



# Implementasi Algoritma *Fuzzy C-Means* dalam Pengelompokan Kabupaten/Kota di Provinsi Jawa Timur Berdasarkan Luas Panen, Produksi, dan Produktivitas Padi

Inayatul Maula, Agung Nilogiri, Luluk Handayani

Program Studi Teknik Informatika, Universitas Muhammadiyah Jember

Email: [inayaatulmaula@gmail.com](mailto:inayaatulmaula@gmail.com)\*, [agungnilogiri@unmuhjember.ac.id](mailto:agungnilogiri@unmuhjember.ac.id), [luluk.handayani@unmuhjember.ac.id](mailto:luluk.handayani@unmuhjember.ac.id)

## ABSTRAK

Indonesia sebagai negara agraris sangat bergantung pada sektor pertanian, terutama tanaman padi yang menjadi makanan pokok mayoritas penduduk. Tingginya konsumsi dan pertumbuhan penduduk membuat Indonesia masih bergantung pada impor, sehingga produksi beras harus ditingkatkan. Oleh karena itu, perlu dilakukan upaya untuk memetakan atau mengelompokkan daerah dengan tingkat produktivitas tanaman padi yang belum optimal di provinsi Jawa Timur. Data penelitian yang digunakan adalah data luas panen (ha), produksi (ton), dan produktivitas padi (ton/ha) yang bersumber dari website Badan Pusat Statistik Jawa Timur sebanyak 38 data kabupaten/kota dari Tahun 2020-2023. Data tersebut diolah menggunakan algoritma *Fuzzy C-Means* dilanjutkan penentuan *cluster* optimum menggunakan *Silhouette Coefficient*. Algoritma *Fuzzy C-Means* memungkinkan setiap data dapat memiliki derajat keanggotaan lebih dari satu *cluster*, sehingga menghasilkan pengelompokan yang lebih akurat dan representatif sesuai karakteristik data. Dari hasil uji 2 *cluster* hingga 10 diperoleh jumlah *cluster* optimum yaitu 6 *cluster*, dengan nilai *Silhouette Coefficient* 0,6902.

**Kata Kunci:** padi, pengelompokan, produksi, *Fuzzy C-Means*, *Silhouette Coefficient*

## ABSTRACT

Indonesia as an agrarian country is highly dependent on the agricultural sector, especially rice which is the staple food of the majority of the population. The high consumption and population growth make Indonesia still dependent on imports, so rice production must be increased. Therefore, it is necessary to map or group areas with suboptimal productivity levels of rice plants in East Java province. The research data used is data on harvest area (ha), production (tons), and rice productivity (tons/ha) sourced from the website of the East Java Central Bureau of Statistics as much as 38 district / city data from 2020-2023. The data is processed using the *Fuzzy C-Means* algorithm followed by determining the optimum cluster using the *Silhouette Coefficient*. The *Fuzzy C-Means* algorithm allows each data to have a membership degree of more than one cluster, resulting in a more accurate and representative grouping according to the characteristics of the data. From the test results of 2 clusters to 10, the optimum number of clusters is 6 clusters, with a *Silhouette Coefficient* value of 0.69021.

**Keywords:** rice, clustering, production, *Fuzzy C-Means*, *Silhouette Coefficient*

## 1. PENDAHULUAN

Indonesia dikenal sebagai negara agraris memiliki sektor pertanian yang berkontribusi signifikan terhadap PDB nasional, dengan padi sebagai salah satu komoditas utama (Putri & Fahira, 2021). Produksi padi menjadi sangat penting mengingat perannya dalam menjaga ketahanan pangan nasional. Tingginya ketergantungan masyarakat terhadap beras dan pertumbuhan populasi penduduk menjadikan Indonesia harus mampu memproduksi padi dalam jumlah yang cukup guna memenuhi kebutuhan seluruh rakyatnya dan memastikan stabilitas pangan di seluruh negeri (Wijayanto & Fathoni, 2021). Keberhasilan pemenuhan beras dalam negeri menjadikan kunci kemandirian pangan Indonesia, sehingga pemerintah tidak memerlukan tindakan untuk mengimpor beras dari negara lain. Akan tetapi dalam kenyataannya, Indonesia membutuhkan impor beras karena sulit untuk mencapai swasembada (Paipan & Abrar, 2020).

