

Peningkatan Kinerja K-Means Clustering pada Data Penggunaan ChatGPT oleh UMKM

Improving K-Means Clustering Performance on ChatGPT Usage Data by MSMEs

Endang Supriyati¹, Tri Listyorini², Mohammad Iqbal³

^{1,2}Program Studi Teknik Informatika, Universitas Muria Kudus

³Program Studi Teknik Elektro, Universitas Muria Kudus

E-mail: ¹endang.supriyati@umk.ac.id, ²trilistyorini@umk.ac.id, ³mohammad.iqbal@umk.ac.id

Abstrak

Di era digital saat ini, Usaha Mikro, Kecil Dan Menengah (UMKM) semakin banyak menggunakan teknologi *chatbot* seperti ChatGPT untuk berinteraksi dengan pelanggan. Namun, mengekstraksi wawasan yang bermakna dari sejumlah besar data yang dihasilkan oleh interaksi ini masih merupakan sebuah tantangan. Penelitian ini mengusulkan pendekatan untuk meningkatkan analisis data penggunaan ChatGPT oleh UMKM menggunakan kombinasi K-Means *clustering* dan *Independent Component Analysis* (ICA) untuk pengurangan dimensi. Pendekatan ini bertujuan untuk menemukan pola dan *cluster* tersembunyi dalam data dengan mereduksi ruang fitur menggunakan ICA dan kemudian menggunakan K-Means *clustering*. Selain itu, penelitian ini mempertimbangkan kombinasi metode *Elbow* untuk mengoptimalkan pemilihan jumlah *cluster* yang optimal. Melalui eksperimen dan evaluasi, ditunjukkan efektivitas metode yang diusulkan, sehingga memberikan UMKM alat yang ampuh untuk mendapatkan wawasan berharga dari data penggunaan ChatGPT. Penelitian ini menggunakan data dari 50 UMKM di Kabupaten Kudus. Hasil evaluasi kinerja K-means menunjukkan untuk data yang relatif kecil, penggunaan ICA kurang mempengaruhi kinerja K-Means. Skenario uji coba data dengan ICA menunjukkan skor *silhouette* yang lebih rendah, yaitu 0.19 dan 0.18 untuk masing-masing data pelatihan dan pengujian, sedangkan skenario uji coba data tanpa ICA menunjukkan skor *silhouette* yang lebih tinggi, yaitu 0.30 dan 0.23.

Kata kunci: K-Means *clustering*, metode *Elbow*, ICA, ChatGPT, UMKM

Abstract

In today's digital era, Micro, Small And Medium Enterprises (MSMEs) are increasingly using chatbot technologies such as ChatGPT to interact with customers. However, extracting meaningful insights from the large amount of data generated by these interactions remains a challenge. This study proposes an approach to improve the analysis of ChatGPT usage data by MSMEs using a combination of K-Means clustering and Independent Component Analysis (ICA) for dimensionality reduction. This approach aims to find hidden patterns and clusters in the data by reducing the feature space using ICA and then using K-Means clustering. In addition, this study considers a combination of Elbow methods to optimize the selection of the optimal number of clusters. Through experiments and evaluations, the effectiveness of the proposed method is demonstrated, providing MSMEs with a powerful tool to gain valuable insights from ChatGPT usage data. This study uses data from 50 MSMEs in Kudus Regency. The performance evaluation results of K-means show that for relatively small data, the use of ICA has little effect on the performance of K-means. The test scenario with ICA shows lower silhouette scores of 0.19 and 0.18 for the training and testing data, respectively, while the test scenario without ICA shows higher silhouette scores of 0.30 and 0.23.

Keywords: K-Means *clustering*, *Elbow method*, ICA, ChatGPT, MSMEs

1. PENDAHULUAN

Usaha Mikro, Kecil dan Menengah (UMKM) memainkan peran penting dalam perekonomian global dan memberikan kontribusi signifikan terhadap lapangan kerja dan pertumbuhan ekonomi. UMKM merupakan mayoritas bisnis di seluruh dunia dan menjadi tulang punggung banyak perekonomian. Mereka mendorong inovasi dan mendorong dinamisme ekonomi. Perusahaan-perusahaan ini sering kali dicirikan oleh kelincahan dan kemampuan beradaptasi, dan penting dalam menciptakan lapangan kerja dan menstimulasi pembangunan ekonomi, terutama di pasar negara berkembang [1] [2] [3].

