

# ANALISIS ALGORITMA K-MEANS DAN FUZZY C-MEANS UNTUK PERSENTASE RATA-RATA PENGELUARAN UNTUK MAKANAN DAN BUKAN MAKANAN

**Fabian Darrell Widyadhana Reswara**

Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara,  
Jl. Letjen S. Parman no. 1, Jakarta, 11440, Indonesia  
E-mail: [fabian.535210062@stu.untar.ac.id](mailto:fabian.535210062@stu.untar.ac.id)

## ABSTRAK

Pengelompokan data adalah salah satu metode penting dalam analisis data, terutama untuk memahami pola dan hubungan dalam data sosial-ekonomi. Penelitian ini membandingkan performa algoritma *K-Means* dan *Fuzzy C-Means* dalam mengelompokkan persentase rata-rata pengeluaran per kapita per bulan untuk makanan dan bukan makanan di daerah perkotaan. Dataset yang digunakan terdiri dari beberapa wilayah perkotaan dengan variabel pengeluaran makanan dan bukan makanan sebagai fitur utama. Kinerja kedua algoritma dievaluasi menggunakan metrik seperti nilai rata-rata *silhouette*, waktu komputasi, dan interpretasi hasil kluster. Hasil penelitian menunjukkan bahwa algoritma *K-Means* secara keseluruhan menghasilkan hasil yang lebih baik dibandingkan dengan algoritma *Fuzzy C-Means*. Rata-rata *Silhouette Score* terbesar terdapat pada algoritma *K-Means* dan *Fuzzy C-Means* sebesar 0.873 dengan kluster = 2. Namun pada kluster lain, algoritma *K-Means* memiliki hasil yang lebih bagus.

**Kata kunci**—Persentase Pengeluaran, *K-Means*, *Fuzzy C-Means*

## ABSTRACT

Data clustering is one of the important methods in data analysis, especially for understanding patterns and relationships in socio-economic data. This study compares the performance of *K-Means* and *Fuzzy C-Means* algorithms in clustering the percentage of average monthly per capita expenditure on food and non-food in urban areas. The data set consists of several urban areas with food and non-food expenditure variables as the main features. The performance of both algorithms was evaluated using metrics such as average silhouette value, computation time, and interpretation of cluster results. The results show that the *K-Means* algorithm overall produces better results than the *Fuzzy C-Means* algorithm. The largest average Silhouette Score is found in the *K-Means* and *Fuzzy C-Means* algorithm of 0.873 with cluster = 2. But in other clusters, the *K-Means* algorithm has better results.

**Keywords**— Spending Percentage, *K-Means*, *Fuzzy C-Means*

## 1. PENDAHULUAN

Dalam era informasi saat ini, data memiliki peran penting dalam memberikan wawasan yang dapat digunakan untuk pengambilan keputusan strategis. Salah satu bentuk data yang signifikan dalam analisis sosial-ekonomi adalah data pengeluaran per kapita, yang mencerminkan tingkat kesejahteraan masyarakat. Analisis data ini, khususnya yang berkaitan dengan makanan dan bukan makanan di daerah perkotaan, dapat memberikan gambaran tentang pola konsumsi masyarakat yang dapat digunakan untuk merumuskan kebijakan ekonomi yang lebih efektif. *Clustering* adalah teknik *unsupervised learning* yang bertujuan untuk mengelompokkan data berdasarkan kesamaan atribut. Proses ini bertujuan membagi data ke dalam beberapa kelompok atau kluster, di mana data dalam satu kluster memiliki tingkat kemiripan yang tinggi, sedangkan data antar kluster memiliki kesamaan yang rendah [1]. *Clustering* merupakan metode yang digunakan untuk mengelompokkan data dengan karakteristik serupa [2]. Pendekatan ini bekerja dengan mengelompokkan data dalam satu set berdasarkan sifat atau karakteristik yang dimiliki setiap data [3]. *Algoritma clustering* dapat

diaplikasikan di berbagai bidang, misalnya dalam pemasaran untuk mengidentifikasi kelompok pelanggan dengan perilaku serupa berdasarkan data historis dan sifat pelanggan [4]

