Implementasi Metode Ensemble ROCK dalam Pengelompokan UMKM di Kabupaten Malang

Reza Sadiya Purwadwika[1], Kartika Maulida Hindrayani[2], Aviolla Terza Damaliana[3]
Universitas Pembangunan Nasional Veteran Jawa Timur
rezasadiya2@gmail.com

Abstrak: UMKM memiliki peran penting dalam perekonomian nasional, namun masih menghadapi berbagai permasalahan seperti rendahnya pemanfaatan teknologi, keterbatasan akses permodalan, dan lemahnya daya saing. Kompleksitas karakteristik data UMKM yang mencakup variabel numerik dan kategorikal menjadi tantangan dalam analisis dan pemetaan yang akurat. Penelitian ini bertujuan untuk mengelompokkan UMKM di Kabupaten Malang berdasarkan karakteristik usaha dengan pendekatan *ensemble clustering* menggunakan algoritma ROCK. Data terdiri dari 75 entri UMKM yang mencakup variabel numerik (omset, modal, tenaga kerja) dan kategorikal (jenis usaha, penggunaan aplikasi transportasi daring). *Clustering* dilakukan secara terpisah dengan *Agglomerative Hierarchical Clustering* untuk data numerik dan ROCK untuk data kategorikal. Hasil kedua metode digabungkan menggunakan pendekatan *ensemble* untuk memperoleh klaster yang lebih stabil dan representatif. Parameter optimal diperoleh pada *theta* = 0,40 dan k = 7 dengan nilai *Clustering Purity* (CP*) sebesar 1,0000 dan *Silhoutte Coefficient* sebesar 1.0000, menunjukkan pemisahan *cluster* yang baik. *Cluster* akhir menunjukkan perbedaan signifikan dalam skala usaha, pemanfaatan teknologi digital, dan performa ekonomi. Temuan ini diharapkan menjadi dasar dalam merancang kebijakan pengembangan UMKM yang lebih tepat sasaran dan berbasis data.
Kata Kunci : UMKM, *Clustering*, *Ensemble* ROCK, *Silhoutte Coefficient*

Abstract: *(MSMEs) play a vital role in the national economy, yet they continue to face several challenges, such as limited use of technology, restricted access to capital, and weak competitiveness. The complexity of MSME data—which includes both numerical and categorical variables—presents challenges in accurate analysis and mapping. This study aims to cluster MSMEs in Malang Regency based on business characteristics using an ensemble clustering approach with the ROCK algorithm. The dataset consists of 75 MSME entries, including numerical variables (revenue, capital, number of employees) and categorical variables (type of business, use of online transportation apps). Clustering was conducted separately: Agglomerative Hierarchical Clustering (AHC) was used for numerical data, and ROCK was applied for categorical data. The results from both methods were then combined using an ensemble approach to generate more stable and representative clusters. The optimal configuration was found at theta = 0.40 and k = 7, with a Clustering Purity (CP*) value of 1.0000 and a Silhouette Coefficient of 1.0000, indicating well-separated clusters. The final clusters show significant differences in business scale, use of digital technology, and economic performance. These findings are expected to serve as a basis for designing more targeted and data-driven MSME development policies.*
Keywords: *MSME, Clustering, Ensemble ROCK, Silhoutte Coefficient*

1. Pendahuluan

Usaha Mikro, Kecil, dan Menengah (UMKM) adalah usaha yang berperan penting dalam perekonomian negara Indonesia (Wati et al., 2024). Berdasarkan data Kementerian Koperasi dan UKM, jumlah UMKM saat ini mencapai 64,2 juta unit, menyumbang 61,07% terhadap Produk Domestik Bruto (PDB) atau setara dengan Rp8.573,89 triliun, serta menyerap sekitar 117 juta tenaga kerja atau 97% dari total angkatan kerja nasional (Kurniasih et al., 2024). Peran besar UMKM ini menjadikannya tulang punggung perekonomian dan sumber utama penyerapan tenaga kerja, termasuk di Kabupaten Malang.

Di era perkembangan teknologi dan informasi yang semakin pesat, pelaku Usaha Mikro, Kecil, dan Menengah (UMKM) dihadapkan pada peluang sekaligus tantangan bisnis yang semakin kompleks dan kompetitif (Idhom et al., 2024). Persaingan yang semakin sengit menuntut para pelaku usaha untuk mampu beradaptasi dan meningkatkan daya saing secara inovatif. Dalam menjawab tantangan tersebut, pemerintah terus mendorong pemanfaatan teknologi informasi dan pendekatan berbasis data dalam pengembangan UMKM sebagai bagian dari strategi pemulihan ekonomi nasional dan penguatan sektor informal. Dalam konteks ini, penerapan algoritma *machine learning* (ML) diharapkan dapat membantu memecahkan berbagai permasalahan yang dihadapi UMKM, termasuk dalam melakukan analisis dan pemetaan potensi usaha secara lebih sistematis, sehingga dapat meningkatkan kesejahteraan para pelaku usaha (Hindrayani et al., 2021).

Sejalan dengan upaya tersebut, pengembangan UMKM yang didukung oleh Pemerintah Kabupaten Malang perlu didasarkan pada pemetaan potensi usaha secara komprehensif agar kebijakan yang diambil tepat sasaran dan efektif dalam mendorong peningkatan kinerja UMKM di tingkat lokal. Salah satu metode yang dapat digunakan untuk melakukan pemetaan tersebut adalah *clustering*, yaitu teknik pengelompokan data yang memungkinkan identifikasi kelompok UMKM berdasarkan karakteristik tertentu. Dengan demikian, pemetaan potensi usaha melalui *clustering* dapat menjadi dasar yang kuat bagi perumusan kebijakan yang lebih terfokus dan strategis dalam meningkatkan daya saing serta kesejahteraan UMKM di Kabupaten Malang.