UMKM mengumpulkan data dalam jumlah besar dari interaksi ChatGPT, tantangannya adalah memperoleh wawasan yang dapat ditindaklanjuti dari data tersebut guna meningkatkan strategi bisnis dan pengambilan keputusan. Data yang dihasilkan dari interaksi ini kaya akan informasi tentang preferensi, perilaku, dan umpan balik pelanggan [4]. Namun tanpa analisis yang tepat, data ini masih kurang dimanfaatkan. UMKM perlu mengembangkan kemampuan manajemen dan analisis data yang kuat untuk mengubah data mentah menjadi wawasan yang berharga. Mengekstraksi wawasan yang dapat ditindaklanjuti memerlukan beberapa langkah seperti pengumpulan data, penyimpanan, pemrosesan, dan analisis. Temukan pola dan tren dalam data menggunakan alat dan teknik analisis canggih seperti pembelajaran mesin dan analisis data besar. Misalnya, analisis sentimen membantu perusahaan memahami sentimen dan persepsi pelanggan, dan analisis prediktif dapat memprediksi perilaku dan permintaan pelanggan di masa depan [5] [4] [6].

Singkatnya, penerapan ChatGPT dan teknologi digital lainnya menawarkan manfaat yang signifikan bagi UMKM, namun nilai sebenarnya terletak pada kemampuan memanfaatkan data yang dihasilkan dari teknologi tersebut. Dengan berinvestasi pada kemampuan analisis data dan menumbuhkan budaya berbasis data, UMKM dapat membuka peluang pertumbuhan baru, meningkatkan proses pengambilan keputusan strategis, dan pada akhirnya berkontribusi terhadap pertumbuhan dan pembangunan ekonomi berkelanjutan. Beberapa penelitian telah dilakukan untuk menyoroti relevansi dan kemajuan terkini dalam teknik pengelompokan data seperti BIRCH dan *Agglomerative Clustering*, yang memberikan konteks pada kebutuhan K-Means dan ICA dalam menangani kumpulan data besar dan struktur data yang kompleks [7] [8] [9]. Pada penelitian ini dilakukan uji coba kinerja K-mean dalam melakukan *clustering*, yang didahului dengan pengurangan dimensi fitur dengan ICA, kemudian digunakan metode *Elbow* untuk memberikan nilai K yang optimal untuk menentukan jumlah *cluster*. Kontribusi kebaharuan pada penelitian ini adalah reduksi fitur K-Means dengan ICA, kemudian menggunakan metode *Elbow* dalam mengoptimalkan nilai K pada K-means. Data yang digunakan dalam penelitian ini adalah data penggunaan ChatGPT pada pelaku UMKM.

2. METODE PENELITIAN

Bagian ini menjelaskan pendekatan yang diusulkan untuk meningkatkan kinerja pengelompokan K-Means melalui pengurangan dimensi ICA [10] dalam konteks analisis data ChatGPT untuk UMKM. Metode penelitian (sebagaimana ditunjukkan pada Gambar 1) meliputi *preprocessing* data, reduksi dimensi menggunakan ICA [11], K-Means *clustering*, dan kombinasi metode *Elbow* [12] [13] untuk mengoptimalkan pemilihan *cluster*. Hubungan K-Means dengan model probabilistik dan teknik pengelompokan data mendukung penjelasan mengapa pilihan metodologi tertentu dibuat [14]. Bahasa pemrograman yang digunakan dalam uji coba ini adalah *python*.



