

# Penentuan Parameter Tingkat Ke-Fuzzy-an Fuzzy C-Means dan Pengaruhnya Terhadap Proses Algoritma

Ni Putu Ayu Triana<sup>a1</sup> dan L. G. Astuti<sup>a2</sup>

<sup>a</sup>Program Studi Informatika, Universitas Udayana  
Jimbaran, Badung, Bali, Indonesia

<sup>1</sup>pulaudewata17@yahoo.co.id

<sup>2</sup>lg.astuti@unud.ac.id (Corresponding author)

## Abstract

*Fuzzy C-Means (FCM) is an algorithm in the process of data clustering that has limitations in the form of being sensitive to the parameters used so that in some cases, the final solution provided is not an optimal solution. One of the influential parameters is the fuzziness level of the algorithm. This parameter is a random real number greater than 1. The determination of these parameters is adjusted to the data used and evaluated with the condition that it reaches a minimum number of iterations for convergence, a small objective final value, and a DBI cluster validity value close to 0. In this study, Indonesian automotive sales data received the optimal algorithm fuzzy level parameter at a value of 2 with other fixed parameters, such as the number of clusters is 3, the smallest error expected to be is 0.00001, and the maximum iteration is 100.*

**Keywords:** Fuzzy C-Means, Fuzziness Level, DBI, clustering

## 1. Pendahuluan

*Fuzzy C-Means (FCM)* merupakan algoritma *machine learning* yang digunakan dalam proses klasterisasi data, dimana hasil klasterisasi tersebut ditentukan berdasarkan derajat keanggotaan yang dimiliki oleh setiap data [1]. Algoritma ini banyak diimplementasikan untuk memberikan solusi dalam berbagai kasus klasterisasi karena performa dari algoritma yang lebih baik dibandingkan algoritma lainnya dengan menghasilkan validitas pusat klaster yang konvergen. Namun dibalik kelebihanannya, algoritma ini memiliki keterbatasan berupa sensitif terhadap parameter yang digunakan sehingga dalam beberapa kasus, solusi akhir yang diberikan bukan merupakan solusi yang optimal (terjebak dalam *local optimum*). Dalam implementasinya, terdapat 4 parameter yang mempengaruhi solusi akhir perhitungan algoritma tersebut diantaranya jumlah klaster, jumlah iterasi, *error* terkecil yang diharapkan, dan salah satunya tingkat ke-fuzzy-an algoritma [2].

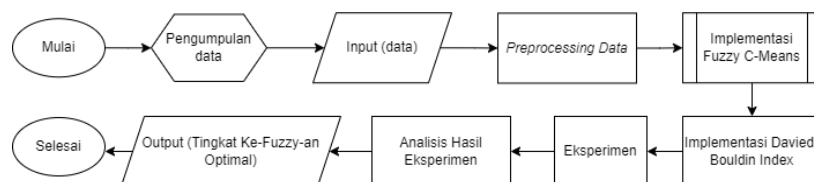
Parameter tingkat ke-fuzzy-an digunakan sebagai pangkat pembobot pada proses algoritma dengan kriteria berupa bilangan *real* acak lebih dari 1. Jika nilai parameter terlalu mendekati 1, maka data yang dikelompokkan akan mengarah ke pengelompokan keras seperti klasifikasi. Namun, apabila nilai parameter terlalu jauh dari 1, maka data yang dikelompokkan akan mengarah ke kaburan klaster [3]. Dalam implementasinya, parameter tingkat ke-fuzzy-an dipilih secara acak dan nilai yang paling banyak digunakan adalah 2. Hal ini karena nilai tersebut menghasilkan hasil optimal solusi akhir algoritma Fuzzy C-Means baik dari waktu algoritma, nilai objektif akhir yang didapatkan, dan validitas hasil klasterisasi. Namun, nilai ini tidak berlaku untuk setiap kasus klasterisasi dengan Fuzzy C-Means, sehingga diperlukan pemilihan parameter tingkat ke-fuzzy-an yang baik. Pemilihan parameter ini dilakukan secara manual dan tidak ada metode yang baku untuk menentukan parameter yang cocok digunakan karena parameter yang digunakan melekat pada data yang akan diklasterisasi [3].