Berdasarkan angka produksi padi sementara yang dirilis oleh Badan Pusat Statistik, Provinsi Jawa Timur adalah penghasil padi terbesar di Indonesia dengan 9.91 juta ton GKG. Pada tahun 2020, dengan luas panen 1.75 juta ha, Provinsi Jawa Timur juga menduduki peringkat pertama dengan 9.94 juta ton GKG. Provinsi ini menjadi penghasil padi terbesar di Indonesia selama empat tahun berturut-turut hingga tahun 2023. Keberlanjutan pencapaian tersebut terlihat dari laporan resmi BPS Jawa

Timur yang memperkirakan total produksi padi pada tahun 2023 mencapai 9,59 juta ton GKG, mengalami peningkatan dibandingkan tahun 2022 yang sebesar 9,53 juta ton GKG (Bappeda Jawa Timur, 2023).

Produksi padi dipengaruhi oleh luas area lahan dan produktivitas padi. Menurut Nainggolan dkk., (2020) berbagai tantangan seperti alih fungsi lahan, rendahnya kualitas irigasi, dan dampak perubahan iklim yang meningkatkan terjadinya kegagalan panen dan bencana alam. Selain itu, produktivitas padi yang belum optimal disebabkan oleh antara lain; pengendalian hama yang kurang efektif, pemilihan varietas yang kurang adaptif, penggunaan pupuk, benih, dan pestisida yang berkualitas rendah, serta sifat fisik tanah tidak optimal akibat struktur yang buruk sehingga menghambat pertumbuhan tanaman padi (Mafor, 2015). Oleh karena itu, perlu adanya tindakan yang dilakukan sebagai referensi pertimbangan terkait kebijakan untuk memetakan tingkat produktivitas tanaman padi di provinsi Jawa Timur. Dengan demikian, penggunaan metode *clustering* dapat menjadi solusi efektif dalam mengidentifikasi daerah dengan hasil produksi padi yang belum optimal.

Algoritma *Fuzzy C-Means Clustering* (FCM) merupakan salah satu teknik pengelompokan yang termasuk dalam kelompok metode *Hard K-Means*. Metode pengelompokan *fuzzy* yang digunakan memungkinkan data menjadi anggota dari seluruh kelas atau *cluster* yang terbentuk dengan derajat keanggotaan 0–1. Dengan menggunakan persamaan *Fuzzy C-Means*, dapat diketahui bagaimana derajat keanggotaan diperbarui untuk setiap data, sementara pusat *cluster* disesuaikan untuk menghasilkan pengelompokan yang optimal (Sanusi dkk., 2019).

Penerapan *Fuzzy C-Means* dalam penelitian ini merujuk pada studi yang dilakukan oleh Firmansyah dkk., (2023) yang membandingkan kinerja *Fuzzy C-Means* dengan *K-Means* pada data intensitas bencana alam di Indonesia Tahun 2017-2021. Hasil optimasi dengan metode *Elbow* menunjukkan bahwa jumlah *cluster* yang optimum adalah dua. Setelah analisis menggunakan metode *K-Means* dan *Fuzzy C-Means* dengan *Silhouette Coefficient*, diperoleh nilai rata-rata *Silhouette Coefficient* untuk algoritma *Fuzzy C-Means* menunjukkan nilai validasi yang lebih tinggi dibandingkan *K-Means*, yaitu sebesar 0.8403194.

Berdasarkan uraian di atas, adapun tujuan dari penelitian ini adalah untuk menganalisis data dan mengetahui hasil “Implementasi Algoritma *Fuzzy C-Means* Dalam Pengelompokan Kabupaten/Kota di Jawa Timur Berdasarkan Luas Panen, Produksi, dan Produktivitas Padi” menggunakan data BPS Jawa Timur dari tahun 2020–2023.

## 2. KAJIAN PUSTAKA

Beberapa penelitian terdahulu yang relevan sebagai kajian untuk mengetahui keterkaitan dengan penelitian yang akan dilakukan ditunjukkan pada Tabel 1.