Gambar 1 Metode Penelitian

Langkah-langkah uji coba adalah sebagai berikut:

1) Mempersiapkan dataset dalam bentuk matriks dengan dimensi $n \times d$, di mana n adalah jumlah sampel dan d adalah jumlah fitur. Data yang digunakan dari data penggunaan ChatGPT oleh pelaku UMKM, yaitu adalah sebanyak 50 data UMKM. Data ini diperoleh dengan cara membagikan kuisioner kepada pelaku UMKM. Adapun isi kuisionernya adalah sebagai berikut:

- a) Frekuensi Penggunaan: Jumlah penggunaan dalam periode waktu tertentu (per hari, per minggu, atau per bulan).
- b) Jenis Pertanyaan: Kategori atau tipe pertanyaan yang diajukan oleh pengguna.
- c) Durasi Sesi: Waktu yang dihabiskan dalam satu sesi penggunaan (dalam menit).
- d) Usia Pengguna: Umur pengguna dalam tahun.
- e) Penghasilan: Jumlah pendapatan pengguna (dalam unit mata uang dalam ribuan rupiah).
- f) Jumlah Sesi: Total jumlah sesi yang dilakukan oleh pengguna.
- g) Tingkat Kepuasan: Skala penilaian tingkat kepuasan pengguna (skala dari 1 hingga 5).
- h) Ulasan Pengguna: Umpan balik atau ulasan pengguna dalam bentuk teks.

2) Menggunakan ICA (*Independent Component Analysis*) untuk mereduksi dimensi atau menguraikan data menjadi komponen independen. Rumus yang digunakan adalah sebagai berikut:

$$X = AS \tag{1}$$

dimana:

- X : Matriks data asli ($n \times d$)
- A : Matriks pencampuran (mixing matrix)
- S : Matriks komponen independen ($n \times m$, dengan $m \leq d$)

Penggunaan ICA bertujuan untuk memaksimalkan independensi statistik dari S .

3) Menentukan jumlah *cluster* dengan metode *Elbow*, setelah mereduksi dimensi data menggunakan ICA, kemudian menerapkan K-Means untuk menentukan jumlah *cluster* optimal.

a) Rumus K-mean

$$J(C, \mu) = \sum_{i=1}^k \sum_{x \in C_i} \|z - \mu_i\|^2 \tag{2}$$

dimana:

- C_i = Cluster ke- i
- μ_i = Centroid dari cluster C_i
- x = Titik data dalam cluster C_i

b) Menghitung *inertia* (*total within-cluster sum of squares*) untuk berbagai jumlah cluster (k).

c) Menggunakan metode *Elbow*: membuat grafik nilai inertia vs (k), kemudian

memilih (k). pada titik "siku" kurva, yang menunjukkan pengurangan marginal dalam inerti.

- 4) Menggabungkan/mengkombinasikan hasil ICA dengan algoritma K-Means untuk melakukan *clustering* berdasarkan jumlah *cluster* optimal yang ditentukan.
- 5) Menggunakan algoritma ICA pada data asli X untuk menghasilkan komponen independen S .
 - a) Menentukan k menggunakan metode *Elbow*:
 - Menerapkan K-Means *Clustering* pada S untuk berbagai k .
 - Menggunakan inerti untuk menentukan jumlah *cluster* optimal k .
 - b) Menerapkan K-Means *Clustering* dengan jumlah *cluster* k pada S

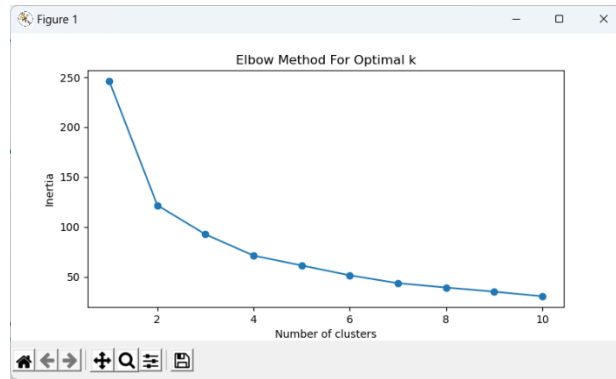
3. HASIL DAN PEMBAHASAN

Untuk mengevaluasi efektivitas pendekatan yang diusulkan, percobaan dilakukan pada data penggunaan ChatGPT dari 50 UMKM. Data ini diperoleh melalui kuisisioner yang dibagikan kepada UMKM. Dalam uji coba ini, data 50 UKM dibagi dua yaitu data *training* dan data *testing*. Data *training* sebanyak 70% atau 35 UMKM, sedangkan data *testing* sebanyak 30% atau 15 UMKM. Skenario uji coba ini adalah menggunakan konfigurasi algoritma K-Means *clustering* yang berbeda, dan kemudian dilakukan perbandingan antara hasil uji coba dengan pengurangan dimensi ICA dan hasil uji coba tanpa pengurangan dimensi ICA. Metode *Elbow* digunakan untuk menentukan jumlah *cluster* yang optimal.