Dua algoritma *clustering* yang paling sering digunakan adalah *K-Means* dan *Fuzzy C-Means*. Meskipun keduanya berbasis pada optimasi menggunakan centroid, pendekatan yang diambil berbeda. *K-Means* menggunakan metode pembagian tegas (*hard clustering*), sedangkan *Fuzzy C-Means* memungkinkan data menjadi anggota beberapa kluster secara bersamaan (*soft clustering*). Evaluasi hasil *clustering* akan dilakukan menggunakan Silhouette Score, yang mengukur tingkat keseragaman dalam setiap kluster, dan *Davies-Bouldin Index* (DBI), yang mengevaluasi tingkat pemisahan antar kluster. Penelitian ini bertujuan untuk membandingkan performa algoritma *K-Means* dan *Fuzzy C-Means* dalam *clustering* data pengeluaran per kapita untuk makanan dan bukan makanan di daerah perkotaan, serta mengidentifikasi kelebihan dan kekurangan masing-masing algoritma.

## 2. METODE PENELITIAN

Penelitian ini bertujuan untuk mengelompokkan data ke dalam kelompok atau kluster berdasarkan kesamaan atau kemiripan data. Dalam penelitian ini, akan dibandingkan metode *clustering* antara algoritma *K-Means* dan *Fuzzy C-Means*. *Silhouette score* dan *Davies-Bouldin Index* akan menjadi pembantu untuk menilai mana algoritma yang lebih baik.

### 2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari sumber statistik resmi yaitu Badan Pusat Statistik (BPS), yang mencakup persentase rata-rata pengeluaran per kapita untuk makanan dan bukan makanan di beberapa wilayah perkotaan melalui link <https://www.bps.go.id/id/statistics-table/1/OTQwIzE=/persentase-rata-rata-pengeluaran-per-kapita-sebulan-untuk-makanan-dan-bukan-makanan-di-daerah-perkotaan-menurut-provinsi--2007-2024.html>. Data ini mengandung persentase rata-rata pengeluaran per kapita sebulan untuk makanan dan bukan makanan setiap provinsi dari Indonesia. Data ini merupakan data dari tahun 2007 sampai dengan tahun 2024.

### 2.2 Algoritma K-Means

K-Means adalah algoritma *clustering* berbasis *centroid* yang digunakan untuk mengelompokkan data ke dalam sejumlah kluster berdasarkan kesamaan atribut. Algoritma *K-Means* merupakan salah satu metode dalam teknik klustering yang populer [5]. Algoritma ini termasuk dalam kategori *unsupervised learning* karena tidak memerlukan label data untuk proses pelatihannya. Tujuannya adalah meminimalkan variansi dalam kluster dan memaksimalkan jarak antar kluster. K-Means merupakan teknik analisis data yang dikenal dalam data mining, digunakan untuk memodelkan data secara *unsupervised*, dan menjadi metode partisi untuk mengelompokkan data [6]. Algoritma ini dapat diterapkan pada berbagai jenis data, baik numerik maupun kategorikal [7]. Kluster sendiri adalah kumpulan objek data yang memiliki kesamaan dalam satu grup, sementara objek dengan karakteristik berbeda berada di kelompok lain [8]. Dengan kemampuan untuk mempartisi data ke dalam kluster [9]. *K-means* salah satu metode data *clustering* non hirarki yang berusaha mempartisi data yang ada di dalam bentuk satu atau lebih cluster atau kelompok [10]. Metode ini akan mengelompokkan data ke dalam kelompok yang memiliki karakteristik data yang sama sedangkan data dengan karakteristik yang berbeda akan ditambahkan ke kelompok lainnya [11]. Algoritma *K-Means* bekerja dengan membagi data ke dalam k kluster, di mana nilai k ditentukan oleh pengguna. Proses algoritma dimulai dengan pembentukan kluster awal, yang kemudian diperbaiki secara iteratif hingga kluster stabil dan tidak ada perubahan lebih lanjut [12]. Dalam menentukan kedekatan antar data, algoritma ini menggunakan matrik jarak, seperti jarak Euclidean. Berikut ini adalah langkah-langkah algoritma *k-means*:

1. Inialisasi Centroid
2. Assign Data ke Kluster

3. Pembaruan Centroid
4. Iterasi

Kelebihan algoritma *K-Means* adalah efisiensi komputasi, sederhana, dan skalabilitas dimana algoritma ini dapat diterapkan pada dataset yang besar dengan dimensi yang rendah. Kekurangan algoritma *K-Means* adalah sensitif terhadap inisialisasi *centroid*, kesulitan dengan kluster yang tidak berbentuk bulat, rentan terhadap outlier, dan harus menentukan jumlah kluster yang tepat. Algoritma *K-Means* dapat digunakan untuk segmentasi pasar, pengolahan citra, dan analisis data sosial. *K-Means* adalah salah satu algoritma yang sangat populer karena kecepatan dan kesederhanaannya, meskipun memiliki keterbatasan pada data dengan struktur yang lebih kompleks atau ketika *outlier* sangat mempengaruhi hasil clustering.