Tabel 1. Penelitian terdahulu

No	Nama Peneliti (Tahun)	Judul	Metode	Hasil
1.	Firmansyah dkk., 2023	<i>Clustering</i> Untuk Menentukan Indeks Kesejahteraan Rakyat di Provinsi Jawa Tengah 2022 Menggunakan Metode <i>Fuzzy C-Means</i>	<i>Fuzzy C-Means</i>	Hasil uji validasi <i>cluster</i> menggunakan <i>elbow method</i> dan <i>silhouette score</i> diperoleh empat <i>cluster</i> optimal
2.	Paembonan dkk., 2021	Penerapan Metode <i>Silhouette Coefficient</i> Untuk Evaluasi <i>Clustering</i> Obat	<i>K-Means</i> dan <i>Silhouette Coefficient</i>	Menghasilkan dua <i>cluster</i> optimal dengan nilai akurasi sebesar 0,4854
3.	Ulinnuha (2020)	<i>Provincial Clustering in Indonesia Based on Plantation Production Using Fuzzy C-Means</i>	<i>Fuzzy C-Means</i> dan <i>Silhouette Index</i>	Menghasilkan tiga <i>cluster</i> optimal dengan nilai akurasi sebesar 0,8432

#### A. Luas Panen, Produktivitas, dan Produksi Padi

Luas panen, produktivitas, dan produksi padi merupakan tiga aspek yang saling berkaitan dalam sistem pertanian. Luas panen padi menurut data Badan Pusat Statistik (BPS) pada tahun 2021 mengacu pada luas tanaman padi yang dipanen setelah cukup umur dengan hasil minimal 11% dari kondisi normal. Luas panen dihitung menggunakan Kerangka Sampel Area (KSA) untuk memastikan data yang lebih akurat, cepat, dan modern (Kementerian Pertanian, 2022). Di sisi lain, produktivitas padi menggambarkan rata-rata hasil produksi per satuan luas lahan, dihitung berdasarkan jumlah gabah kering giling (GKG) per hektar dalam satu musim tanam, dan dapat ditingkatkan melalui penerapan sistem tanam yang optimal serta pengaturan jarak tanam yang tepat (Dinas Pertanian dan Pangan, 2022). Sementara itu, produksi padi adalah total hasil panen dalam kurun waktu tertentu. Peningkatan produksi dipengaruhi oleh berbagai faktor, seperti luas lahan yang diolah, penggunaan pupuk, ketersediaan tenaga kerja, serta sistem pengairan yang memadai (Rengganis dkk., 2021). Ketiga komponen ini saling memengaruhi, dimana peningkatan atau penurunan luas panen dan produktivitas akan berdampak langsung pada jumlah produksi, sehingga pengelolaan yang baik terhadap ketiganya sangat penting untuk menjamin keberlanjutan produksi pangan, khususnya beras.

#### B. *Clustering*

*Clustering* adalah teknik dalam data mining yang digunakan untuk menyelesaikan masalah data tanpa label atau kelas yang diketahui. Proses *clustering* bekerja dengan cara mengelompokkan kumpulan data ke dalam beberapa *cluster*. Objek-objek yang ada di dalam *cluster* tersebut memiliki tingkat kemiripan yang tinggi dan tingkat kemiripan yang rendah. *Clustering* termasuk dalam *unsupervised learning* dan juga dikenal dengan istilah data segmentasi, karena metode ini mempartisi banyak data ke dalam banyak grup berdasarkan karakteristiknya (Jumaidi, 2018). Menurut Aselelino & Wijayanto (2023) metode *clustering* secara umum dapat dibagi menjadi dua yaitu *Hierarchical Clustering* dan *Partitional Clustering*.

##### 1) *Partitional Clustering*

*Partitional clustering* adalah metode yang membagi data menjadi sejumlah *non-overlapping* subset atau *cluster* tanpa struktur *hierarki* antara satu dengan lainnya. Secara umum metode ini memiliki fungsi tujuan untuk meminimalisir jarak (*dissimilarity*) dari seluruh data ke pusat *cluster* masing-masing. Contohnya *K-Means*, *Fuzzy C-Means* dan *Mixture Modelling*.

##### 2) *Hierarchical Clustering*

*Hierarchical clustering* adalah metode yang menciptakan *hierarki cluster*. Metode ini menggabungkan dua kelompok yang terdekat di setiap iterasinya ataupun pembagian dari seluruh set data ke dalam satu *cluster*. Contohnya *Single Linkage*, *Complete Linkage*, dan *Average Linkage*.

#### C. *Rstudio*

RStudio adalah *Integrated Development Environment* (IDE) untuk bahasa pemrograman R yang menyediakan fungsi dan ekspresi untuk komputasi statistik dan grafis. Hal ini memungkinkan untuk menulis kode yang ringkas, mudah dibaca, dan dapat melakukan berbagai analisis seperti klasifikasi, regresi, penambahan data dan pembelajaran mesin (R Core Team, 2022).