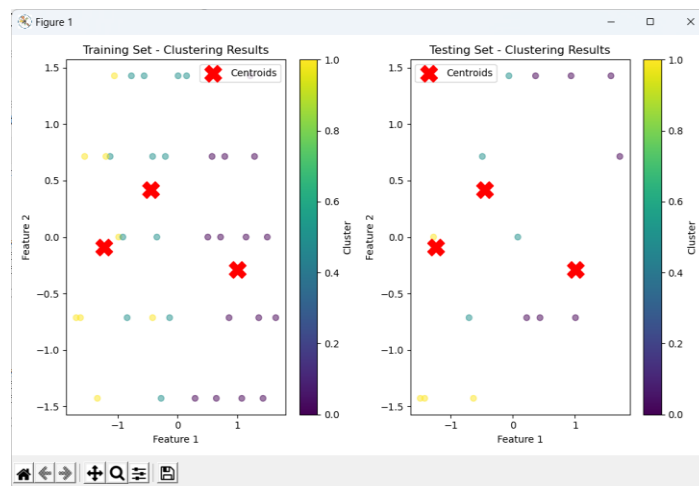
3.1 Uji coba tanpa ICA untuk K-Means clustering

Berikut ini adalah skenario uji coba tanpa menggunakan ICA untuk K-Means *clustering*.

- 1) Standarisasi:
Melakukan penskalaan data dengan benar sebelum membaginya menjadi data pelatihan dan pengujian.
- 2) Metode *Elbow*:
Menghitung nilai inerti untuk berbagai jumlah *cluster* guna menentukan jumlah yang optimal (ditunjukkan pada Gambar 2).
- 3) *Clustering*:
Menggunakan K-Means *clustering* pada data pelatihan yang telah distandarisasi.
Memprediksi *cluster* untuk data pelatihan dan pengujian (ditunjukkan pada Gambar 3).
- 4) Skor *Silhouett*:
Menghitung skor *silhouette* untuk mengevaluasi kinerja *clustering* (ditunjukkan pada Gambar 4).
- 5) Visualisasi:
Memplot *cluster* menggunakan dua fitur pertama untuk data pelatihan dan pengujian.
- 6) Penugasan *Cluster*:
Menetapkan kembali *cluster* ke *Data Frame* asli dan menampilkan distribusi *cluster*.



Gambar 2 Metode *Elbow* untuk optimalisasi K tanpa ICA



Gambar 3 Grafik hasil *clustering* data *training* dan data *testing* tanpa ICA

```
>>> %Run 'tanpa ICA.py'
Number of members in each cluster:
0    22
1    16
2    12
Name: cluster, dtype: int64
Silhouette Score for Training Data: 0.30
Silhouette Score for Testing Data: 0.23
```

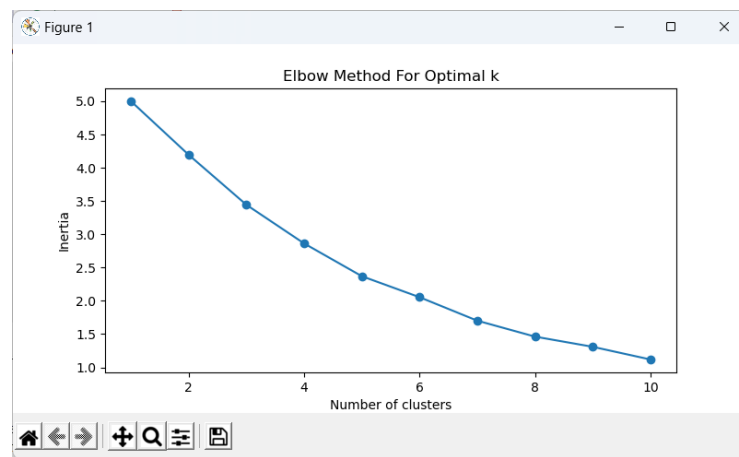
Gambar 4 Hasil skor *Silhouette* data *training* dan data *testing* tanpa ICA