### 2.3 Algoritma Fuzzy C-Means

*Fuzzy C-Means* (FCM) adalah algoritma *clustering* berbasis *centroid* yang mengadopsi pendekatan *soft clustering*, memperluas konsep *K-Means*. Algoritma ini bekerja dengan menggunakan derajat keanggotaan untuk menentukan kluster yang sesuai bagi setiap data [13]. diperkenalkan oleh Dunn pada tahun 1973 dan disempurnakan oleh Bezdek pada tahun 1981, FCM merupakan salah satu metode clustering berbasis fuzzy [14]. Tidak seperti *K-Means* yang membagi data secara eksklusif ke satu kluster (*hard clustering*), FCM memungkinkan setiap data memiliki tingkat keanggotaan pada beberapa kluster sekaligus, menjadikannya lebih cocok untuk data dengan batas kluster yang samar atau tidak tegas. Selain itu, FCM dapat digabungkan dengan metode lain seperti RFM, AHP, SAW, dan TOPSIS untuk menghasilkan rekomendasi yang mendukung proses pengambilan keputusan [15]. *Fuzzy C-Means* mengelompokkan data dengan meminimalkan fungsi objektif berdasarkan jarak data ke *centroid cluster*, sambil mempertimbangkan derajat keanggotaan data dalam kluster. Metode *Fuzzy C-Means* mengacu pada seberapa besar kemungkinan data dapat menjadi anggota kluster [16]. Derajat keanggotaan ini dihitung menggunakan nilai probabilitas, yang disebut matriks keanggotaan ( $U$ ). Berikut ini adalah langkah-langkah algoritma *Fuzzy C-Means*:

1. Inisialisasi
2. Pembaruan Centroid
3. Pembaruan Derajat Keanggotaan
4. Iterasi

Kelebihan *Fuzzy C-Means* adalah fleksibilitas, presisi lebih tinggi, dan bisa digunakan untuk aplikasi seperti segmentasi citra serta clustering data kompleks. Kekurangan *Fuzzy C-Means* merupakan kompleksitas komputasi, sensitive terhadap parameter Fuzzy, dan rentan terhadap outlier. *Fuzzy C-Means* bisa digunakan untuk segmentasi citra, analisis data sosial dan ekonomi, serta identifikasi pola genetik atau klasifikasi penyakit.

### 2.4 Evaluasi Model

Berikut ini adalah nilai-nilai yang digunakan untuk evaluasi model:

#### 1. *Silhouette Score*:

*Silhouette Score* merupakan metrik yang digunakan untuk mengevaluasi kualitas hasil *clustering*. Metrik ini menentukan seberapa efektif suatu data dikelompokkan ke dalam kluster tertentu dengan membandingkan jaraknya terhadap data lain dalam kluster yang sama serta terhadap kluster terdekat lainnya. Nilai *Silhouette Score* memberikan gambaran tentang tingkat keseragaman data dalam kluster dan sejauh mana kluster tersebut terpisah dari kluster lainnya. Untuk menilai kualitas pengelompokan, perhitungan *silhouette* dapat dilakukan untuk setiap kluster secara individual maupun untuk keseluruhan kluster hasil suatu algoritma clustering [17].

Untuk setiap data  $i$  dalam dataset, *Silhouette Score* ( $S(i)$ ) dihitung berdasarkan dua nilai utama:

- a.  $a(i)$ : Rata-rata jarak data  $i$  ke semua data lain dalam kluster yang sama (*intra-cluster distance*).
- b.  $b(i)$ : Rata-rata jarak data  $i$  ke semua data dalam kluster terdekat lainnya (*nearest-cluster distance*).