Dalam penelitian sebelumnya, pemilihan parameter tingkat ke-fuzzy-an Fuzzy C-Means ditentukan hanya dengan melihat hasil akurasi dari klasterisasi yang dilakukan [4]. Sementara dalam penelitian ini, penentuan parameter tingkat ke-fuzzy-an Fuzzy C-Means ditentukan berdasarkan jumlah iterasi yang dibutuhkan, nilai objektif akhir, dan nilai validitas klasterisasi. Tujuan penelitian ini adalah mengetahui parameter tingkat ke-fuzzy-an Fuzzy C-Means yang optimal digunakan dan pengaruhnya terhadap proses algoritma.

## 2. Metode Penelitian

### 2.1. Desain Penelitian

Pada penelitian ini, akan dilakukan penentuan parameter tingkat ke-fuzzy-an yang optimal digunakan dalam data penjualan automotif Indonesia pada bulan Januari – Agustus 2022. Penentuan parameter ini dilakukan menggunakan 8 skenario eksperimen dengan tahapan yang dapat dilihat pada gambar 1.



Gambar 1. Desain Penelitian

### 2.2. Pengumpulan Data

Dalam penelitian ini, data yang digunakan merupakan data penjualan automotif Indonesia pada bulan Januari – Agustus 2022. Data tersebut dikumpulkan melalui *website* Asosiasi Industri Automotif Indonesia (<https://www.gaikindo.or.id/>). Jumlah data yang didapatkan adalah 104 data dengan 9 fitur yakni nama *brand* dari automotif yang terjual dan jumlah penjualan dari bulan Januari - Agustus 2022. Tipe data dari sembilan fitur tersebut dijabarkan dalam tabel 1.

Tabel 1. Daftar Fitur Data dan Tipe Data

Fitur	Tipe Data
nama_brand	Nominal
januari	Numerik
februari	Numerik
maret	Numerik
april	Numerik
mei	Numerik
juni	Numerik
juli	Numerik
agustus	Numerik

### 2.3. Pengolahan Data Awal

Data penjualan automotif Indonesia pada bulan Januari – Agustus 2022 yang digunakan dalam penelitian ini diberikan pengolahan data tahap awal (*pre-processing*) berupa penghapusan baris kosong dalam data dan menghilangkan nilai duplikasi dalam data berdasarkan fitur nama *brand* automotif. Dalam data yang digunakan pada penelitian ini, tidak terdapat nilai kosong maupun duplikasi data setelah dilakukan *pre-processing* menggunakan bahasa pemrograman Python (gambar 2 dan gambar 3).

```

    Menghapus baris yang memiliki nilai kosong (NaN)
    df = df.dropna()
    df
    nama_brand  januari  februari  maret  april  mei  juni  juli  agustus
    0  Honda All New City    57    101    136    58    34    16    20    73
    1  Honda All New Civic    80    158    64    57    30    0    87    4
    2  Hyundai HMID Ioniq EV Prime    0    0    0    0    0    0    0    4
    3  Hyundai HMID Ioniq EV Signature    17    1    6    2    1    2    0    12
    4  Hyundai HMID Genesis G80 EV    0    0    1    0    0    9    0    0
    ...
    99  Mitsubishi Fuso FN 61 FL HD (GX2)    1    10    19    6    14    12    26    66
    100  Hino ZY - HR    27    33    20    33    16    17    8    92
    101  Isuzu GVR 34 J HR ABS    26    22    53    29    42    32    2    86
    102  Mercedes Benz PC Axor 2528 C    37    74    96    86    40    92    38    5
    103  Toyota New Agya 1.2 G Mi 2020    39    36    64    14    33    39    29    12
    104 rows x 9 columns
    
```

Gambar 2. Hasil Data Proses Penghapusan Baris Kosong

menghapus nilai duplikasi

```
[5] df.drop_duplicates(subset="nama_brand",
keep=False, inplace=True)
df
```

	nama_brand	januari	februari	maret	april	mei	juni	juli	agustus
0	Honda All New City	57	101	136	58	34	16	20	73
1	Honda All New Civic	80	158	64	57	30	0	87	4
2	Hyundai HMID Ioniq EV Prime	0	0	0	0	0	0	0	4
3	Hyundai HMID Ioniq EV Signature	17	1	6	2	1	2	0	12
4	Hyundai HMID Genesis G80 EV	0	0	1	0	0	9	0	0
...	...	...	...	...	...	...	...	...	...
99	Mitsubishi Fuso FN 61 FL HD (6X2)	1	10	19	6	14	12	26	66
100	Hino ZY - HR	27	33	20	33	16	17	8	32
101	Isuzu GVR 34 J HR ABS	26	22	53	29	42	32	2	86
102	Mercedes Benz PC Axor 2528 C	37	74	96	86	40	92	38	5
103	Toyota New Ayga 1.2 G Mi 2020	39	36	64	14	33	39	29	12

