(3), 3261–3268. <https://doi.org/10.36040/jati.v8i3.9655>
- Andi Akram Nur Risal, Dyah Darma Andayani, Muh Ilham Suherman, & Andi Baso Kaswar. (2024). Utilizing the K-Means Clustering Algorithm for Analyzing Student Achievement Assessment at SMK Negeri 1 Gowa. *Journal of Embedded Systems, Security and Intelligent Systems*, 05(March), 60–67. <https://doi.org/10.59562/jessi.v5i1.2178>
- Bagustio, A. P., Purnamasari, A. I., & Irfan Ali. (2024). Analisis Data Penjualan Menggunakan Algoritma K-Means Clustering Pada Toko Kecantikan Putri. *PROSISKO: Jurnal Pengembangan Riset Dan Observasi Sistem Komputer*, 11(2), 159–167. <https://doi.org/10.30656/prosisko.v11i2.7928>
- Basalamah, A. T., & Setyadi, R. (2023). Penerapan Algoritma K-Means Clustering Pada Tingkat Penyelesaian Pendidikan Di Provinsi Indonesia. *Jurnal Informatika Dan Teknologi Komputer*, 4(2), 114–121. <https://ejurnalunsam.id/index.php/jicom/>
- Chen, L., Zhong, Q., & Li, Z. (2023). Analysis of spatial characteristics and influence mechanism of human settlement suitability in traditional villages based on multi-scale geographically weighted regression model: A case study of Hunan province. *Ecological Indicators*, 154(May), 110828. <https://doi.org/10.1016/j.ecolind.2023.110828>
- Hariyanto, Y., Primadewi, A., & Mukhtar, H. (2025). Analisis Kepuasan Masyarakat terhadap Pelayanan Publik menggunakan K-Means Clustering. *Journal of Information System Research (JOSH)*, 8(3), 460–469. <https://doi.org/10.47065/josh.v6i2.6577>
- Iskandar, N. R., Guntara, R. G., & Maesaroh, S. S. (2023). Implementasi Pengelolaan Transaksi Usaha Berbasis Android pada UMKM Menggunakan Metode Pengembangan Scrum. 1(6), 903–913.
- Linarti, U., Soleliza Jones, A. H., Zahrotun, L., & Rahmawati, A. (2024). Penerapan Metode K-Medoids Guna Pengelompokan Data Usaha Mikro, Kecil dan Menengah (UMKM) Bidang Kuliner Di Kota Yogyakarta. *Jurnal Ilmu Komputer Dan Sistem Informasi (JIKOMSI)*, 7(1), 37–45. <https://doi.org/10.55338/jikoms.v7i1.2194>
- Lubis, M. B., Khofifah Siregar, N., Maulida, S., Marpaung, W. T., & Marpaung, S. F. (2023). Penerapan Etika Bisnis Islam Dalam Meningkatkan Usaha Mikro Kecil dan Menengah (UMKM). *Penerapan Etika Bisnis (Melati, Dkk.) Madani: Jurnal Ilmiah Multidisiplin*, 1(12), 93–97. <https://doi.org/10.5281/zenodo.10427936>
- Mauliadi, R. (2022). Data Mining Menggunakan Algoritma K-Means Clustering dalam Analisis Tingkat Potongan Harga Terhadap Harga Jual Sepeda Motor Honda. *Jurnal Informatika Ekonomi Bisnis*, 4, 7–9. <https://doi.org/10.37034/inf.v4i4.156>
- Mohammad Ferdiansyah, & Umi Chotijah. (2024). Implementasi Algoritme K-Means++ Untuk Clustering Penjualan Bahan Bangunan. *Jurnal Ilmiah Teknik Informatika Dan Komunikasi*, 4(1), 181–193. <https://doi.org/10.55606/juitik.v4i1.767>
- Petrus, R., Suryani, E., & Bajari, M. (2023). Analisis Penanganan Pengaduan Konsumen Pada Situs Konsumen Indonesia Menggunakan Metode K-Means Cluster. ... : *Jurnal Penelitian Bisnis ...*, 1(1). <http://e-journal.nalanda.ac.id/index.php/SAMMAJIVA/article/view/99>
- Suara, A., Sanjaya, A., & Pamungkas, D. P. (2022). Implementasi Metode Double Moving Average Untuk Prediksi Produksi Sabun. *Seminar Nasional Inovasi Teknologi*, 224–229. <https://proceeding.unpkediri.ac.id/index.php/inotek/article/download/2587/1615/9644>
- Tajrin, T., Said, M., & Sinaga, A. R. S. (2025). Analisis Kepuasan Pelanggan Terhadap Bengkel

Ahass Karya Servis Di kota Medan Menggunakan Metode K-Means Clustering. *Management Studies and Entrepreneurship Journal (MSEJ)*, 6(4 SE-Articles), 4582–4595. <https://journal.yrpiiku.com/index.php/msej/article/view/8036>

Wisanta, E. H., & Marlim, Y. N. (2021). Analisis Algoritma K-Means Untuk Clustering Kepuasan Pelayanan: Mall Pelayanan Publik Pekanbaru. *Seminar Nasional Informatika (SENATIKA) Prosiding SENATIKA 2021*, 223–228.