Berikut ini adalah formula untuk Silhouette Score untuk data  $i$  [18]:

$$S(i) = \frac{b - a}{\max(a(i), b(i))} \quad (1)$$

Kelebihan *Silhouette Score* adalah evaluasi kuantitatif, independent dari jumlah kluster, dan identifikasi kluster yang kurang baik. Kekurangan *Silhouette Score* adalah komputasi intensif dan kurang cocok untuk kluster non-konveks. Nilai *silhouette score* yang tinggi menandakan bahwa clustering tersebut baik, sementara nilai yang rendah menunjukkan bahwa ada penyebaran yang tidak homogen dalam satu atau lebih kelompok [19].

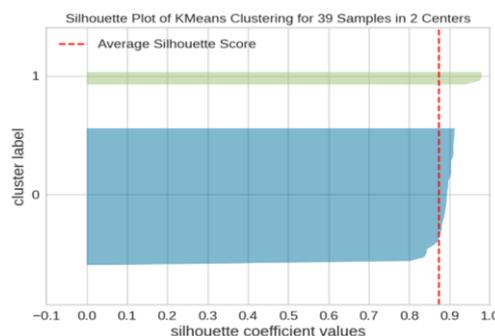
2. Davies Bouldin Index:
3. David L. Davies dan Donald W. Bouldin memperkenalkan *Davies-Bouldin Index* (DBI) pada tahun 1979 sebagai metode untuk mengoptimalkan jarak antar kluster (*inter-cluster distance*) sekaligus meminimalkan jarak antar titik dalam kluster (*intra-cluster distance*) [20]. DBI merupakan matrik evaluasi internal yang digunakan untuk menilai kualitas clustering berdasarkan tingkat keseragaman dalam kluster serta tingkat pemisahan antar kluster [21]. Pengujian menggunakan DBI membantu dalam mengevaluasi akurasi kluster yang dihasilkan dalam penelitian [22]. *Davies Bouldin Index* digunakan untuk mengukur kualitas dari hasil clustering [23]. Metrik ini menghitung rasio antara penyebaran (*scatter*) dalam kluster dan jarak antar centroid kluster, sehingga memberikan gambaran tentang seberapa baik kluster terbentuk dan terpisah satu sama lain. Selain itu, DBI sering digunakan untuk menentukan jumlah kluster optimal setelah proses clustering selesai [24]. Davies Bouldin Index didasarkan pada dua aspek utama yaitu penyebaran dalam kluster ( $S_i$ ) untuk mengukur keseragaman data dalam satu kluster dan jarak antar kluster ( $D_{i,j}$ ) untuk mengukur jarak antara centroid dua kluster yang berbeda. Berikut ini adalah rumus DBI [25]:

$$\frac{1}{k} \sum_{i=1}^k \max R_{i,j} \quad (2)$$

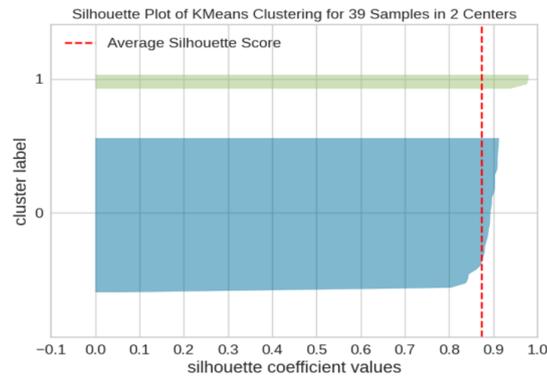
Jika DBI menunjukkan nilai yang rendah, nilai tersebut mengindikasikan kluster yang baik, dengan penyebaran dalam kluster yang kecil dan jarak antar kluster yang besar. Jika DBI menunjukkan nilai yang tinggi, nilai tersebut mengindikasikan kluster yang buruk, Dimana data dalam kluster sangat tersebar atau kluster terlalu dekat satu sama lain. Kelebihan DBI adalah otomatisasi, kesederhanaan, dan efisiensi komputasi. Kekurangan DBI adalah sensitif terhadap bentuk kluster, rentan terhadap outlier, dan tidak mempertimbangkan struktur internal kluster. DBI bisa dipakai untuk evaluasi algoritma clustering, pemilihan jumlah cluster optimal, dan validasi model clustering.