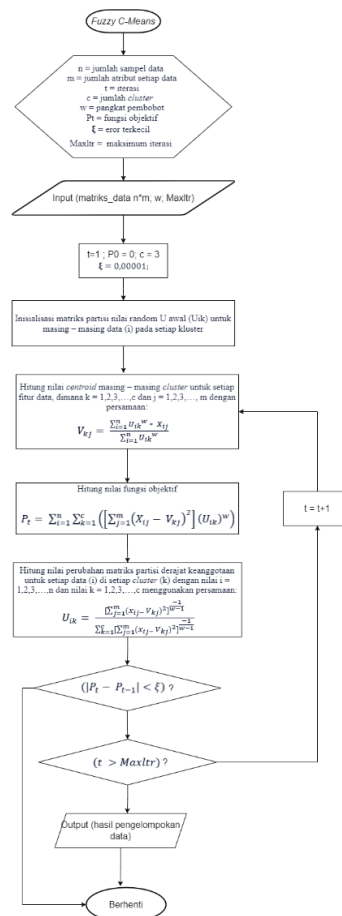
104 rows x 9 columns

Gambar 3. Hasil Data Proses Penghapusan Duplikasi Data

#### 2.4. Analisis Rentang Parameter Tingkat Ke-Fuzzy-an

Rentang parameter tingkat ke-fuzzy-an yang digunakan dalam penelitian ini adalah dari rentang 1.5 hingga 5 dengan jarak lompatan 0.5. Hal ini karena nilai tersebut banyak digunakan dalam berbagai kasus dan tidak memiliki tingkat kekaburan (*fuzziness*) dan tingkat keras pengelompokan yang kuat [5].

#### 2.5. Tahapan Algoritma Fuzzy C-Means



Gambar 4. Tahapan Algoritma Fuzzy C-Means

## 2.6. Evaluasi

*Davied-Bouldin Index* (DBI) adalah metode evaluasi untuk mengukur validitas kluster hasil metode klusterisasi dengan menguji tingkat maksimal jarak inter-kluster dan pada saat yang sama juga menguji tingkat minimal jarak antar kluster. Semakin kecil nilai DBI yang diperoleh (non-negatif  $\geq 0$ ), maka semakin baik kluster yang diperoleh dari pengelompokan yang dilakukan oleh algoritma klusterisasi [5]. Persamaan dari DBI dapat dilihat pada persamaan (1).

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (1)$$

Keterangan:

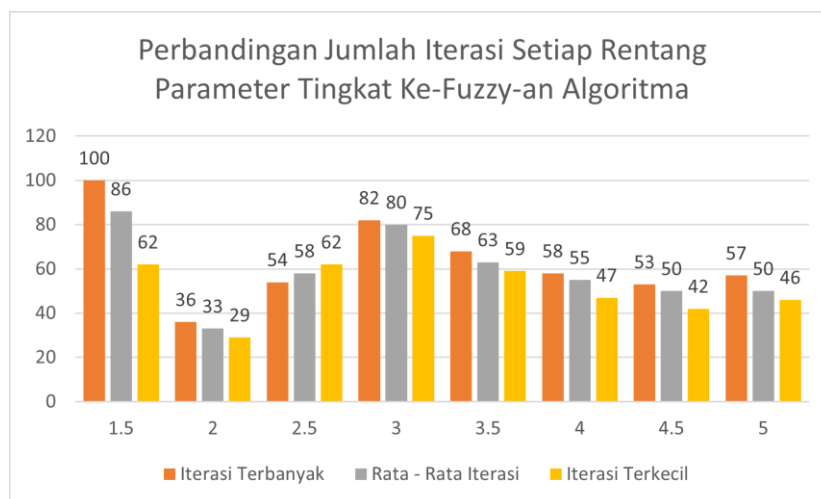
K : jumlah kluster

## 3. Hasil dan Pembahasan

Dalam penelitian ini, terdapat 8 skenario eksperimen yang dilakukan yaitu melakukan klusterisasi data menggunakan Fuzzy C-Means dengan nilai parameter tingkat ke-fuzzy-an algoritma diantaranya 1.5, 2, 2.5, 3, 3.5, 4, 4.5, dan 5. Setiap skenario dilakukan sebanyak 10 kali dan dilakukan pencatatan jumlah iterasi dan nilai objektif akhir yang didapat untuk mencapai konvergen. Serta, dilakukan pula proses evaluasi validitas menggunakan DBI pada masing-masing skenario. Hasil klusterisasi yang optimal adalah skenario yang memiliki hasil kombinasi jumlah iterasi yang sedikit, nilai akhir fungsi objektif yang kecil, dan nilai validitas menggunakan DBI yang kecil. Sementara itu, implementasi algoritma Fuzzy C-Means dalam penelitian ini dilakukan menggunakan bahasa pemrograman python dengan tahapan seperti pada gambar 4. Adapun parameter tetap yang digunakan dalam penelitian ini diantaranya jumlah kluster sebanyak 3, jumlah maksimum iterasi sebanyak 100, dan nilai *error* yang diharapkan adalah 0.00001.

### 3.1. Pengaruh Pada Jumlah Iterasi

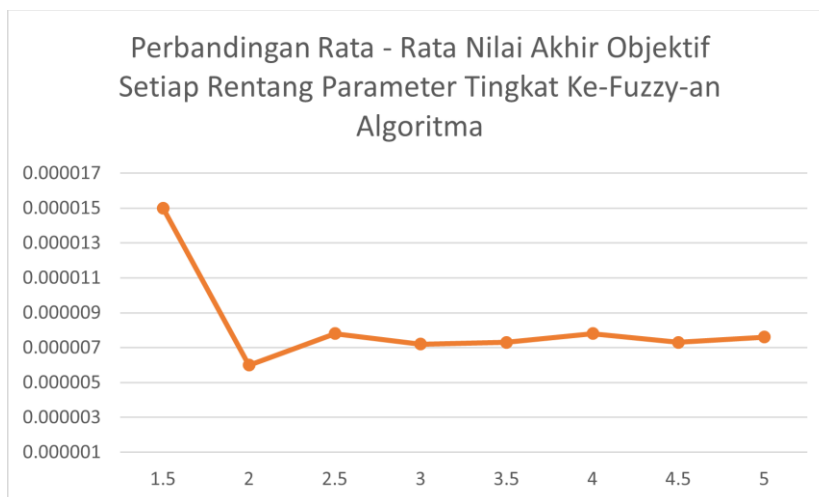
Pada gambar 5, ditampilkan perbandingan jumlah iterasi setiap rentang parameter tingkat ke-fuzzy-an Fuzzy C-Means. Secara keseluruhan, nilai parameter 2 memiliki jumlah iterasi minimum dalam 3 kategori iterasi untuk mencapai konvergen dengan iterasi terkecil sebanyak 29, iterasi terbesar sebanyak 36, dan rata-rata iterasi sebanyak 33.



**Gambar 5.** Perbandingan Jumlah Iterasi Setiap Rentang Parameter Tingkat Ke-Fuzzy-an Algoritma

### 3.2. Pengaruh Pada Nilai Objektif Akhir

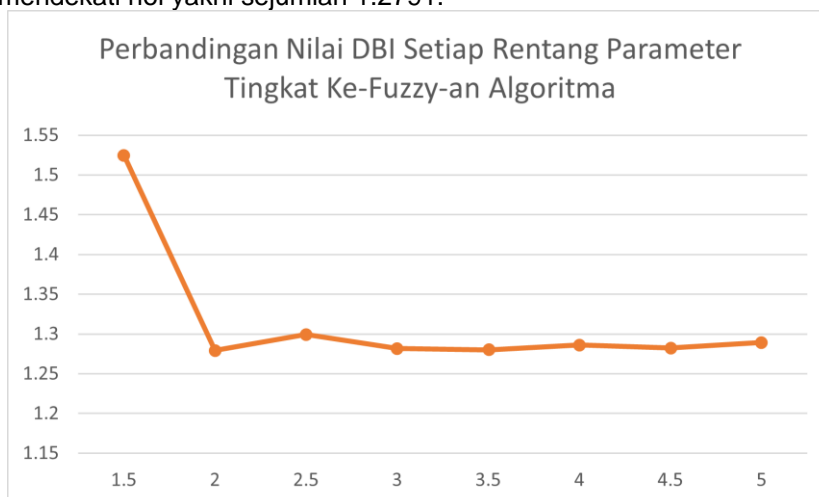
Dalam gambar 6, rata-rata nilai akhir objektif yang optimal dicapai oleh nilai parameter 2 yakni 0.000006. Nilai ini merupakan rata-rata nilai akhir objektif minimum dibandingkan rentang parameter lainnya.