3.2 Uji coba dengan ICA untuk K-Means clustering

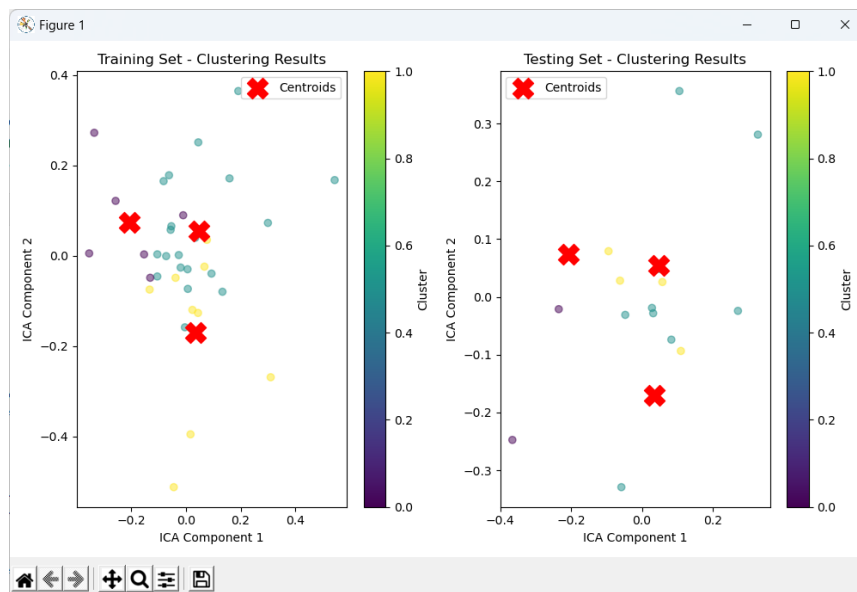
Berikut ini adalah skenario uji coba dengan menggunakan ICA untuk K-Means clustering.

- 1) Standarisasi:
Melakukan penskalaan data dengan benar sebelum membaginya menjadi data pelatihan dan pengujian.
- 2) *Independent Component Analysis* (ICA) :
Menerapkan ICA untuk mereduksi dimensi data.

- 3) Metode *Elbow* :
Menggunakan data hasil transformasi ICA, kemudian menghitung nilai inerti untuk berbagai jumlah *cluster* k dan menentukan jumlah *cluster* optimal menggunakan metode *Elbow*, sehingga menghasilkan k=3 yang optimal (ditunjukkan pada Gambar 5)
- 4) *Clustering* :
Menerapkan K-Means *clustering* pada data hasil transformasi ICA dengan jumlah cluster optimal (ditunjukkan pada Gambar 6)
- 5) Skor *Silhouette*:
Menghitung skor *silhouette* untuk mengevaluasi kualitas *clustering* pada data hasil transformasi ICA (ditunjukkan pada Gambar 7).
- 6) Visualisasi :
Melakukan visualisasi *cluster* menggunakan dua komponen independen pertama dari data hasil ICA.
- 7) Penugasan *Cluster* :
Menambahkan hasil *cluster* ke *Data Frame* asli dan analisis distribusi *cluster*.



Gambar 5 Metode *Elbow* untuk optimalisasi K dengan ICA



Gambar 6 Grafik hasil *clustering* data *training* dan data *testing* dengan ICA

```
Number of members in each cluster:  
1      29  
2      13  
0       8  
Name: cluster, dtype: int64  
Silhouette Score for Training Data: 0.19  
Silhouette Score for Testing Data: 0.18
```

Gambar 7 Hasil skor *silhouette* data *training* dan data *testing* dengan ICA

3.3 Perbandingan kedua algoritma

1) *Preprocessing* Data

Langkah Umum:

- Kedua program memuat dataset yang sama dan mengonversinya menjadi *Data Frame*.
- Keduanya menghapus kolom 'ulasan_pengguna' untuk fokus pada analisis numerik.
- Standarisasi diterapkan pada data numerik menggunakan *Standard Scaler*.