Penelitian sebelumnya mengenai pengelompokkan UMKM telah dilakukan oleh (Saskya & Apriyanto, 2022) dengan menggunakan data UMKM di Kelurahan Pangongangan, Kota Madiun. Pada penelitian tersebut, pengelompokkan UMKM dilakukan dengan metode *fuzzy c-means* yang hanya memanfaatkan variabel numerik, yaitu omset, aset, dan tenaga kerja. Namun, pendekatan yang hanya menggunakan variabel numerik memiliki keterbatasan, salah satunya adalah ketidakmampuan dalam merepresentasikan data kategorikal yang juga memegang peran penting dalam pengelompokan UMKM. Akibatnya, hasil pengelompokan kurang representatif terhadap kondisi nyata UMKM yang kompleks dan heterogen.

Data UMKM memiliki karakteristik kompleks yang terdiri dari variabel numerik dan kategorikal, sehingga diperlukan metode *clustering* yang dapat mengintegrasikan kedua jenis data tersebut. Salah satu pendekatan yang dapat digunakan adalah metode *ensemble clustering*, yaitu teknik pengelompokan yang menggabungkan hasil *clustering* dari beberapa algoritma untuk memperoleh kelompok yang lebih stabil dan representatif. Metode *Ensemble Robust Clustering using Links* (ROCK) merupakan salah satu metode yang relevan untuk tujuan ini. Ketika metode *Robust Clustering using Links* (ROCK) digabungkan dalam pendekatan *ensemble*, kemampuannya dalam mengolah data menjadi lebih kuat dan stabil.

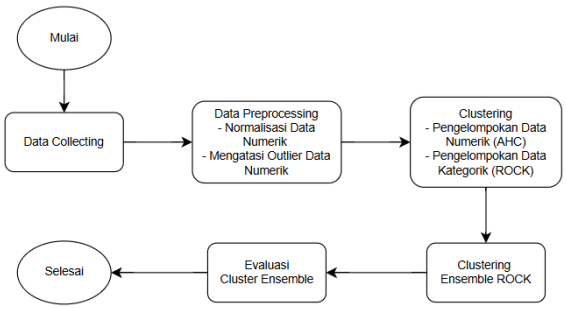
Pemilihan *Ensemble* ROCK dibandingkan dengan metode *clustering* lainnya didasarkan pada keunggulannya dalam menangani variabel kategorikal secara efektif melalui pendekatan kemiripan (*similarity*). Selain itu, metode ini menunjukkan fleksibilitas yang tinggi dalam menggabungkan hasil pengelompokan dari data numerik dan kategorikal secara simultan. Keunggulan lainnya adalah kemampuannya dalam menghasilkan klaster yang lebih konsisten serta lebih tahan terhadap pengaruh *noise* dan *outlier*.

Penelitian terkait penerapan metode *Ensemble* ROCK telah dilakukan oleh (Jannah et al., 2024) yang mengelompokkan data Perguruan Tinggi Swasta (PTS) di Kota Semarang. Penelitian tersebut menggunakan metode hirarki *agglomerative* untuk menganalisis data numerik dan metode ROCK untuk menganalisis data kategorikal. Selanjutnya hasilnya digabungkan dan dilakukan pengelompokan ulang menggunakan metode *ensemble* ROCK sehingga diperoleh final *cluster* sebanyak tiga kategori dengan nilai *threshold* sebesar 0,2.

Berdasarkan latar belakang tersebut, penelitian ini akan melakukan implementasi metode *Ensemble* ROCK untuk mengelompokkan UMKM di Kabupaten Malang dengan mengombinasikan algoritma *Agglomerative Hierarchical Clustering* untuk analisis data numerik dan metode ROCK untuk analisis data kategorikal. Diharapkan penelitian ini dapat meningkatkan kualitas dan ketepatan hasil pengelompokan UMKM di Kabupaten Malang sehingga dapat membantu pemerintah daerah dalam merumuskan kebijakan dan program pemberdayaan UMKM yang lebih tepat sasaran.

2. Metode Penelitian

2.1 Alur Penelitian



Gambar 1. Alur Penelitian

1. Data *Collecting*

   Mengumpulkan 75 data UMKM di Kabupaten Malang melalui kuisioner.
2. Data *Preprocessing*

   Melakukan normalisasi *Z-Score* pada data numerik dan penanganan *outlier* untuk meningkatkan kualitas data.
3. *Clustering*

   Pada tahap ini dilakukan dua metode *clustering* yang berbeda, yaitu AHC untuk data numerik dan ROCK untuk data kategorik.
4. *Clustering Ensemble* ROCK

   Tahapan ini merupakan proses penggabungan hasil *clustering* dari data numerik dan kategorik yang sebelumnya telah dikelompokkan menggunakan metode AHC dan ROCK. *Clustering Ensemble* dilakukan untuk menghasilkan *cluster* akhir yang lebih stabil dan representatif, dengan mempertimbangkan informasi dari kedua jenis data. Metode ROCK digunakan kembali pada tahap ini karena

kemampuannya dalam menangani data kategorik serta kemiripan berbasis links, yang dapat dimanfaatkan untuk mengintegrasikan hasil *clustering* dari berbagai sumber data.

5. Evaluasi *Cluster Ensemble*

   Tahap ini mengevaluasi hasil *clustering* dari *ensemble* ROCK dengan *Davies-Bouldin Index*.

