

# CLUSTERING KETIDAKCUKUPAN KONSUMSI PANGAN PER KABUPATEN/KOTA MENGGUNAKAN ALGORITMA K-MEANS DAN FUZZY C-MEANS

**Jonathan Suryadi**

Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara,  
Jln. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia  
E-mail: jonathan.535210072@stu.untar.ac.id

## ABSTRAK

Ketidakcukupan konsumsi pangan merupakan masalah yang signifikan di Indonesia, dipengaruhi oleh faktor sosial-ekonomi dan kondisi geografis. Penelitian ini bertujuan untuk mengidentifikasi pola distribusi ketidakcukupan konsumsi pangan per kabupaten/kota dengan menggunakan algoritma K-Means dan Fuzzy C-Means. Data yang digunakan berasal dari Badan Pusat Statistik, mencakup prevalensi ketidakcukupan konsumsi pangan di seluruh Indonesia. Hasil penelitian menunjukkan bahwa algoritma K-Means menghasilkan pengelompokan dengan Silhouette Score yang lebih tinggi untuk jumlah cluster kecil, sementara Fuzzy C-Means memberikan stabilitas yang lebih baik pada cluster yang lebih kompleks. Temuan ini memberikan wawasan penting untuk pengambilan kebijakan distribusi pangan berbasis data yang lebih efektif.

**Kata kunci**— Ketidakcukupan Konsumsi Pangan, K-Means, Fuzzy C-Means, Clustering, Silhouette Score.

## ABSTRACT

*Food consumption inadequacy is a significant problem in Indonesia, influenced by socio-economic factors and geographical conditions. This study aims to identify the distribution pattern of food consumption inadequacy per district/city using K-Means and Fuzzy C-Means algorithms. The data used comes from the Central Bureau of Statistics, covering the prevalence of inadequate food consumption throughout Indonesia. The results show that K-Means algorithm produces clustering with higher Silhouette Score for small number of clusters, while Fuzzy C-Means provides better stability on more complex clusters. These findings provide important insights for more effective data-driven food distribution policy making.*

**Keywords**— Food Consumption Inadequacy, K-Means, Fuzzy C-Means, Clustering, Silhouette Score.

## 1. PENDAHULUAN

Ketidakcukupan konsumsi pangan merupakan tantangan besar yang berdampak pada kesehatan masyarakat dan pembangunan ekonomi, terutama di daerah-daerah dengan keterbatasan akses pangan. Faktor-faktor sosial-ekonomi dan kondisi geografis masing-masing kabupaten/kota di Indonesia dapat memengaruhi tingkat konsumsi pangan. Oleh karena itu, penting untuk memahami pola distribusi dan konsumsi pangan di berbagai wilayah dengan pendekatan berbasis data yang dapat memberikan gambaran yang lebih jelas tentang ketidakmerataan tersebut. Salah satu metode yang dapat diterapkan adalah algoritma clustering seperti K-Means dan Fuzzy C-Means, yang memungkinkan untuk mengelompokkan kabupaten/kota berdasarkan kemiripan pola konsumsi pangan mereka, meskipun dengan batasan yang berbeda. [1] [2]

Dalam hal ini, K-Means mengelompokkan data berdasarkan kedekatan titik pusat cluster, sementara Fuzzy C-Means memungkinkan data memiliki kemungkinan ganda untuk masuk dalam lebih dari satu cluster. Hal ini sangat berguna terutama ketika data memiliki ketidakpastian dan tidak ada batas yang jelas di antara cluster, yang sering kali terjadi dalam analisis konsumsi pangan di wilayah dengan keragaman sosial-ekonomi yang tinggi. Beberapa studi sebelumnya telah berhasil menunjukkan keunggulan algoritma ini dalam berbagai bidang, termasuk analisis ketidakcukupan pangan, di mana algoritma tersebut membantu memahami pola distribusi dan ketidakmerataan konsumsi. [3] [4]