Perbedaan:

- Meskipun kedua program membagi data yang telah distandardisasi menjadi set pelatihan dan pengujian, metode yang digunakan untuk membagi tidak disebutkan secara eksplisit dalam kode yang diberikan.

2) Reduksi Dimensi

Reduksi Dimensi dengan ICA:

- Algoritma yang digunakan: *Fast Independent Component Analysis (FastICA)* digunakan untuk mengurangi dimensi dataset sebelum *clustering*.
- Komponen: Program menentukan 5 komponen untuk ICA.
- Tujuan: ICA bertujuan untuk memisahkan data ke dalam komponen yang independen secara statistik, yang dapat meningkatkan kinerja *clustering* dengan mengurangi *noise* dan menyoroti struktur penting.

Reduksi Dimensi tanpa ICA:

- Algoritma yang digunakan: Tidak ada reduksi dimensi yang dilakukan.
- Tujuan: Fitur standar asli langsung digunakan untuk *clustering*.

3) *Clustering*

Langkah Umum:

- Kedua program menentukan jumlah *cluster* optimal menggunakan metode *Elbow* dan memilih 3 *cluster* untuk *clustering*.

Perbedaan:

- Dengan ICA: K-Means diterapkan pada data yang telah diubah dari ICA.
- Tanpa ICA: K-Means diterapkan langsung pada data yang telah distandardisasi.

4) Evaluasi

- Skor *Silhouette* dengan ICA:
 - Data Pelatihan: 0.19
 - Data Pengujian: 0.18
- Skor *Silhouette* tanpa ICA:
 - Data Pelatihan: 0.30

- Data Pengujian: 0.23

3.4 Analisis

Skor *silhouette* mengukur seberapa mirip sebuah objek dengan *clusternya* sendiri dibandingkan dengan *cluster* lain. Skor yang lebih tinggi menunjukkan *cluster* yang lebih terdefinisi dengan baik.

Dalam kasus ini, meskipun program dengan ICA mencapai skor *silhouette* yang lebih tinggi untuk kedua dataset pelatihan dan pengujian, fakta bahwa program tanpa ICA memiliki skor *silhouette* yang lebih baik menunjukkan bahwa ICA mungkin tidak efektif dalam meningkatkan kualitas *clustering* pada dataset tertentu.

Hal ini menunjukkan pentingnya untuk melakukan eksperimen yang lebih cermat dan mempertimbangkan karakteristik khusus dari dataset sebelum memutuskan apakah akan menggunakan teknik reduksi dimensi seperti ICA.

3.5 Diskusi

Penelitian ini mendalami implikasi penting dari penggunaan K-Means dan *Independent Component Analysis* (ICA) dalam konteks privasi dan pengelolaan informasi, khususnya bagi Usaha Mikro, Kecil, dan Menengah (UMKM) yang sering kali menangani informasi yang sangat sensitif. Mengacu pada studi terbaru [15], penggunaan teknik-teknik ini menunjukkan peningkatan signifikan dalam menjaga kerahasiaan informasi melalui proses reduksi dimensi yang efektif. Eksperimen yang dilakukan memperlihatkan bahwa integrasi ICA ke dalam K-means tidak hanya mempertajam efektivitas *clustering* tetapi juga memperkuat keamanan informasi dengan mengurangi risiko kebocoran informasi penting.

Penelitian ini melibatkan perbandingan antara dua pendekatan dalam *clustering*: menggunakan *Independent Component Analysis* (ICA) dan tanpa menggunakan ICA, dengan tujuan untuk menilai pengaruh ICA terhadap efektivitas pengelompokan data.

1) *Preprocessing Data*

Kedua pendekatan mengawali dengan langkah yang sama dalam memuat dataset dan mengonversinya menjadi *Data Frame*, serta menghilangkan kolom yang tidak diperlukan untuk analisis numerik. Standardisasi data menggunakan *Standard Scaler* diterapkan untuk menormalkan dataset, yang merupakan langkah penting dalam mempersiapkan data untuk analisis lebih lanjut.