2.2 Normalisasi *Z-Score*

Data yang memiliki selisih rentang yang signifikan dapat menghasilkan model dengan akurasi rendah. Normalisasi data merupakan salah satu upaya untuk mengatasi masalah ini. Proses normalisasi tidak mengubah informasi yang terkandung dalam data, namun membantu menyesuaikan rentang nilainya sehingga memudahkan dalam analisis dan penggunaan model. Salah satu metode normalisasi yang cukup baik untuk menyeimbangkan skala data yakni menggunakan *Z-score* (Fahrudin et al., 2021). Metode ini dihitung dengan melakukan pencarian nilai ukuran penyimpangan data dari hasil nilai dari rata – rata yang diukur dalam satuan standar deviasi. Berikut adalah rumus metode normalisasi *Z-Score* (Selayanti, et al., 2025)

$$Z = \frac{X - \mu}{\sigma}$$

2.3 *Agglomerative Hierarchy Clustering* (AHC)

Metode *Agglomerative Hierarchy Clustering* (AHC) digunakan untuk mengelompokkan data numerik. Pengelompokan data numerik dilakukan berdasarkan ukuran ketidakmiripan. Ukuran Ketidakmiripan yang biasa digunakan adalah *jarak Euclidean*. Misalkan terdapat dua objek dengan x = $(x_1, x_2, \dots, x_n)$ dan y = $(y_1, y_2, \dots, y_n)$, maka jarak *Euclidean* antara dua objek tersebut adalah (Roux, 2018):

$$d_{(x,y)} = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} = \sqrt{(x-y)^T(x-y)}$$

Dalam AHC terdapat 3 metode *cluster* hirarki yang dapat digunakan yaitu *single linkage*, *complete linkage*, dan *average linkage*. *Single linkage* mengelompokkan berdasarkan jarak terdekat antara anggota *cluster*, *complete linkage* berdasarkan jarak terjauh antara anggota *cluster*, *average linkage* berdasarkan jarak rata-rata antar objek dari semua pasangan titik data dalam dua *cluster* (Belinda et al., 2019).

a. *Single Linkage*

$$d_{(uv)w} = min(d_{uw}, d_{vw})$$

b. *Complete Linkage*

$$d_{(uv)w} = max(d_{uw}, d_{vw})$$

c. *Average Linkage*

$$d_{(uv)w} = \frac{d_{uw} + d_{vw}}{n_{(uv)} n_w}$$

2.4 *Robust Clustering using Links* (ROCK)

Metode *Robust Clustering using Links* (ROCK) digunakan untuk mengelompokkan data kategorik. Berikut adalah tahapan metode ROCK (Dutta et al., 2005):

a. Menghitung nilai kemiripan (*similarity)*

$$sim(X_i, X_j) = \frac{|X_i \cap X_j|}{|X_i \cup X_j|}, i \neq j$$

b. Menentukan tetangga

Penentuan pengamatan X_i dan X_j sebagai tetangga yaitu berdasarkan nilai $sim(X_i, X_j) \geq \theta$. Nilai *threshold (theta)* yang digunakan berkisar antara 0 sampai 1 menyesuaikan data yang ada.

c. Menghitung nilai *link*

Perhitungan nilai *link* untuk semua kemungkinan pasangan dari n objek dapat menggunakan matriks A. Matriks A adalah matriks berukuran $nn \times nn$ yang bernilai 1 (jika X_i dan X_j dinyatakan tetangga) serta bernilai 0 jika X_i dan X_j (dinyatakan bukan tetangga).

d. Menentukan Goodness Measure

$$g(C_i, C_j) = \frac{link(C_i, C_j)}{(n_i + n_j)^{1+2f(\theta)} - {n_i}^{1+2f(\theta)} - {n_j}^{1+2f(\theta)}}$$

Dimana n_i dan n_j adalah jumlah anggota dalam kelompok ke-*i* dan kelompok ke-*j*, sedangkan

$$f(\theta) = \frac{1-\theta}{1+\theta}$$

2.5 Validasi *Cluster* Numerik

Validasi *cluster* pada data numerik digunakan untuk menentukan metode pengelompokan *agglomerative hierarchy clustering* dan jumlah *cluster* yang paling optimum.

a. *Pseudo-F*

Perhitungan nilai *Pseudo-F* digunakan untuk menentukan jumlah cluster yang paling optimum setelah dilakukan pengelompokan (*clustering*). Semakin tinggi nilai *Pseudo-F* mengindikasikan jumlah *cluster* paling optimum (Maulina et al., 2025). Perhitungan validitas untuk menentukan jumlah cluster optimum dapat dituliskan dalam persamaan berikut :

Sum of Square Total (SST):

$$SST = \sum_{l=1}^{m}\sum_{i=1}^{n}(x_{il} - \bar{x}_l)^2$$

Sum of Square Within Group (SSW):

$$SSW = \sum_{c=1}^{C}\sum_{l=1}^{m}\sum_{i=1}^{n_c}(x_{ilc} - \bar{x}_{lc})^2$$

Sum of Square Between Group (SSB):

$$SSB = SST - SSW$$

Selanjutnya, dilakukan perhitungan nilai R^2 berikut :

$$R^2 = \frac{SSB}{SST} = \frac{[SST - SSW]}{SST}$$

Selanjutnya, dilakukan perhitungan nilai *Pseudo-F* berikut :

$$Pseudo - F = \frac{\frac{R^2}{C-1}}{\frac{1-R^2}{n-C}}$$

b. *Internal Cluster Dispersion Rate (icdrate)*
Perhitungan nilai *internal cluster dispersion rate (icdrate)* digunakan untuk menentukan metode pengelompokan yang paling optimum. Semakin rendah nilai *internal cluster dispersion rate (icdrate)* mengindikasikan metode pengelompokan yang paling optimum (Fitriana, 2021). Perhitungan nilai internal cluster dispersion rate (icdrate) berikut :

$$icdrate = 1 - \frac{SSB}{SST} = 1 - R^2$$

2.6 Validasi *Cluster* Kategorik

Validasi *cluster* pada data kategorik digunakan untuk menentukan jumlah *cluster* yang paling optimum. Pengukuran validasi *cluster* pada data kategorik dapat menggunakan perhitungan nilai *compactness*. Semakin tinggi nilai CP* mengindikasikan *cluster* yang dihasilkan semakin baik (Belinda et al., 2019). Perhitungan nilai *compactness* berikut :

$$CP* = \frac{1}{N}\sum_{k=1}^{K} n_k \left(\frac{\sum x_i, x_j \in C_k sim(x_i, x_j)}{n_k(n_k-1)/2}\right),$$

2.7 Cluster Ensemble

Model *ensemble* adalah kombinasi dua model atau lebih dalam satu fungsi (Muhaimin et al., 2021). *Cluster ensemble* merupakan metode pengelompokan data dengan menggabungkan hasil pengelompokan dari beberapa metode berbeda sehingga diperoleh solusi gabungan sebagai solusi akhir. Langkah-langkah algoritma CEBMDC (*Cluster Ensemble Based Mixed Data Clustering*) telah dijelaskan oleh He, Xu, dan Deng (2004).

a. Memisahkan data campuran menjadi dua bagian data yaitu data kategorik murni dan data numerik murni
b. Melakukan pengelompokan dengan algoritma pengelompokan untuk masing-masing jenis data tersebut.
c. Menggabungkan hasil pengelompokan (output) dari data numerik dan data kategorik. Penggabungan ini disebut proses *ensemble*.
d. Melakukan pengelompokan ensemble menggunakan algoritma pengelompokan data kategorik untuk mendapatkan cluster akhir.