Dengan menggunakan algoritma K-Means dan Fuzzy C-Means, penelitian ini bertujuan untuk memberikan wawasan baru tentang bagaimana pola konsumsi pangan dapat dikelompokkan berdasarkan data yang ada, serta memberikan dasar bagi kebijakan distribusi pangan yang lebih terarah dan berbasis data di Indonesia. [5]

## 2. METODE PENELITIAN

### 2.1. Dataset

Data yang digunakan dalam penelitian ini berasal dari Badan Pusat Statistik (BPS), khususnya data Prevalensi Ketidacukupan Konsumsi Pangan per Kabupaten/Kota. Dataset ini mencakup informasi tentang persentase kabupaten/kota di Indonesia yang mengalami ketidacukupan konsumsi pangan, dengan berbagai variabel terkait yang dapat digunakan untuk analisis lebih lanjut.

### 2.2. Algoritma Klasifikasi

Berbagai algoritma clustering digunakan dalam penelitian ini untuk mengelompokkan kabupaten/kota berdasarkan pola konsumsi pangan mereka. Clustering adalah teknik analisis data yang digunakan untuk menemukan pola atau struktur dalam data tanpa menggunakan label atau kategori yang sudah ditentukan sebelumnya. Dua algoritma clustering utama yang digunakan dalam penelitian ini adalah K-Means dan Fuzzy C-Means, yang keduanya memungkinkan pengelompokan data berdasarkan kedekatannya dengan titik pusat (*centroid*) atau menggunakan logika fuzzy untuk menangani ketidakpastian dalam data. [6]

#### 2.2.1. K-Means

K-Means adalah algoritma *clustering* yang mengelompokkan data ke dalam  $k$  cluster berdasarkan kedekatannya dengan titik pusat (*centroid*). Berikut adalah langkah-langkah utama dalam algoritma K-Means:

1. Inisialisasi *Centroid*: Tentukan jumlah cluster ( $k$ ) yang diinginkan dan pilih centroid awal. Pemilihan *centroid* dapat dilakukan secara acak atau menggunakan metode K-Means++, yang lebih efisien dalam memilih titik pusat yang optimal.
2. Penugasan Data ke Cluster: Setiap titik data akan dialokasikan ke cluster yang memiliki centroid terdekat. Untuk mengukur kedekatannya, biasanya digunakan jarak Euclidean antara data dan centroid, yang dihitung dengan rumus berikut:

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

di mana:

- $d(x,c)$  : Jarak Euclidean antara data  $x$  dan  $c$ .
- $x_i$  dan  $c_i$  : Nilai fitur ke- $i$  dari titik data  $x$  dan  $c$ .
- $n$  : Jumlah fitur pada data.

3. Update Centroid: Setelah data dialokasikan ke cluster, centroid dihitung ulang berdasarkan rata-rata posisi data dalam cluster tersebut. Formula pembaruan centroid untuk cluster  $k$  adalah:

$$c_k = \frac{1}{|C_k|} \sum_{x \in C_k} x$$

di mana:

- $c_k$  : Centroid dari cluster  $k$ ,
- $C_k$  : Set data yang tergolong dari cluster  $k$ ,
- $|C_k|$  : Jumlah dari elemen cluster  $k$ ,
- $X$  : Titik data dalam cluster  $k$ .

4. Iterasi: Proses ini berlanjut dengan langkah 2 dan 3 sampai posisi centroid tidak berubah secara signifikan, yang menandakan bahwa algoritma telah konvergen.

K-Means banyak digunakan dalam berbagai aplikasi karena kesederhanaannya dan kemampuannya dalam memproses data besar. Namun, kelemahan dari algoritma ini adalah ketergantungan pada pemilihan jumlah cluster yang tepat dan sensitivitas terhadap pemilihan centroid awal yang acak. Beberapa penelitian baru-baru ini mengkaji varian K-Means dan evaluasi performanya, terutama dalam menangani data besar dan beragam. [7] [8]

### 2.2.2. Algoritma Fuzzy C-Means

*Fuzzy C-Means* (FCM) adalah algoritma clustering berbasis fuzzy logic, yang memungkinkan data memiliki keanggotaan dalam lebih dari satu cluster. Hal ini sangat berguna ketika data tidak memiliki batas yang jelas antar cluster, seperti dalam analisis data konsumsi pangan yang memiliki ketidakpastian. Algoritma ini memberikan derajat keanggotaan untuk setiap data pada setiap cluster, dengan menggunakan fungsi keanggotaan fuzzy untuk mengukur seberapa besar data tersebut terasosiasi dengan suatu cluster.