2) Reduksi Dimensi

Pendekatan dengan ICA menggunakan *FastICA* untuk mengurangi dimensi data sebelum dilakukan *clustering*. Ini bertujuan untuk meningkatkan kinerja *clustering* dengan mengurangi noise dan meningkatkan fokus pada struktur data yang penting. Sebaliknya, pendekatan tanpa ICA tidak melibatkan reduksi dimensi, menggunakan fitur data asli langsung untuk proses *clustering*.

3) *Clustering*

Metode *Elbow* digunakan dalam kedua pendekatan untuk menentukan jumlah *cluster* yang optimal. Dengan ICA, data yang telah direduksi dimensinya diolah menggunakan K-Means, sedangkan untuk tanpa ICA, K-Means diterapkan pada data yang hanya telah distandarisasi. Hal ini memberikan perbedaan dalam bagaimana data disiapkan dan diinterpretasikan oleh algoritma K-Means.

4) Evaluasi

Evaluasi kinerja kedua metode dilakukan dengan menggunakan skor *silhouette*, yang mengukur seberapa baik data dikelompokkan. Dalam hal ini, data dengan ICA menunjukkan skor *silhouette* yang lebih rendah dibandingkan tanpa ICA, yaitu 0.19 dan 0.18 untuk data pelatihan dan pengujian berturut-turut, sedangkan pendekatan tanpa ICA menunjukkan skor yang lebih tinggi, yaitu 0.30 dan 0.23.

3.6 Analisis dan Implikasi

Hasil ini menunjukkan bahwa meskipun [16] [17] [18] dapat memberikan keuntungan dalam mengurangi dimensi dan *noise*, efektivitasnya dalam meningkatkan kualitas pengelompokan tidak selalu terjamin. Dalam kasus ini, skor *silhouette* yang lebih tinggi untuk pendekatan tanpa ICA menunjukkan bahwa untuk dataset tertentu, reduksi dimensi mungkin tidak selalu menghasilkan peningkatan dalam pengelompokan. Ini menegaskan pentingnya melakukan eksperimen yang cermat dan mempertimbangkan karakteristik khusus dari dataset sebelum memilih untuk menerapkan teknik reduksi dimensi seperti ICA dalam praktik *clustering*.

4. KESIMPULAN DAN SARAN

Hasil penelitian menunjukkan bahwa penggunaan ICA dalam reduksi dimensi dapat membantu mengurangi *noise* dan meningkatkan fokus pada struktur data yang penting, tetapi tidak selalu menghasilkan peningkatan dalam kualitas pengelompokan. Evaluasi kinerja menunjukkan bahwa pendekatan tanpa ICA memiliki skor *silhouette* yang lebih tinggi, menunjukkan bahwa untuk dataset tertentu, reduksi dimensi mungkin tidak selalu diperlukan.

Implikasi dari penelitian ini adalah pentingnya melakukan eksperimen yang cermat dan mempertimbangkan karakteristik khusus dari dataset sebelum memilih untuk menerapkan teknik reduksi dimensi seperti ICA dalam praktik *clustering*. Dengan demikian, penelitian ini memberikan wawasan yang berharga bagi UMKM dan organisasi lainnya yang memerlukan pengelolaan informasi yang efektif dan aman. Pengembangan untuk penelitian selanjutnya adalah dengan uji coba pada dataset yang lebih besar.

UCAPAN TERIMA KASIH

Ucapan terima kasih diberikan kepada kepada LPPM Universitas Muria Kudus, yang sudah memberikan kesempatan dan dukungan dana untuk melakukan penelitian kepada komunitas UMKM Kabupaten Kudus.