#### D. *Normalisasi Data*

Normalisasi data adalah proses mengubah data ke dalam bentuk normal dengan tujuan menyederhanakan data, menyederhanakan basis data, dan membuatnya lebih pendek. Ada beberapa metode normalisasi data yang umum digunakan, seperti normalisasi *Z-score*, normalisasi *min-max* dan normalisasi *decimal scaling* (Kusnaldi dkk., 2022). Pada penelitian ini menggunakan normalisasi *min-max* dengan persamaan sebagai berikut (Izonin dkk., 2022)

$$x' = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (1)$$

- $x'$  = nilai hasil normalisasi  
 $x_i$  = nilai tertentu yang akan dinormalisasikan.  
 $\min(x)$  = nilai minimal dari atribut  
 $\max(x)$  = nilai maksimal dari atribut

#### E. Fuzzy C-Means

*Fuzzy C-Means* (FCM) adalah suatu teknik pengklasteran data yang di setiap *cluster* tersebut setiap titik-titik data memiliki nilai derajat keanggotaan (Rizal dkk., 2022). Pada algoritma ini, posisi pusat *cluster* awal belum akurat. Setiap data memiliki derajat keanggotaan terhadap setiap *cluster*. Melalui iterasi yang melibatkan penyesuaian pusat *cluster* dan nilai derajat keanggotaan data, dapat diamati bahwa pusat *cluster* akan mengarah ke lokasi yang lebih tepat. Iterasi ini didasarkan pada upaya meminimalkan fungsi objektif yang mencerminkan jarak dari titik data tertentu ke pusat *cluster*, yang dibobotkan oleh derajat keanggotaan dari titik data tersebut (Rahakbauw dkk., 2017).

- 1) Masukkan data yang akan di *cluster*  $X$ , ke dalam bentuk matriks berukuran  $n \times m$  ( $n$  = jumlah sampel data dan  $m$  = atribut setiap data).  $X_{ij}$  = data ke- $i$  ( $i = 1,2,3, \dots, n$ ) atribut ke-  $j$  ( $j = 1,2, \dots, m$ )
- 2) Tentukan beberapa parameter yang diperlukan, di antaranya:
  - Jumlah *cluster* ( $c$ )
  - Pangkat/bobot ( $w$ )
  - Maksimum iterasi ( $MaxIter$ )
  - *Error* terkecil ( $\xi$ )
  - Fungsi objektif awal ( $P_0 = 0$ )
  - Iterasi awal ( $t = 1$ )
- 3) Bangkitkan bilangan random  $\mu_{ik}$ , sebagai elemen-elemen matriks partisi awal  $U$  (derajat keanggotaan pada *cluster*) dengan  $i = 1,2,3, \dots, n$ ;  $k = 1,2,3, \dots, c$ . Posisi dan nilai matriks dihasilkan secara acak, dengan nilai keanggotaan terletak dalam interval 0 hingga 1. Hitung jumlah setiap kolom,

$$Q_i = \sum_{k=1}^c \mu_{ik} \text{ dengan } j = 1,2,3, \dots, n \quad (2)$$

Hitung:

$$\mu_{ik} = \frac{\mu_{ik}}{Q_{ik}} \quad (3)$$

- 4) Hitung pusat *cluster* ke-  $k$ :  $V_{kj}$  dengan  $k = 1,2,3, \dots, c$  dan  $j = 1,2,3, \dots, m$

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w \times X_{ij})}{\sum_{i=1}^n (\mu_{ik})^w} \quad (4)$$

Dengan,

- $V_{kj}$  = pusat *cluster*  
 $j$  = iterasi  
 $\mu_{ik}$  = perubahan matriks partisi  
 $X_{ij}$  = atribut

- 5) Hitung fungsi objektif pada iterasi ke- $t$ . Fungsi objektif dipakai sebagai kondisi iterasi untuk mendapatkan pusat *cluster* yang tepat. Sehingga diperoleh kesamaan data untuk masuk ke *cluster* mana pada langkah akhir:

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left( \left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] (\mu_{ik})^w \right) \quad (5)$$

- 6) Hitung terjadinya perubahan nilai matriks partisi:

$$\mu_{ik} = \frac{\left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{\frac{-1}{w-1}}}{\sum_{k=1}^c \left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{\frac{-1}{w-1}}} \quad (6)$$

dengan  $i=1,2,3, \dots, n$  dan  $k=1,2,3, \dots, c$

- 7) Cek kondisi hingga berhenti dengan persamaan:

Jika;  $(|Pt - (Pt-1)| < \xi)$  atau  $(t > MaxIter)$  maka berhenti;  
 Jika tidak :  $t = t + 1$ , mengulang ke langkah empat

F. *Silhouette Coefficient*

*Silhouette Coefficient* adalah metrik yang digunakan untuk menghitung jumlah *cluster* optimal dalam data mining. Koefisien ini mengevaluasi objek pada setiap *cluster* dengan memperhatikan nilai *silhouette*. Rentang nilai *Silhouette Coefficient* tiap objek dapat bervariasi antara -1 sampai dengan 1 (Ulinnuha, 2020). Nilai dari *Silhouette Coefficient* menginterpretasikan seberapa dekat suatu data pada satu *cluster* terhadap tetangga pada *cluster* yang lain. Semakin besar nilai *silhouette* (mendekati 1), maka semakin baik *cluster* tersebut (Paembonan dkk., 2021).

$$S_i = \frac{b_i - a_i}{\max[a_i, b_i]} \tag{7}$$

$$SC = \frac{\sum_{i=1}^n S_i}{n} \tag{8}$$

Dengan,

$s_i$  = nilai *silhouette* dari satu titik objek  $i$

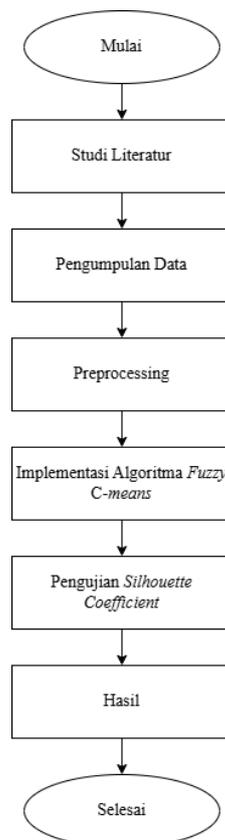
$SC$  = nilai maksimum *mean*  $s_i$  pada seluruh data

$a_i$  = rata-rata jarak objek  $i$  dengan objek lain di *cluster* yang sama

$b_i$  = nilai terkecil antara objek  $i$  dengan objek lain di *cluster* berbeda

### 3. METODE PENELITIAN

Metode penelitian menggunakan algoritma *Fuzzy C-Means*, dan pengujian metode *Silhouette Coefficient* untuk menentukan *cluster* optimum. Tahapan-tahapan yang dilakukan pada penelitian adalah sebagai berikut:



Gambar 1. Tahapan penelitian

A. Studi Literatur

Pada fase awal penelitian ini, dimulai dengan mengidentifikasi dan menganalisis masalah yang akan diteliti yaitu potensi produksi wilayah pertanian di Jawa Timur belum optimal karena ketimpangan dalam luas panen, produksi, dan produktivitas padi di berbagai kabupaten/kota. Selanjutnya, menetapkan latar belakang permasalahan dan menelaah berbagai literatur, seperti jurnal atau artikel yang relevan dengan masalah. Solusi yang dapat dilakukan dengan menggunakan algoritma *Fuzzy C-Means* untuk mengelompokkan kabupaten/kota berdasarkan tiga indikator utama, dan pengujian *Silhouette Coefficient* untuk mengevaluasi kualitas kelompok, sehingga dapat mengidentifikasi wilayah dengan potensi pertanian yang perlu dioptimalkan.

B. Pengumpulan Data

Pengumpulan data menggunakan data sekunder yang diperoleh melalui publikasi pada website Badan Pusat Statistik Provinsi Jawa Timur. Dalam publikasi tersebut, terdapat sejumlah data per kabupaten/kota di Jawa Timur mengenai tanaman padi dan digunakan sebanyak 38 kabupaten/kota tahun 2020-2023 berdasarkan luas panen, produksi dan produktivitas padi.