2.8 Evaluasi *Cluster Ensemble*

Evaluasi *cluster ensemble* bertujuan untuk mengukur kualitas pemisahan *cluster* yang terbentuk. Salah satu metrik yang digunakan adalah *Silhoutte Coefficient*, yang menilai kompak dan terpisahnya *cluster* hasil *ensemble*. *Davies-Bouldin Index* (DBI) adalah metrik evaluasi yang digunakan untuk menilai kualitas hasil *clustering*. Semakin mendekati angka 1 pada *Silhouette Coefficient*, maka semakin tinggi kualitas pengelompokkan pada satu kelompok tersebut. Sebaliknya, semakin mendekati nilai -1 pada *Silhouette Coefficient* (Rahmawati et al., 2024). Perhitungan dapat dituliskan berikut:

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

3. Hasil dan Pembahasan

3.1 Deskripsi Data

Data yang digunakan dalam penelitian ini berasal dari pelaku Usaha Mikro, Kecil, dan Menengah (UMKM) di Kabupaten Malang, dengan total sebanyak 75 data. Data mencakup informasi terkait karakteristik usaha dan pelaku usaha. Variabel-variabel yang digunakan dalam penelitian ini dapat dilihat pada Tabel 1 berikut:

Tabel 1. Variabel Penelitian

Variabel	Jenis Variabel	Tipe Data
Omzet	Numerik	*Integer*
Modal	Numerik	*Integer*
Jumlah Tenaga Kerja	Numerik	*Integer*
Jenis Usaha	Kategorik 0 = Mamin 1 = Oleh-oleh	*Boolean*
Terdaftar Online Usaha	Kategorik 0 = Ya 1 = Tidak	*Boolean*

3.2 Data Preprocessing

Pra-pemrosesan data adalah proses menyiapkan data agar menjadi data yang sudah tetap, sebelum digunakan sebagai data pelatihan (Fahrudin et al., 2017). Berikut adalah proses nya:

a. Statistika Deskriptif Pada Data Numerik

Statistika deskriptif bertujuan untuk menggambarkan karakteristik data.

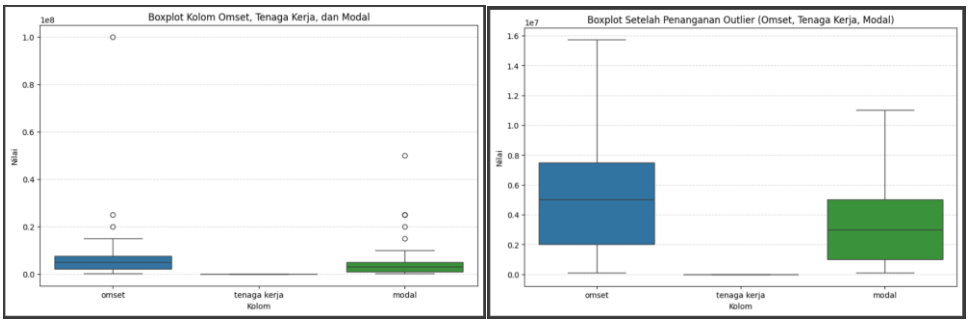
Tabel 2. Statistika Deskriptif Data Numerik

	omset	tenaga kerja	modal
count	75	75	75
mean	6.862.667	2,89	5.272.000
std	11.987.899	1,39	7.684.427
min	100.000	1	100.000
25%	2.000.000	2	1.000.000
50%	5.000.000	3	3.000.000
75%	7.500.000	4	5.000.000
max	100.000.000	7	50.000.000

Berdasarkan tabel 2 diketahui bahwa rata-rata omset UMKM sebesar Rp6.862.667 dengan sebaran yang tinggi, terlihat dari simpangan baku yang besar dan selisih nilai minimum-maksimum yang jauh. Modal juga memiliki distribusi yang lebar, dengan rata-rata Rp5.272.000 dan maksimum hingga Rp50.000.000. Sementara itu, tenaga kerja rata-rata sebanyak 3 orang. Nilai-nilai ekstrem pada omset dan modal mengindikasikan adanya *outlier*, sehingga diperlukan normalisasi dan penanganan lanjutan sebelum analisis lebih lanjut.

b. Menangani *Outlier* Pada Data Numerik

Pada tahap ini melakukan penangangan *outlier* yang ditemukan pada variabel numerik. Penanganan ini bertujuan menjaga kestabilan data dan mencegah pengaruh negatif dari nilai ekstrem pada analisis selanjutnya.