Langkah-langkah dalam Fuzzy C-Means adalah sebagai berikut:

1. Inisialisasi: Tentukan jumlah cluster ( $c$ ) dan parameter fuzzy ( $m$ ), yang mengontrol derajat keanggotaan.
2. Penugasan Keanggotaan: Setiap titik data diberikan derajat keanggotaan pada setiap cluster berdasarkan kedekatannya dengan centroid dan posisi centroid lainnya. Fungsi keanggotaan dihitung dengan rumus berikut:

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left( \frac{d(x_i, c_j)}{d(x_i, c_k)} \right)^{\frac{2}{m-1}}}$$

di mana:

- $u_{ij}$  : Keanggotaan data  $x_i$  pada cluster  $c_j$ ,
- $d(x_i, c_j)$  : Jarak Euclidean antara data  $x_i$  dan centroid  $c_j$ ,
- $m$  : Parameter fuzzy, yang lebih besar dari 1, biasanya 2,
- $C$  : Jumlah cluster.

3. Update Centroid: Centroid dihitung ulang dengan memperhatikan derajat keanggotaan fuzzy setiap titik data. Rumus pembaruan centroid adalah:

$$c_j = \frac{\sum_{i=1}^n u_{ij}^m \cdot x_i}{\sum_{i=1}^n u_{ij}^m}$$

di mana:

- $c_j$  : Centroid dari cluster  $j$ ,
- $u_{ij}$  : Keanggotaan data  $x_i$  pada cluster  $c_j$ ,
- $m$  : Paramete fuzzy.

4. Iterasi: Proses ini diulang hingga konvergensi tercapai, yaitu ketika perubahan keanggotaan dan posisi centroid sangat kecil.

Keunggulan utama dari Fuzzy C-Means adalah kemampuannya untuk menangani ketidakpastian dan fleksibilitas dalam mengelompokkan data yang tidak terpisah dengan jelas. [9] [10]

### 2.3. Metode Evaluasi

Evaluasi model clustering dalam penelitian ini menggunakan *Silhouette Score*, yang merupakan metrik utama untuk mengukur kualitas hasil *clustering*. *Silhouette Score* mengukur

seberapa baik setiap titik data terkluster dalam cluster yang sesuai dibandingkan dengan cluster lainnya. Nilai Silhouette Score berkisar antara -1 hingga +1, dengan nilai yang mendekati +1 menunjukkan bahwa data dikelompokkan dengan baik dan nilai yang mendekati -1 menunjukkan bahwa data lebih cocok berada dalam cluster lain. Metrik ini sangat berguna karena tidak memerlukan data label yang sudah ada dan dapat memberikan gambaran yang jelas mengenai kualitas cluster yang terbentuk, baik untuk K-Means maupun Fuzzy C-Means. Dalam penelitian ini, Silhouette Score digunakan untuk mengevaluasi performa kedua algoritma clustering dengan berbagai jumlah cluster (k) dan diulang dengan beberapa eksperimen untuk memastikan hasil yang konsisten. [11]

### 3. HASIL DAN PEMBAHASAN

Pada penelitian ini, dilakukan eksperimen clustering menggunakan dua algoritma, yaitu K-Means dan Fuzzy C-Means, untuk menganalisis pola konsumsi pangan per kabupaten/kota di Indonesia. Masing-masing algoritma diuji dengan 10 eksperimen untuk berbagai jumlah cluster (k) mulai dari k=2 hingga k=7. Hasil eksperimen menunjukkan perbandingan antara kedua algoritma dalam hal kualitas clustering yang diukur dengan *Silhouette Score*.