#### 4. HASIL DAN PEMBAHASAN

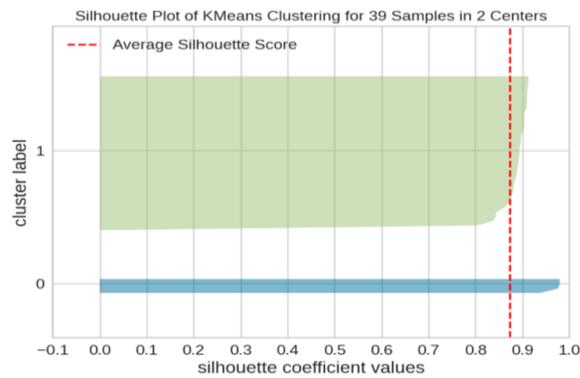
Berikut ini adalah contoh plot kluster untuk algoritma K-Means dengan beberapa kali eksperimen untuk kluster = 2.



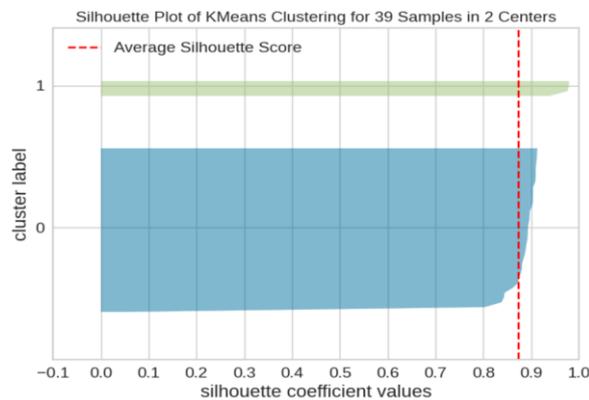
**Gambar 1.** *Silhouette plot* untuk algoritma K-Means dengan kluster = 2 untuk eksperimen 1



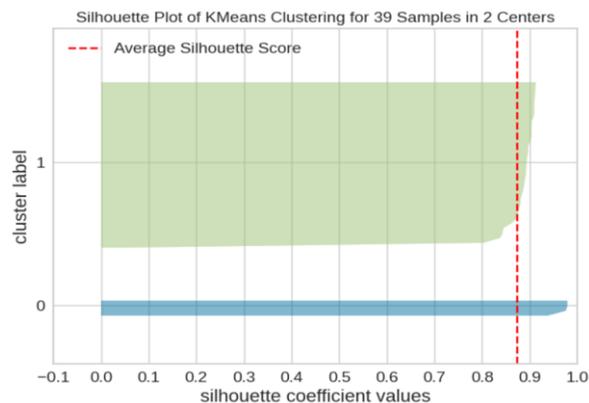
**Gambar 2.** *Silhouette plot* untuk algoritma *K-Means* dengan kluster = 2 untuk eksperimen 2



**Gambar 3.** *Silhouette plot* untuk algoritma *K-Means* dengan kluster = 2 untuk eksperimen 3



**Gambar 4.** *Silhouette plot* untuk algoritma *K-Means* dengan kluster = 2 untuk eksperimen 4



**Gambar 5.** *Silhouette plot* untuk algoritma *K-Means* dengan kluster = 2 untuk eksperimen 5

Berikut ini adalah hasil perbandingan nilai *Silhouette Score* untuk algoritma *K-Means* dan *Fuzzy C-Means* untuk jumlah kluster = 2.

**Tabel 1.** Perbandingan antar algoritma dengan kluster = 2

Eksperimen ke	Silhouette Score K-Means	Silhouette Score Fuzzy C-Means	Davies-Bouldin Index K-Means	Davies-Bouldin Index Fuzzy C-Means
1	0.873	0.873	0.111	0.111
2	0.873	0.873	0.111	0.111
3	0.873	0.873	0.111	0.111
4	0.873	0.873	0.111	0.111
5	0.873	0.873	0.111	0.111
Rata-rata	0.873	0.873	0.111	0.111

Berikut ini adalah hasil perbandingan nilai *Silhouette Score* untuk algoritma *K-Means* dan *Fuzzy C-Means* untuk jumlah kluster = 3.

**Table 2.** Perbandingan antar algoritma dengan kluster = 3

Eksperimen ke	Silhouette Score K-Means	Silhouette Score Fuzzy C-Means	Davies-Bouldin Index K-Means	Davies-Bouldin Index Fuzzy C-Means
1	0.848	0.275	0.093	1.001
2	0.848	0.847	0.093	0.094
3	0.848	0.275	0.093	1.006
4	0.848	0.847	0.093	0.094
5	0.848	0.275	0.093	1.006
Rata-rata	0.848	0.275	0.093	1.0035

Berikut ini adalah hasil perbandingan nilai *Silhouette Score* untuk algoritma *K-Means* dan *Fuzzy C-Means* untuk jumlah kluster = 4.