Gambar 3. *Outlier*

Berdasarkan gambar 3 diketahui bahwa tidak ditemukan *outlier* pada variabel tenaga kerja, namun pada variabel modal dan omset terdapat nilai-nilai ekstrem. Nilai *outlier* tersebut langsung diganti dengan nilai kuartil terdekat berdasarkan metode *Interquartile Range* (IQR), yaitu Q1 untuk nilai yang di bawah batas bawah dan Q3 untuk nilai yang di atas batas atas.

c. Normalisasi *Z-Score* Pada Data Numerik

Setelah penanganan *outlier*, langkah selanjutnya adalah melakukan normalisasi pada data numerik, khususnya variabel modal, omset, dan tenaga kerja. Normalisasi dilakukan menggunakan metode *Z-Score*, yang berfungsi untuk mengubah skala data menjadi distribusi dengan rata-rata (*mean*) 0 dan standar deviasi 1. Normalisasi ini penting untuk memastikan bahwa setiap variabel memiliki kontribusi yang seimbang dalam proses analisis atau pemodelan,

3.3 Clustering

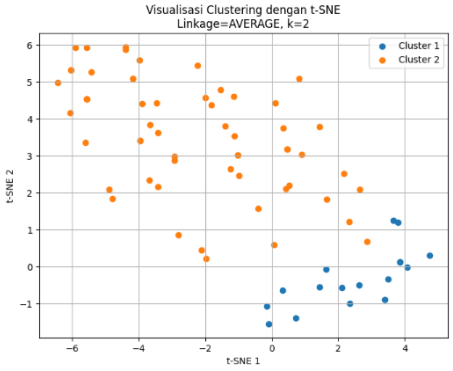
a. *Agglomerative Hierarchical Clustering* (AHC) numerik

Clustering terhadap data numerik dilakukan menggunakan metode *Agglomerative Hierarchical Clustering* (AHC) dengan tiga pendekatan linkage, yaitu *single, complete, dan average*. Evaluasi dilakukan menggunakan dua metrik utama, yaitu nilai *Pseudo-F dan* ICD *rate* (*Intra Cluster Distance Rate*). Berikut adalah gambar output :

```
=== Validasi Clustering Berdasarkan Pseudo-F dan ICD Rate ===

>>> LINKAGE: SINGLE
  K=2 → Pseudo-F: 6.2763, ICD rate: 2.7625
  K=3 → Pseudo-F: 6.1377, ICD rate: 2.5630
  K=4 → Pseudo-F: 4.7641, ICD rate: 2.4973
  K=5 → Pseudo-F: 11.0485, ICD rate: 1.8390
  K=6 → Pseudo-F: 16.2982, ICD rate: 1.3755

>>> LINKAGE: COMPLETE
  K=2 → Pseudo-F: 41.5323, ICD rate: 1.9121
  K=3 → Pseudo-F: 38.0771, ICD rate: 1.4579
  K=4 → Pseudo-F: 39.9909, ICD rate: 1.1153
  K=5 → Pseudo-F: 32.3047, ICD rate: 1.0541
  K=6 → Pseudo-F: 41.1634, ICD rate: 0.7532

>>> LINKAGE: AVERAGE
  K=2 → Pseudo-F: 50.4550, ICD rate: 1.7739
  K=3 → Pseudo-F: 39.5883, ICD rate: 1.4288
  K=4 → Pseudo-F: 31.3630, ICD rate: 1.2902
  K=5 → Pseudo-F: 47.8681, ICD rate: 0.8031
  K=6 → Pseudo-F: 40.5154, ICD rate: 0.7622

=== HASIL AKHIR CLUSTERING TERBAIK ===
Jumlah klaster optimum : 2
Metode linkage terbaik : AVERAGE
Nilai Pseudo-F tertinggi: 50.4550
Nilai ICD terkecil      : 1.7739
```

Gambar 4. AHC Numerik

Berdasarkan gambar 4 diketahui hasil validasi menunjukkan bahwa metode *linkage average* dengan jumlah *cluster* sebanyak dua (k=2) memberikan performa terbaik, dengan nilai *Pseudo-F* tertinggi sebesar 50.4550 dan ICD *rate* sebesar 1.7739. Nilai *Pseudo-F* yang tinggi mengindikasikan pemisahan antar cluster yang baik, sementara nilai ICD rate yang rendah menunjukkan kekompakan cluster yang terbentuk. Berikut adalah gambar visualisasi cluster yang terbentuk :



Gambar 5. Visualisasi *Cluster* AHC Numerik

Berdasarkan gambar 5 diketahui visualisasi hasil *clustering* menggunakan t-SNE memperkuat hasil evaluasi tersebut. Terlihat bahwa data terbagi menjadi dua *cluster* yang cukup jelas dan terpisah satu sama lain. *Cluster* pertama cenderung berada di sisi kiri ruang t-SNE, sedangkan *cluster* kedua tersebar di sisi kanan, tanpa adanya tumpang tindih yang signifikan. Hal ini menunjukkan bahwa metode AHC dengan *linkage average* mampu memisahkan data dengan baik secara visual dan statistik.

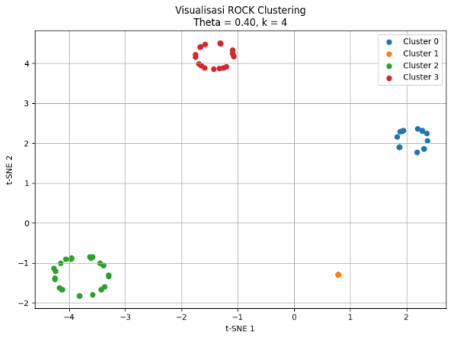
b. *Robust Clustering using Links* (ROCK) Kategorik

Clustering terhadap data kategorikal dilakukan menggunakan metode ROCK *(Robust Clustering using Links*), yang dievaluasi berdasarkan parameter *theta*, jumlah *cluster* (k), dan nilai CP*. Eksperimen dilakukan pada rentang nilai *theta* dari 0.05 hingga 0.90 dan jumlah cluster dari 2 hingga 7. Berikut adalah gambar output:

```
Theta=0.40, k=7 → CP* Total = 1.0000
```

Gambar 6. ROCK Kategorik

Berdasarkan gambar 6 diketahui hasil evaluasi menunjukkan bahwa nilai CP* tertinggi, yaitu 1.0000, secara konsisten dicapai pada kombinasi k=4 untuk sebagian besar nilai *theta*, termasuk *theta* terkecil yaitu 0.40. Kombinasi *theta*=0.40 dan k=4 dipilih sebagai konfigurasi terbaik karena mampu menghasilkan *cluster* yang paling murni dan stabil terhadap perubahan parameter. Berikut adalah gambar visualisasi *cluster* yang terbentuk :