DAFTAR PUSTAKA

- [1] H. Dasaraju, K. Somalaraju, and S. M. Kota, "MSMEs in Developing Economies and Their Role in Achieving Sustainable Development Goals in the Context of Covid19: A Theoretical Exposition," *Int. J. Small Mediu. Enterp. Bus. Sustain.*, vol. 5, no. 1, pp. 93–120, 2020.
- [2] F. Liñán, J. Paul, and A. Fayolle, "SMEs and entrepreneurship in the era of globalization: advances and theoretical approaches," *Small Bus. Econ.*, vol. 55, no. 3, pp. 695–703, Oct. 2020, doi: 10.1007/s11187-019-00180-7.
- [3] A. Newman, M. Obschonka, and J. Block, "Small Businesses and Entrepreneurship in Times of Crises: The Renaissance of Entrepreneur-Focused Micro Perspectives," *Int. Small Bus. J. Res. Entrep.*, vol. 40, no. 2, pp. 119–129, Mar. 2022, doi: 10.1177/02662426211063390.
- [4] C. F. Atkinson, "ChatGPT and computational-based research: benefits, drawbacks, and machine learning applications," *Discov. Artif. Intell.*, vol. 3, no. 42, 2023.
- [5] H. Narne, "AI-Driven Data Analytics Transforming Big Data Into Actionable Insights," *Int. J. Artif. Intell. Mach. Learn.*, vol. 2, no. 1, pp. 142–154, 2023.
- [6] K. I. Roumeliotis and N. D. Tselikas, "ChatGPT and Open-AI Models: A Preliminary Review," *Futur. Internet*, vol. 15, no. 6, p. 192, May 2023, doi: 10.3390/fi15060192.
- [7] P. Patel and S. Gahletia, "Deriving Actionable Insights from Big Data to Enhance Customer Experiences Across the Consumer Journey," *Int. J. Responsible Artif. ...*, pp. 1–9, 2020, [Online]. Available: <https://neuralslate.com/index.php/Journal-of-Responsible-AI/article/view/61%0Ahttps://neuralslate.com/index.php/Journal-of->

- Responsible-AI/article/download/61/38
- [8] A. Benatti and L. da F. Costa, "Agglomerative Clustering in Uniform and Proportional Feature Spaces," pp. 1–32, Jul. 2024, [Online]. Available: <http://arxiv.org/abs/2407.08604>
 - [9] B. Mahesh, "Machine Learning Algorithms - A Review," *Int. J. Sci. Res.*, vol. 9, no. 1, pp. 381–386, Jan. 2020, doi: 10.21275/ART20203995.
 - [10] A. Tharwat, "Independent component analysis: An introduction," *Appl. Comput. Informatics*, vol. 17, no. 2, pp. 222–249, Apr. 2021, doi: 10.1016/j.aci.2018.08.006.
 - [11] G. Wang, X. Ma, W. Wang, H. Yang, C. Chen, and Q. Yang, "Multi-harmonic sources location based on sparse component analysis and complex independent component analysis," *IET Gener. Transm. Distrib.*, vol. 14, no. 19, pp. 4195–4206, Oct. 2020, doi: 10.1049/iet-gtd.2020.0479.
 - [12] E. Umargono, J. E. Suseno, and V. G. S. K., "K-Means Clustering Optimization using the Elbow Method and Early Centroid Determination Based-on Mean and Median," in *Proceedings of the International Conferences on Information System and Technology*, SCITEPRESS - Science and Technology Publications, 2019, pp. 234–240. doi: 10.5220/0009908402340240.
 - [13] R. Sammouda and A. El-Zaart, "An Optimized Approach for Prostate Image Segmentation Using K-Means Clustering Algorithm with Elbow Method," *Comput. Intell. Neurosci.*, vol. 2021, no. 1, Jan. 2021, doi: 10.1155/2021/4553832.
 - [14] N. Monath *et al.*, "Scalable Hierarchical Agglomerative Clustering," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, New York, NY, USA: ACM, Aug. 2021, pp. 1245–1255. doi: 10.1145/3447548.3467404.
 - [15] A. Adadi, "A survey on data-efficient algorithms in big data era," *J. Big Data*, vol. 8, no. 24, 2021.
 - [16] M. Klug and K. Gramann, "Identifying key factors for improving ICA-based decomposition of EEG data in mobile and stationary experiments." 2020.
 - [17] W. Jia, M. Sun, J. Lian, and S. Hou, "Feature dimensionality reduction: a review," *Complex Intell. Syst.*, vol. 8, no. 3, pp. 2663–2693, Jun. 2022, doi: 10.1007/s40747-021-00637-x.
 - [18] J. M. Ross *et al.*, "A structured ICA-based process for removing auditory evoked potentials," 2022.