**Table 3.** Perbandingan antar algoritma dengan kluster = 4

Eksperimen ke	Silhouette Score K-Means	Silhouette Score Fuzzy C-Means	Davies-Bouldin Index K-Means	Davies-Bouldin Index Fuzzy C-Means
1	0.383	0.360	0.556	0.615
2	0.393	0.360	0.571	0.615
3	0.462	0.360	0.452	0.615
4	0.406	0.360	0.533	0.615
5	0.339	0.360	0.609	0.615
Rata-rata	0.361	0.36	0.5825	0.615

Berikut ini adalah hasil perbandingan nilai *Silhouette Score* untuk algoritma *K-Means* dan *Fuzzy C-Means* untuk jumlah kluster = 5.

**Tabel 4.** Perbandingan antar algoritma dengan kluster = 5

Eksperimen ke	Silhouette Score K-Means	Silhouette Score Fuzzy C-Means	Davies-Bouldin Index K-Means	Davies-Bouldin Index Fuzzy C-Means
1	0.353	0.351	0.694	0.699
2	0.353	0.351	0.694	0.699
3	0.295	0.351	0.720	0.699
4	0.354	0.351	0.694	0.699
5	0.287	0.351	0.720	0.699
Rata-rata	0.32	0.351	0.707	0.699

Berikut ini adalah hasil perbandingan nilai *Silhouette Score* untuk algoritma *K-Means* dan *Fuzzy C-Means* untuk jumlah kluster = 6

**Tabel 5.** Perbandingan antar algoritma dengan kluster = 6

Eksperimen ke	Silhouette Score K-Means	Silhouette Score Fuzzy C-Means	Davies-Bouldin Index K-Means	Davies-Bouldin Index Fuzzy C-Means
1	0.216	0.269	1.106	0.890
2	0.288	0.326	0.872	0.588
3	0.291	0.269	0.864	0.890
4	0.337	0.269	0.656	0.890
5	0.260	0.269	0.815	0.890
Rata-rata	0.238	0.269	0.9605	0.89

Berikut ini adalah hasil perbandingan nilai *Silhouette Score* untuk algoritma *K-Means* dan *Fuzzy C-Means* untuk jumlah kluster = 7.

**Tabel 6.** Perbandingan antar algoritma dengan kluster = 7

Eksperimen ke	Silhouette Score K-Means	Silhouette Score Fuzzy C-Means	Davies-Bouldin Index K-Means	Davies-Bouldin Index Fuzzy C-Means
1	0.214	0.244	1.008	0.767
2	0.199	0.202	1.050	1.010
3	0.289	0.244	0.812	0.767
4	0.286	0.202	0.753	1.010
5	0.276	0.202	0.890	1.010
Rata-rata	0.245	0.223	0.949	0.8885

Berikut ini merupakan rata-rata dari setiap kluster serta algoritma *K-Means* dan *Fuzzy C-Means*

**Table 7.** Perbandingan rata-rata antar algoritma

Jumlah Cluster	Rata-Rata Silhouette Score K-Means	Rata-Rata Silhouette Score Fuzzy C-Means	Rata-Rata Davies-Bouldin Index K-Means	Rata-Rata Davies-Bouldin Index Fuzzy C-Means
2	0.873	0.873	0.111	0.111
3	0.848	0.275	0.093	1.0035
4	0.361	0.36	0.5825	0.615
5	0.32	0.351	0.707	0.699
6	0.238	0.269	0.9605	0.89
7	0.245	0.223	0.949	0.8885

Melalui tabel di atas, rata-rata *Silhouette Score* terbesar terdapat pada algoritma *K-Means* dan *Fuzzy C-Means* sebesar 0.873 dengan kluster = 2. Rata-rata *Davies-Bouldin Index* terbesar terdapat pada algoritma *K-Means* dengan nilai sebesar 0.9605. Rata-rata *Silhouette Score* terendah terdapat pada algoritma *Fuzzy C-Means* dengan nilai sebesar 0.223. Rata-rata *Davies Bouldin Index* terendah terdapat pada algoritma *K-Means* dan *Fuzzy C-Means* dengan nilai sebesar 0.111. Secara keseluruhan, algoritma *K-Means* memiliki performa yang lebih baik dikarenakan memiliki *Silhouette Score* yang lebih besar. Algoritma *K-Means* juga lebih baik dibandingkan dengan algoritma *Fuzzy C-Means* dikarenakan nilai *Davies-Bouldin Index* yang lebih rendah. Jumlah kluster terbaik ada pada kluster = 2 karena memiliki nilai yang sama menggunakan algoritma *K-Means* ataupun menggunakan algoritma *Fuzzy C-Means*.