Gambar 7. Visualisasi *Cluster* ROCK Kategorik

Berdasarkan gambar 7 diketahui visualisasi hasil *clustering* menggunakan metode ROCK dengan parameter *theta* = 0.40 dan jumlah cluster (k) = 4, yang divisualisasikan melalui teknik t-SNE. Terlihat bahwa keempat *cluster* terbentuk secara terpisah dengan sangat jelas di ruang dua dimensi, tanpa adanya tumpang tindih antar kelompok. Hal ini menunjukkan bahwa metode ROCK berhasil memisahkan data kategorikal dengan baik. *Cluster* 0 (biru), *Cluste*r 2 (hijau), dan *Cluster* 3 (merah) membentuk kelompok yang padat dan terfokus pada wilayah tertentu, mencerminkan kohesi internal yang kuat. Sementara itu, *Cluster* 1 (oranye)

hanya terdiri dari satu titik data, yang mengindikasikan adanya data yang memiliki karakteristik sangat berbeda dari yang lain. Secara keseluruhan, visualisasi ini menunjukkan bahwa konfigurasi parameter yang digunakan mampu menghasilkan pemisahan klaster yang optimal dan representatif.

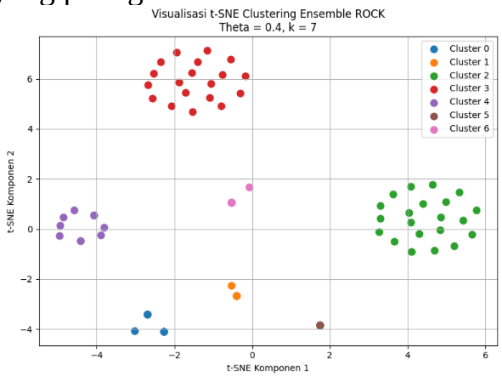
3.4 Clustering Ensemble ROCK

Pendekatan *ensemble* ini menggabungkan hasil *clustering* dari beberapa model atau konfigurasi parameter untuk meningkatkan stabilitas dan akurasi hasil akhir. Evaluasi dilakukan terhadap kombinasi nilai parameter *theta*, jumlah *cluster* (k), dan CP* sebagai metrik utama. Eksperimen dilakukan pada rentang nilai *theta* dari 0.05 hingga 0.90 dan jumlah cluster dari 2 hingga 7.

```
Theta=0.40, k=7 → CP* Total = 1.0000
```

Gambar 8. *Ensemble* ROCK

Berdasarkan gambar 8 diketahui hasil evaluasi menunjukkan bahwa konfigurasi *theta* = 0.40 dan k = 7 memberikan performa terbaik dengan nilai CP* tertinggi sebesar 1,0000 Konfigurasi ini dipilih sebagai hasil *ensemble* terbaik karena menghasilkan *cluster* yang paling murni dan konsisten terhadap variasi parameter.



Gambar 9. Visualisasi *Ensemble* ROCK

Berdasarkan gambar 9 diketahui visualisasi hasil *ensemble clustering* dengan t-SNE menunjukkan bahwa data terbagi ke dalam tujuh *cluster* dengan penyebaran yang relatif jelas dan tidak saling tumpang tindih. *Cluster* 2 (merah) dan *cluster* 3 (hijau) mendominasi sebagian besar data dan membentuk kelompok besar yang terpisah secara tegas. Sementara itu, *cluster* 0 (biru), *cluster* 4 (ungu), dan *cluster* 6 (*pink*) membentuk kelompok kecil yang juga terpisah dengan baik. Adapun *cluster* 1 (oranye) dan *cluster* 5 (coklat) hanya terdiri dari satu atau dua titik, yang mengindikasikan keberadaan data dengan karakteristik unik. Meskipun terdapat variasi ukuran antar *cluster*, metode *Ensemble* ROCK dengan parameter *theta* = 0,4 dan k = 7 mampu mengidentifikasi pola dan struktur minoritas secara akurat, serta membentuk *cluster* yang kompak dan tidak tumpang tindih.

Berdasarkan hasil *clustering* yang dihasilkan oleh metode *ensemble* ROCK, diperoleh empat kelompok UMKM yang memiliki karakteristik berbeda satu sama lain. Ciri khas dari masing-masing *cluster*, dapat diketahui dengan analisis lebih lanjut melalui distribusi variabel kategorik serta statistik deskriptif dari variabel numerik yang terdapat dalam setiap *cluster*. Analisis ini memberikan gambaran menyeluruh mengenai pola-pola umum yang muncul dalam masing-masing kelompok, seperti jenis usaha yang dominan, pemanfaatan layanan ojek online, rata-rata omzet, modal, serta jumlah tenaga kerja.

Tabel 3. Profilisasi *Cluster Ensemble* ROCK

cluster_ensemble_rock	Rata-rata modal	Rata-rata tenaga kerja	Rata-rata omset	Dominasi Ojol	Dominasi Jenis
1	10.285.714	3.43	9.428.571	ya	mamin
2	9.400.000	4.4	10.350.000	ya	oleh
3	2.209.523	2.81	3.976.190	tidak	oleh
4	1.736.842	2.21	3.931.578	tidak	mamin
5	2.300.000	2.5	3.500.000	ya	mamin
6	9.400.000	4.2	11.300.000	tidak	oleh
7	4.125.000	3	6.187.500	ya	oleh

Berdasarkan tabel 3 menunjukkan profilisasi *cluster ensemble* ROCK, diketahui bahwa masing-masing *cluster* menunjukkan karakteristik yang berbeda secara signifikan. *Cluster* 1 dan *Cluster* 2 merupakan kelompok dengan performa ekonomi yang relatif tinggi, ditunjukkan oleh rata-rata modal dan omzet yang besar, serta telah memanfaatkan layanan ojek online (*ojol*). Keduanya juga memiliki tenaga kerja yang lebih banyak dibandingkan *cluster* lainnya, dengan dominasi jenis usaha makanan dan minuman (*Cluster* 1) serta oleh-oleh (*Cluster* 2).