Tabel 2. Data tanaman padi Provinsi Jawa Timur 2020-2023

Kab/Kota	LP	PT	PR	...	LP	PT	PR
	2020	2020	2020		2023	2023	2023
Kab Pacitan	19360	43,07	83377,94		17923,4	51,88	92993
Kab Blitar	35172	57,38	201803,8		36340,17	66,1	240224
Kab Kediri	36470	59,95	218624,9		31268,33	58,7	183534
Kab Malang	48475	58,4	283096,8		43918,59	63,61	279366
...	...	...	...		...	...	...
Kota Batu	771	66,84	5153,04		727,29	66,86	4863

C. Preprocessing

Normalisasi data dengan metode *min-max* adalah teknik *preprocessing* yang bertujuan untuk merubah skala data numerik ke rentang tertentu, biasanya 0 hingga 1, sehingga setiap fitur memiliki kontribusi yang setara dalam analisis atau pemodelan menggunakan persamaan (1).

Tabel 3. Data yang sudah dinormalisasi

Kab/Kota	LP	PT	PR	...	LP	PT	PR
	2020	2020	2020		2023	2023	2023
Kab Pacitan	0,128	0,000	0,091		0,124	0,153	0,113
Kab Ponorogo	0,462	0,623	0,451		0,474	0,458	0,490
Kab Trenggalek	0,131	0,489	0,120		0,154	0,162	0,140
Kab Tulungagung	0,247	0,725	0,252		0,285	0,437	0,292
Kab Blitar	0,236	0,582	0,227		0,257	0,762	0,298
Kab Kediri	0,245	0,687	0,246		0,220	0,445	0,227
Kab Malang	0,328	0,624	0,320		0,311	0,655	0,347
...	...	...	...		...	...	...
Kota Batu	0,000	0,967	0,001		0,001	0,794	0,002

#### D. Implementasi Algoritma *Fuzzy C-Means*

Tahapan pertama yang dilakukan setelah menormalisasi data yaitu menghitung pusat *cluster*. Selanjutnya, menghitung fungsi objektif berdasarkan jarak antara data dan pusat *cluster* menggunakan derajat keanggotaan *fuzzy*. Kemudian, pusat *cluster* diperbarui menggunakan rata-rata berbobot data, dengan bobot berasal dari matriks keanggotaan. Setelah pembaruan, perubahan nilai matriks keanggotaan dihitung untuk memantau perubahan antar iterasi. Iterasi berlanjut dengan menghitung kembali fungsi objektif, pembaruan matriks keanggotaan, dan pusat *cluster*. Proses ini diulang hingga salah satu kondisi berhenti terpenuhi, yaitu perubahan matriks keanggotaan antar iterasi sangat kecil atau jumlah iterasi telah mencapai batas maksimal. Hasil akhir berupa pusat *cluster* dan matriks keanggotaan *fuzzy*, yang merepresentasikan pembagian data ke dalam *cluster* secara *fuzzy*.

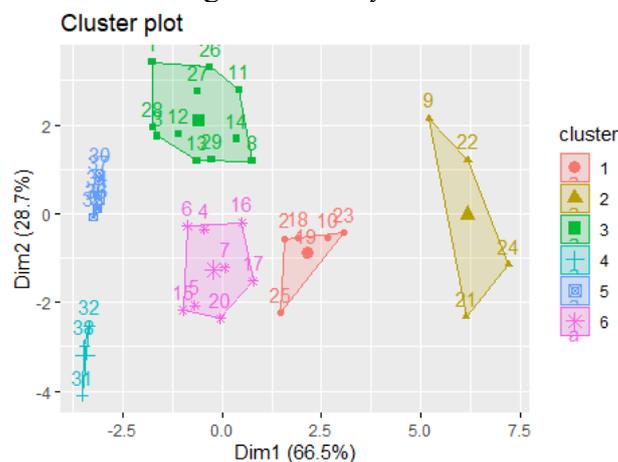
#### E. Uji *Silhouette Coefficient*

Setelah dilakukan perhitungan menggunakan *Fuzzy C-Means* dari 2 *cluster* hingga 10 *cluster*, kemudian menentukan *cluster* optimal menggunakan metode *Silhouette Coefficient*. Langkah awal dalam mencari *cluster* yang optimal dengan melakukan perhitungan nilai rata-rata jarak antara satu objek terhadap seluruh objek lainnya yang berada dalam *cluster* yang sama menggunakan rumus (7).