**Gambar 6.** Perbandingan Rata-Rata Nilai Akhir Objektif Setiap Rentang Parameter Tingkat Ke-Fuzzy-an Algoritma

### 3.3. Pengaruh Pada Validitas Klasterisasi

Dalam gambar, ditampilkan nilai DBI setiap rentang parameter tingkat ke-fuzzy-an algoritma. Berdasarkan hasil tersebut, didapat bahwa parameter 2 memiliki nilai DBI optimal yaitu nilai DBI minimum yang mendekati nol yakni sejumlah 1.2791.



**Gambar 7.** Perbandingan Nilai DBI Setiap Rentang Parameter Tingkat Ke-Fuzzy-an Algoritma

Berdasarkan ketiga hasil tersebut, didapat bahwa nilai parameter 2 merupakan nilai optimal untuk tingkat ke-fuzzy-an pada data penjualan automotif Indonesia pada Januari-Agustus 2022 karena nilai tersebut mencapai jumlah iterasi minimum, rata-rata nilai objektif terkecil, dan memiliki nilai validitas yang baik. Hasil klasterisasi menggunakan parameter tersebut didapat untuk klaster dengan total penjualan tinggi sebanyak 13 *brand*, total penjualan sedang sebanyak 36 *brand*, dan total penjualan rendah sebanyak 55 *brand*.

## 4. Kesimpulan

Berdasarkan paparan penelitian sebelumnya, dapat disimpulkan beberapa hal diantaranya:

1. Nilai parameter tingkat ke-fuzzy-an Fuzzy C-Means yang optimal digunakan untuk data penjualan automotif Indonesia adalah 2, dengan parameter lainnya yakni jumlah klaster

- sebanyak 3, nilai *error* yang diharapkan adalah 0.000001, dan jumlah maksimum iterasi 100.
2. Jika ditinjau dari jumlah iterasi, parameter tingkat ke-fuzzy-an yang optimal digunakan adalah nilai parameter yang menghasilkan minimum jumlah iterasi dalam mencapai konvergen.
  3. Jika ditinjau dari nilai objektif akhir, parameter tingkat ke-fuzzy-an yang optimal digunakan adalah nilai parameter yang menghasilkan minimum nilai akhir objektif (jauh lebih kecil dibandingkan parameter nilai *error* yang diharapkan) dalam mencapai konvergen.
  4. Jika ditinjau dari nilai validitas klusterisasi menggunakan DBI, parameter tingkat ke-fuzzy-an yang optimal digunakan adalah nilai parameter yang menghasilkan minimum nilai validitas (semakin kecil mendekati 0).
  5. Perubahan besar kecil nilai parameter tingkat ke-fuzzy-an Fuzzy C-Means tidak memiliki perubahan linear dalam jumlah iterasi yang diperlukan, nilai akhir objektif yang dihasilkan, serta nilai validitas DBI yang didapat karena keoptimalan parameter tersebut ditentukan berdasarkan kecocokannya terhadap data yang digunakan.

## References

- [1] E. Uum, "Penerapan Algoritma Fuzzy C-Means Untuk Pengelompokan Harga Gabah di Tingkat Penggilingan Berdasarkan Kualitas Gabah," Malang, Universitas Islam Negeri Maulana Malik Ibrahim, 2014.
- [2] E. Prabowo and R. Kurniawan, "Optimasi Algoritma Fuzzy Clustering dengan Menggunakan Algoritma Forest Optimization," *Information System Development (ISD)*, vol. 4, pp. 1-6, 2019.
- [3] A. Gupta, "FuzzyClustering to Identify Clusters at Different Levels of Fuzziness: An Evolutionary Multi-Objective Optimization Approach," DeepAI, San Francissco, 2018.
- [4] W. Anggraeni, "Penentuan Nilai Pangkat Pada Algoritma Fuzzy C-Means," *Factor Exacta*, vol. 8(3), no. 1979-276x, pp. 266-278, 2015.
- [5] W. M.A.R and R. , "A novel point density based validity index for clustering gene expression datasets," *International Journal of Data Mining and Bioninformatics*, vol. 17(1), pp. 66-84, 2017.