Sementara itu, *Cluster* 3 dan *Cluster* 4 menunjukkan karakteristik dengan performa ekonomi yang lebih rendah, dengan rata-rata omzet dan modal paling kecil serta belum memanfaatkan ojol. Kedua *cluster* ini didominasi oleh jenis usaha oleh-oleh (*Cluster* 3) dan makanan-minuman (*Cluster* 4), dan memiliki jumlah tenaga kerja paling sedikit. *Cluster* 5 dan *Cluster* 6 memiliki karakteristik menengah, dengan modal dan omzet yang sedang serta dominasi penggunaan ojol, terutama pada jenis usaha makanan dan minuman. Cluster 7 menunjukkan karakteristik unik dengan modal dan tenaga kerja yang relatif rendah, namun memiliki omzet yang cukup tinggi. Kelompok ini juga menggunakan ojol dan didominasi oleh usaha oleh-oleh.

Secara keseluruhan, profilisasi ini menunjukkan bahwa pemanfaatan teknologi seperti ojol berkorelasi dengan peningkatan performa ekonomi, serta adanya variasi yang cukup jelas antara kelompok usaha makanan dan oleh-oleh dalam hal kebutuhan modal, tenaga kerja, dan capaian omzet. Temuan ini penting untuk menjadi dasar dalam penyusunan strategi pemberdayaan UMKM yang lebih terarah sesuai karakteristik masing-masing kelompok.

3.5 Evaluasi Cluster Ensemble

Evaluasi *cluster ensemble* ROCK ini menggunakan *Silhoutte Coefficient* yang menghasilkan nilai sebesar 1.0000, yang berarti kualitas *cluster* baik dengan tingkat kekompakan yang tinggi dan pemisahan yang jelas antar *cluster*.

3.6 Pembahasan

Hasil utama dari penelitian ini menunjukkan bahwa metode Ensemble ROCK dengan parameter theta = 0.40 dan k = 7 menghasilkan kualitas klaster yang sangat baik, terbukti dari nilai Silhouette Coefficient sebesar 1.0000 dan visualisasi t-SNE yang memperlihatkan pemisahan antar cluster yang jelas. Hal ini menunjukkan bahwa metode ini efektif dalam menangani data campuran numerik dan kategorikal, serta mampu mengidentifikasi karakteristik UMKM secara representatif.

Jika dibandingkan dengan penelitian terdahulu oleh (Saskya & Apriyanto 2022), terdapat peningkatan signifikan dalam hal representasi data. Penelitian

tersebut menggunakan metode fuzzy c-means yang terbatas pada variabel numerik (omzet, aset, dan tenaga kerja), sehingga kurang mampu menggambarkan kondisi nyata UMKM yang heterogen. Dengan pendekatan Ensemble ROCK, penelitian ini memberikan keunggulan fleksibilitas dan stabilitas hasil yang lebih baik karena turut mengintegrasikan variabel kategorikal.

Penerapan metode ini juga sejalan dengan penelitian yang dilakukan oleh (Jannah et al., 2024) mengenai pengelompokan data Perguruan Tinggi Swasta di Semarang. Jika Jannah et al. menggunakan nilai *threshold* sebesar 0,2 untuk memperoleh tiga kategori, penelitian ini menemukan bahwa konfigurasi *theta* 0,40 lebih optimal untuk data UMKM di Kabupaten Malang. Hal ini memperkuat literatur mengenai efektivitas algoritma ROCK dalam *clustering* data kategorikal, sekaligus menunjukkan inovasi integrasi *ensemble* untuk menghasilkan segmentasi yang lebih utuh.

Selain itu, dibandingkan dengan penggunaan metode konvensional seperti Agglomerative Hierarchical Clustering (AHC) yang digunakan oleh (Selayanti et al., 2025) untuk segmentasi wisata kuliner, pendekatan ensemble dalam studi ini menawarkan stabilitas yang lebih tinggi. Meskipun memiliki keterbatasan pada jumlah data yang relatif kecil, secara praktis hasil ini dapat dimanfaatkan oleh pemerintah daerah Kabupaten Malang untuk menyusun strategi pengembangan UMKM yang lebih terarah. Secara akademik, penelitian ini berhasil memperkaya literatur mengenai pengelompokan data campuran dengan pendekatan ensemble yang lebih adaptif.

4. Kesimpulan

Dalam penelitian ini, telah dilakukan analisis *clustering* terhadap data UMKM di Kabupaten Malang dengan pendekatan numerik dan kategorik. Untuk data numerik, proses awal dilakukan dengan penanganan *outlier* menggunakan metode *Interquartile Range* (IQR) serta normalisasi menggunakan *Z-Score* agar data memiliki skala yang seimbang. *Clustering* numerik dilakukan menggunakan metode *Agglomerative Hierarchical Clustering* (AHC) dengan teknik *average linkage* dan menghasilkan dua *cluster* sebagai konfigurasi terbaik. Hasil evaluasi menunjukkan nilai *Pseudo-F* yang tinggi sebesar 50.4550 dan nilai ICD *Rate* yang rendah sebesar 1.7739, yang mencerminkan pemisahan dan kekompakan *cluster* yang baik. Sementara itu, untuk data kategorikal, metode ROCK dengan konfigurasi *theta* = 0.40 dan k = 4 menghasilkan nilai CP* sempurna sebesar 1.0000, menunjukkan pemisahan *cluster* yang sangat baik.

Pendekatan *ensemble* ROCK yang menggabungkan berbagai hasil *clustering* menghasilkan tujuh *cluster* dengan konfigurasi terbaik pada theta = 0.40 dan k = 7, dengan nilai CP* sebesar 1.0000 dan *Silhouette Coefficient* sebesar 1.0000, mengindikasikan kualitas *clustering* yang sangat tinggi dan stabil. Secara karakteristik, *cluster* dengan performa ekonomi lebih tinggi umumnya didominasi oleh UMKM yang telah memanfaatkan layanan digital seperti ojek online (ojol), sedangkan cluster dengan performa lebih rendah cenderung belum terdigitalisasi.