## 5. KESIMPULAN

Algoritma *K-Means* memiliki performa yang lebih baik dibandingkan dengan algoritma *Fuzzy C-Means*. Kedua algoritma dengan kluster = 2 memiliki hasil *Silhouette Score* dan *Davies-Bouldin Index* yang sama. Namun, apabila melihat nilai secara keseluruhan termasuk dari kluster 2-7, algoritma *K-Means* memiliki hasil yang lebih baik. Kemungkinan untuk pengembangan selanjutnya

adalah menggunakan algoritma lain selain *K-Means* dan *Fuzzy C-Means* untuk mencari *Silhouette Score* dan *Davies-Bouldin Index* yang lebih baik.

## UCAPAN TERIMA KASIH

Terima kasih kepada Bu Teny Handayhani S.Kom., M.Kom., Ph.D. yang telah membantu saya dalam menyusun tulisan ilmiah ini serta menjadi dosen pengajar saya untuk mata kuliah Machine Learning sehingga saya bisa menyusun tulisan ilmiah ini dengan baik dan benar.

## DAFTAR PUSTAKA

- [1] M. Triandini, S. Defit And G. W. Nurcahyo, "Data Mining Dalam Mengukur Tingkat Keaktifan Siswa Dalam Mengikuti Proses Belajar Pada Smp It Andalas Cendekia Dengan Menggunakan Metode K-Means Clustering," *Jurnal Informasi Dan Teknologi*, Vol. 3, No. 3, 2021.
- [2] K. P. Simanjuntak And U. Khaira, "Pengelompokan Titik Api Di Provinsi Jambi Dengan *Algoritma Agglomerative Hierarchical Clustering*," *Malcom: Indonesian Journal of Machine Learning and Computer Science*, Vol. 1, No. 1, 2021.
- [3] R. J. Kasim, S. Bahri And S. Amir, "Implementasi Metode K-Means Untuk *Clustering* Data Penduduk Miskin Dengan Systematic Random Sampling," *Prosiding Seminar Nasional Sistem Informasi Dan Teknologi (Sisfotek)*, 2021
- [4] M. R. Sulistio, N. Suarna And O. Nurdiawan, "Analisa Penerapan Metode Clustering X-Means Dalam Pengelompokan Penjualan Barang," *Jurnal Teknologi Ilmu Komputer*, Vol. 1, No. 2, 2023.
- [5] I. Soliani And S. Juanita, "Grouping the Prevalence of Disease Cases by Age In Bandung City Hospitals Using K-Means," *Jurnal Teknik Informatika (Jutif)*, Vol. 3, No. 6, 2022.
- [6] G. B. Kaligis And S. Yulianto, "Analisa Perbandingan Algoritma K-Means, Kmedoids, Dan X-Means Untuk Pengelompokan Kinerja Pegawai," Vol. 1, No. 3, 2022.
- [7] D. Marcelina, A. Kurnia And T., "Analisis Klaster Kinerja Usaha Kecil Dan Menengah Menggunakan *Algoritma K-Means Clustering*," *Malcom: Indonesian Journal of Machine Learning and Computer Science*, Vol. 3, No. 2, 2023.
- [8] Y. Elda, S. Defit, Y. Yunus And R. Syaljumairi, "Klasterisasi Penempatan Siswa Yang Optimal Untuk Meningkatkan Nilai Rata-Rata Kelas Menggunakan K-Means," *Jurnal Informasi Dan Teknologi*, Vol. 3, No. 3, 2021.
- [9] M. D. Permana, A. L. Hananto, E. Novalia, B. Huda And T. Paryono, "Klasterisasi Data Jamaah Umrah Pada Tanurmutmainah Tour Menggunakan Algoritma K-Means," *Jurnal Komtekinfo*, Vol. 10, No. 1, 2023.
- [10] M. D. Chandra, E. Irawan, I. S. Saragih, A. P. Windarto And D. Suhendro, "Penerapan Algoritma K-Means Dalam Mengelompokkan Balita Yang Mengalami Gizi Buruk Menurut Provinsi," *Bios: Jurnal Teknologi Informasi Dan Rekayasa Komputer*, Vol. 2, No. 1, 2021.
- [11] H. Syukron, M. F. Fayyad, F. J. Fauzan, Y. Ikhsani And U. R. Gurning, "Perbandingan K-Means K-Medoids Dan Fuzzy C-Means Untuk Pengelompokan Data Pelanggan Dengan Model Lrfm," *Malcom: Indonesian Journal of Machine Learning and Computer Science*, Vol. 2, No. 2, 2022.
- [12] R. Fauziah And A. I. Purnamasari, "Implementasi Algoritma K-Means Pada Kasus Kekerasan Anak Dan," 2023.
- [13] N. Ulinnuha, "Provincial *Clustering in Indonesia Based on Plantation Production Using Fuzzy C-Means*," *Itsmart: Jurnal Ilmiah Teknologi Dan Informasi*, Vol. 9, No. 1, 2020.
- [14] M. Siregar And M., "Analisis Perilaku Sosial Masyarakat Terhadap Pandemi Menggunakan Metode Fuzzy C-Means," *Journal of Information Systems and Technology (Jistech)*, Vol. 1, No. 2, 2024.
- [15] A. C. Putra And K. D. Hartono, "Optimalisasi Penyaluran Bantuan Pemerintah Untuk Umkm Menggunakan Metode Fuzzy C-Means," *Jurnal Resti*, Vol. 5, No. 3, 2021.
- [16] L. Wulandari And B. O. Yogantara, "*Algorithm Analysis Of K-Means and Fuzzy C-Means for Clustering Countries Based on Economy and Health*," *Faktor Exacta*, Vol. 15, No. 2, 2022.
- [17] S. Mulyadi, F. Insani, S. Agustian And L. Afriyanti, "Pengelompokan Data Pendistribusian Listrik Menggunakan Algoritma Mini Batch K-Means Clustering," *Malcom: Indonesian Journal of Machine Learning and Computer Science*, Vol. 4, No. 3, 2024.
- [18] M. Farid, "Pengelompokan Data Pendistribusian Listrik Menggunakan Algoritma *Density Based Spatial Clustering of Application with Noise (DBSCAN)*," 2024.