#### 3.1. Hasil Eksperimen

Tabel berikut menunjukkan hasil *Silhouette Score* untuk K-Means dan Fuzzy C-Means pada setiap jumlah cluster (k) dan eksperimen yang dilakukan. Rata-rata *Silhouette Score* untuk masing-masing algoritma dihitung setelah 10 eksperimen.

**Tabel 1.** Tabel Hasil Eksperimen

Eksperimen ke	K-Means (k=2)	Fuzzy C-Means (k=2)	K-Means (k=3)	Fuzzy C-Means (k=3)	K-Means (k=4)	Fuzzy C-Means (k=4)	K-Means (k=5)	Fuzzy C-Means (k=5)	K-Means (k=6)	Fuzzy C-Means (k=6)	K-Means (k=7)	Fuzzy C-Means (k=7)
1	0.7857	0.7848	0.5494	0.5475	0.5774	0.5778	0.5799	0.5589	0.5113	0.5243	0.5098	0.5372
2	0.7857	0.7848	0.5494	0.5475	0.5790	0.5778	0.5622	0.5589	0.5500	0.5243	0.5471	0.5372
3	0.7857	0.7848	0.5494	0.5475	0.5740	0.5778	0.5802	0.5589	0.5281	0.5243	0.5282	0.5372
4	0.7848	0.7848	0.5494	0.5475	0.5790	0.5778	0.5795	0.5589	0.5125	0.5243	0.5439	0.5372
5	0.7857	0.7848	0.7366	0.5475	0.5790	0.5778	0.5591	0.5589	0.5142	0.5243	0.5156	0.5372
6	0.7848	0.7848	0.7366	0.5475	0.5740	0.5778	0.5795	0.5589	0.5384	0.5243	0.5272	0.5372
7	0.7848	0.7848	0.7286	0.5475	0.5774	0.5778	0.5802	0.5589	0.5113	0.5243	0.5111	0.5372
8	0.7848	0.7848	0.7286	0.5475	0.5775	0.5778	0.5802	0.5589	0.5469	0.5243	0.5275	0.5372
9	0.7857	0.7848	0.7366	0.5475	0.5774	0.5778	0.5648	0.5589	0.5554	0.5243	0.5289	0.5372
10	0.7848	0.7848	0.5491	0.5475	0.5790	0.5778	0.5802	0.5589	0.5405	0.5243	0.5282	0.5370

Rata-Rata *Silhouette Score*

K-Means : 0.6060

Fuzzy C-Means : 0.5884

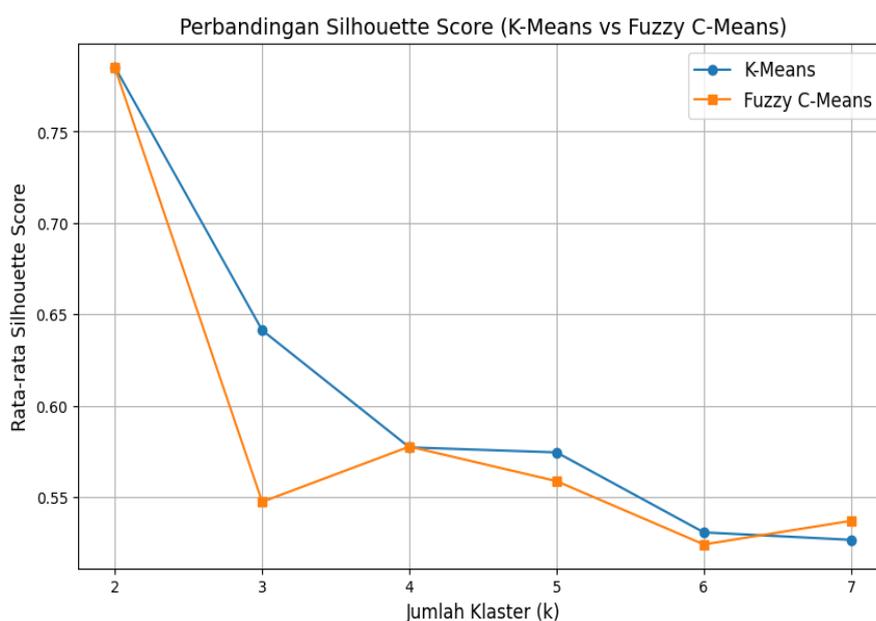
#### 3.2. Perbandingan *Silhouette Score* (K-Means vs Fuzzy C-Means)