Hasil ini penting karena menunjukkan bahwa metode *ensemble* ROCK adaptif terhadap data campuran dan mampu menghasilkan segmentasi UMKM yang representatif. Keunggulan ini menjadikan metode tersebut lebih fleksibel dibandingkan pendekatan konvensional. Meski jumlah data terbatas, temuan ini bermanfaat secara praktis dalam merancang strategi pemberdayaan UMKM berbasis karakteristik *cluster*, serta memberikan kontribusi akademik terhadap pengembangan metode *clustering* untuk data campuran secara lebih adaptif dan terintegrasi.

Daftar Pustaka

Fahrudin, T. M., Riyantoko, P. A., Hindrayani, K. M., & Swari, M. H. P. (2021). Cluster Analysis of Hospital Inpatient Service Efficiency Based on BOR, BTO, TOI, AvLOS Indicators using Agglomerative Hierarchical Clustering. *Telematika*, *18*(2), 194. https://doi.org/10.31315/telematika.v18i2.4786

Fahrudin, T. M., Syarif, I., & Barakbah, A. R. (2017). Data Mining Approach for Breast Cancer Patient Recovery. *EMITTER International Journal of Engineering Technology*, *5*(1), 36–71. https://doi.org/10.24003/emitter.v5i1.190

Hendrawan, M. B., & Gunadi, G. (2025). Segmentasi Layanan Jasa Kantor Akuntan Publik Menggunakan Algoritma K-Means Clustering. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, *7*(2), 268–276. https://doi.org/10.47233/jteksis.v7i2.1912

Idhom, M., Priananda, A. M., Raynaldi, A., Halim, R. N., Pamungkas, S. A., & Wardana, A. C. (2024). Upaya Rebranding Sebagai Bentuk Kepedulian Terhadap UMKM. *Journal of Community Service (JCOS)*, *2*(4), 124–131. https://doi.org/10.56855/jcos.v2i4.1112

Jannah, B., Utami, I. T., & Hakim, A. R. (2024). Metode Ensemble Robust Clustering Using Links (Rock) Untuk Pengelompokan Perguruan Tinggi Swasta (Pts) Di Kota Semarang. *Jurnal Gaussian*, *12*(3), 445–452. https://doi.org/10.14710/j.gauss.12.3.445-452

Kurniasih, R., Wasiyanti, S., & Utami, L. D. (2025). *Optimalisasi Transaksi Keuangan Menggunakan Zahir Accounting Versi 6 . 0 Pada Bengkel Rahmat Cimandala Raya Motor*. *5*(1), 1–8.

Maulida Hindrayani, K., Anjani, A., & Lina Nurlaili, A. (2021). Penerapan Machine Learning pada Penjualan Produk UMKM : Studi Literatur. *Prosiding Seminar Nasional Sains Data*, *1*(01), 19–23. https://doi.org/10.33005/senada.v1i01.7

Muhaimin, A., Aji Riyantoko, P., Prabowo, H., & Trimono, T. (2021). Negative Binomial Time Series Regression – Random Forest Ensemble in Intermittent Data. *Internasional Journal of Data Science, Engineering, and Anaylitics*, *1*(2), 36–42. https://doi.org/10.33005/ijdasea.v1i2.10

Nourma Maulina, M., Arya Saputra, B., Statistika, D., & Sains dan Matematika, F. (2025). Optimalisasi Cluster Pada Cluster Hierarki Menggunakan Pseudo F-Statistics Calinski Harabasz untuk Ketahanan Pangan. *Jurnal Gaussian*, *14*(1), 1–12. https://doi.org/10.14710/j.gauss.14.1.01-12

Nur, I., & Fitriana, L. (2021). Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Keluarga Sehat Menggunakan Metode Klaster Hirarki dan Non Hirarki. *Jurnal Paradigma: Jurnal Multidisipliner Mahasiswa Pascasarjana Indonesia*, *2*(1), 27–36. https://journal.ugm.ac.id/paradigma/article/view/66072

Rahmawati, T., Wilandari, Y., & Kartikasari, P. (2024). Analisis Perbandingan Silhouette Coefficient Dan Metode Elbow Pada Pengelompokkan Provinsi Di Indonesia Berdasarkan Indikator Ipm Dengan K-Medoids. *Jurnal Gaussian*, *13*(1), 13–24. https://doi.org/10.14710/j.gauss.13.1.13-24

Saskya, L. F., & Apriyanto, R. A. N. (2022). Implementasi Fuzzy C-Means Clustering Dalam Pengelompokan Umkm Di Kelurahan Pangongangan Kota Madiun. *Power Elektronik : Jurnal Orang Elektro*, *11*(2), 204. https://doi.org/10.30591/polektro.v12i1.3713

Selayanti, N., Putri, S. A., & Fahrudin, T. M. (2025). Analisis Segmentasi Sentra Wisata Kuliner untuk Optimalisasi Omzet UMKM di Surabaya Menggunakan Metode Agglomerative Hierarchical Clustering. *JoMMiT : Jurnal Multi Media Dan IT*, *8*(2), 113–122. https://doi.org/10.46961/jommit.v8i2.1351

Sofyan, S. (2017). Peran UMKM (Usaha Mikro, Kecil, Dan Menengah) Dalam Perekonomian Indonesia. *Jurnal Bilancia*, *11*(1), 33–59. file:///C:/Users/Asus/Downloads/298-Article Text-380-1-10-20180728-3.pdf

Studi, P. S. (2019). Nadira Sri Belinda, Izzati Rahmi Hg, Hazmira Yozza. *Jurnal Matematika*

UNAND, *VIII*(2), 108–119.

Wati, D. L., Septianingsih, V., Khoeruddin, W., & Al-Qorni, Z. Q. (2024). Peranan UMKM (Usaha Mikro Kecil dan Menengah) Dalam Meningkatkan Kesejahteraan Rakyat. *Jurnal Dinamika Ekonomi Syariah*, *3*(1), 265–285.