- [19] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," *Jurnal Ilmiah Informatika Dan Ilmu Komputer (Jima-Ilkom)*, Vol. 3, No. 1, 2024.
- [20] W. Putri And M. Afdal, "Penerapan Algoritma K-Means Untuk Pengelompokan Data Penyandang Disabilitas Di Kabupaten Rokan Hilir," *Ijirse: Indonesian Journal of Informatic Research and Software Engineering*, Vol. 3, No. 1, 2023.
- [21] M. Ulfah, A. S. Irtawaty, S. Mulyanto, Y. Kurniawan And Z., "Pengelompokan Daerah Berdasarkan Faktor Dampak Bencana Tanah Longsor Menggunakan Algoritma K-Medoids," Vol. 9, No. 1, 2023.
- [22] S. Ramadhani, D. Azzahra And T. Z., "Comparison Of K-Means And K-Medoids Algorithms in Text Mining Based on Davies Bouldin Index Testing for Classification of Student's Thesis," *Jurnal Teknologi Informasi & Komunikasi Digital Zone*, Vol. 13, No. 1, 2022.
- [23] B. Kristanto, A. T. Zy And M. Fatchan, "Analisis Penentuan Karyawan Tetap Dengan *Algoritma K-Means Dan Davies Bouldin Index*," *Bulletin of Information Technology (Bit)*, Vol. 4, No. 1, 2023.
- [24] D. Purliantoro And I. Ayesha, "Klasterisasi Data Mining K-Means Dengan *Indeks Davies Bouldin* Berdasarkan Hasil Peramalan Produksi Tanaman Biofarmaka Di Provinsi Indonesia Menggunakan Arima," *Journal of Scientech Research and Development*, Vol. 5, No. 1, 2023.
- [25] A. M. Nusrang, M. K. Aidid And Z. Rais, "*K-Means Cluster Analysis for Grouping Districts in South Sulawesi Province Based on Village Potential*," *Arrus Journal of Mathematics and Applied Science*, Vol. 2, No. 2.