Perbandingan *Silhouette Score* antara K-Means dan Fuzzy C-Means dilakukan untuk melihat kualitas pengelompokan pada berbagai jumlah cluster (k). Berdasarkan hasil eksperimen, dapat dilihat bahwa K-Means memberikan *Silhouette Score* yang lebih tinggi daripada Fuzzy C-Means pada sebagian besar jumlah cluster yang lebih kecil (k=2 dan k=3). Hal ini menunjukkan bahwa K-

Means lebih efektif dalam menghasilkan cluster yang lebih terpisah dan lebih padat pada jumlah cluster yang lebih sedikit.

Namun, ketika jumlah cluster meningkat ( $k=6$  dan  $k=7$ ), penurunan nilai Silhouette Score lebih tajam pada K-Means, yang menunjukkan bahwa algoritma ini kurang efektif dalam menangani data yang lebih kompleks atau cluster yang lebih banyak. Sebaliknya, Fuzzy C-Means menunjukkan stabilitas yang lebih baik di seluruh eksperimen, meskipun dengan Silhouette Score yang sedikit lebih rendah dibandingkan K-Means.

Gambar berikut menunjukkan perbandingan *Silhouette Score* rata-rata untuk kedua algoritma pada berbagai jumlah cluster. Dari grafik tersebut, terlihat bahwa K-Means unggul pada jumlah cluster yang lebih kecil, tetapi Fuzzy C-Means cenderung lebih konsisten pada jumlah cluster yang lebih besar. Perbandingan *Silhouette Score* dapat dilihat pada Gambar 1.



**Gambar 1.** Gambar Perbandingan Silhouette Score

### 3.3. Pembahasan Hasil

Berdasarkan hasil yang diperoleh dari K-Means dan Fuzzy C-Means, dapat dilihat bahwa K-Means cenderung memberikan Silhouette Score yang lebih tinggi dibandingkan Fuzzy C-Means, terutama pada jumlah cluster yang lebih kecil ( $k=2$ ). Hal ini menunjukkan bahwa K-Means lebih efektif dalam mengelompokkan data ketika jumlah cluster yang lebih sedikit digunakan, menghasilkan cluster yang lebih terpisah dan lebih padat.

Namun, seiring bertambahnya jumlah cluster ( $k=6$  dan  $k=7$ ), skor untuk K-Means mengalami penurunan yang lebih signifikan, menunjukkan bahwa K-Means lebih rentan terhadap data yang lebih kompleks atau jumlah cluster yang lebih tinggi. Di sisi lain, Fuzzy C-Means menunjukkan kestabilan yang lebih baik meskipun dengan skor yang lebih rendah, yang menunjukkan keunggulan algoritma ini dalam menangani ketidakpastian dan data yang tumpang tindih.

Secara keseluruhan, meskipun K-Means unggul dalam hal pemisahan cluster pada jumlah cluster yang lebih kecil, Fuzzy C-Means memberikan hasil yang lebih stabil dan fleksibel pada jumlah cluster yang lebih besar. Kedua algoritma ini memiliki kekuatan dan kelemahannya masing-masing, yang harus dipertimbangkan berdasarkan karakteristik data yang digunakan.

#### 4. KESIMPULAN

Dalam penelitian ini, dilakukan eksperimen clustering menggunakan K-Means dan Fuzzy C-Means untuk menganalisis data ketidacukupan konsumsi pangan di Indonesia. Berdasarkan hasil eksperimen dengan *Silhouette Score*, K-Means menunjukkan skor yang lebih tinggi dibandingkan Fuzzy C-Means pada jumlah cluster yang lebih kecil ( $k=2$  dan  $k=3$ ), namun mengalami penurunan pada jumlah cluster yang lebih banyak. Di sisi lain Fuzzy C-Means menunjukkan hasil yang lebih stabil meskipun dengan skor yang sedikit lebih rendah.

Secara keseluruhan, K-Means lebih efektif pada data dengan jumlah cluster sedikit, sementara Fuzzy C-Means lebih fleksibel dalam menangani data yang lebih kompleks dan tumpang tindih. Pemilihan algoritma yang tepat bergantung pada karakteristik data dan tujuan analisis.

#### UCAPAN TERIMA KASIH

Penulis ingin menyampaikan penghargaan yang tulus kepada Fakultas Teknologi Informasi Universitas Tarumanagara, serta kepada dosen-dosen yang telah memberikan bimbingan. Ucapan terima kasih juga ditujukan kepada keluarga dan teman-teman yang telah memberikan dukungan moral dan motivasi, sehingga penulis dapat menyelesaikan penelitian ini dengan baik.

#### DAFTAR PUSTAKA

- [1]. A. Kumar and S. S. Sodhi, "Comparative Analysis of Fuzzy C- Means and K-Means Clustering in the Case of Image Segmentation," *2021 8th International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 194-200, 2021
- [2]. I. Tahyudin, G. Firmansyah, A. G. Ivansyah, W. Ma'arifah and L. Lestari, "Comparison of K-Means Algorithms and Fuzzy C-Means Algorithms for Clustering Customers Dataset," *2022 1st International Conference on Smart Technology, Applied Informatics, and Engineering (APICS)*, pp. 211-216, 2022.
- [3]. S. a. H. D. Li, "A Comparative Study on k-Means Clustering with Different Cluster Representations," *2023 42nd Chinese Control Conference (CCC)*, pp. 7959-7964, 2024.
- [4]. J. L. a. M. F. a. R. A. a. B. A. a. D. P. Corcuera Bárcena, "Federated c-means and Fuzzy c-means Clustering Algorithms for Horizontally and Vertically Partitioned Data," *IEEE Transactions on Artificial Intelligence*, pp. 1-15, 2024.
- [5]. I. a. C. M. a. C. M. E. a. S. A. E. a. d. I. T. D. I. a. A. S. M. a. A. I. Mahnoor and Shafi, "A Review of Approaches for Rapid Data Clustering: Challenges, Opportunities, and Future Directions," *IEEE Access*, vol. 12, pp. 138086-138120, 2024.
- [6]. A. a. K. L. S. Pugazhenthii, "Selection of Optimal Number of Clusters and Centroids for K-means and Fuzzy C-means Clustering: A Review," *2020 5th International Conference on Computing, Communication and Security (ICCCS)*, pp. 1-4, 2020.
- [7]. R. S. a. M. S. I. Mohiuddin Ahmed, "The k-means Algorithm: A Comprehensive Survey and Performance Evaluation," *Electronics*, 2020, vol. 9, 2020.
- [8]. K. P. a. Y. M.-S. Sinaga, "Unsupervised K-Means Clustering Algorithm," *IEEE Access*, vol. 8, pp. 80716-80727, 2020.
- [9]. W. M. a. K. Xu, "Enhancement of the Classification Performance of Fuzzy C-Means through Uncertainty Reduction with Cloud Model Interpolation," *Mathematics*, 2024, vol. 12, 2024.
- [10]. S. a. L. D. a. Z. Z. a. P. R. Zhou, "A New Membership Scaling Fuzzy C-Means Clustering Algorithm," *IEEE Transactions on Fuzzy Systems*, vol. 29, pp. 2810-2818, 2021.
- [11]. A. P. a. Sengupta, "Benchmarking Studies Aimed at Clustering and Classification Tasks Using K-Means, Fuzzy C-Means and Evolutionary Neural Networks," *Mach. Learn. Knowl. Extr.*, 2021, vol. 3, 2021.