### 4. HASIL DAN PEMBAHASAN

#### A. *Fuzzy C-Means* pada Rstudio

Data uji yang dimasukkan ke RStudio diolah menggunakan perintah *Fuzzy C-Means* untuk 2 hingga 10 *cluster*. Hasil pemrosesan perintah tersebut muncul di bagian *console*, menampilkan *output* seperti jumlah iterasi, pusat *cluster*, nilai fungsi objektif, dan derajat keanggotaan masing-masing objek terhadap setiap *cluster*. *Cluster* yang telah diuji dapat divisualisasikan dalam bentuk plot dengan simbol dan warna yang berbeda untuk setiap *cluster*. Di bawah ini adalah representasi hasil yang diperoleh dari enam *cluster* algoritma *Fuzzy C-Means* di Rstudio:



Gambar 2. Hasil plot 6 cluster

#### B. Hasil *Silhouette Coefficient* pada Rstudio

Setelah melakukan *clustering* dengan *Fuzzy C-Means*, langkah berikutnya adalah menentukan jumlah *cluster* optimal dengan metode *Silhouette Coefficient*, yang dihitung untuk jumlah *cluster* antara 2 hingga 10. Hasil perhitungan *Silhouette Coefficient* berupa nilai desimal jika, nilai *Silhouette Coefficient* mendekati angka 1 maka pengelompokan pada suatu *cluster* semakin membaik atau optimum. Hasil penerapan metode *Silhouette Coefficient* pada Tabel 4 sebagai berikut.

Tabel 4. Hasil Silhouette Coefficient Pada rstudio

<i>Cluster</i>	<i>Silhouette Coefficient</i>
2	0.6644
3	0.5882
4	0.6539
5	0.6452
6	0.6902
7	0.6585
8	0.6806
9	0.5239
10	0.5020

Dalam metode ini, nilai *cluster* yang digunakan untuk menentukan *cluster* optimal adalah nilai *Silhouette Coefficient* yang paling mendekati nilai 1. Karena nilai *Silhouette Coefficient* pada Tabel 3 menunjukkan bahwa pengujian pada 6 *cluster* memiliki nilai yang mendekati 1 yaitu 0.6902 jika dibandingkan dengan *cluster* lainnya, maka *cluster* 6 memiliki kualitas yang paling baik dalam mengelompokkan data secara jelas. Berikut anggota pada masing-masing *cluster* dan karakteristiknya ditunjukkan pada Tabel 5.

Tabel 5. Hasil pengelompokan kabupaten/kota pada 6 *cluster*

<i>Cluster</i>	Kabupaten/Kota
1	Kabupaten Ponorogo, Banyuwangi, Nganjuk, Madiun, Tuban dan Gresik
2	Kabupaten Ngawi, Jember, Bojonegoro dan Lamongan
3	Kabupaten Pacitan, Trenggalek, Lumajang, Bondowoso, Situbondo, Probolinggo, Pasuruan, Bangkalan, Sumenep, Pamekasan dan Sampang
4	Kota Blitar, Malang dan Batu
5	Kota Kediri, Probolinggo, Pasuruan, Mojokerto, Madiun dan Surabaya
6	Kabupaten Tulungagung, Blita, Kediri, Malang, Sidoarjo, Mojokerto, Jombang dan Magetan

Karakteristik pada masing-masing adalah sebagai berikut:

- 1) *Cluster 1*: Wilayah yang termasuk dalam *cluster 1* memiliki luas panen yang besar dengan tingkat produktivitas yang sedang dan total produksi yang tinggi mencerminkan luasnya lahan pertanian, meskipun efisiensi hasil per hektar tidak terlalu menonjol.
- 2) *Cluster 2*: *Cluster* ini mencakup wilayah dengan luas panen yang sangat besar dan produksi total tertinggi dibandingkan *cluster* lainnya. Namun, produktivitasnya sedikit lebih rendah dibanding beberapa *cluster* lain
- 3) *Cluster 3*: Wilayah dalam *cluster 3* memiliki luas panen yang kecil hingga sedang, disertai produktivitas dan produksi total yang rendah. Kondisi ini mungkin mencerminkan keterbatasan potensi lahan atau adanya hambatan teknis dalam meningkatkan hasil pertanian.
- 4) *Cluster 4*: *Cluster* ini mencakup wilayah dengan luas panen yang sangat kecil, tetapi memiliki produktivitas tertinggi dibandingkan *cluster* lainnya. Meskipun demikian, keterbatasan luas lahan menyebabkan produksi total tetap sangat rendah.
- 5) *Cluster 5*: Wilayah pada *cluster 5* memiliki luas panen yang kecil dengan tingkat produktivitas yang sedang dan produksi total yang rendah. Serupa dengan *cluster 4*, wilayah dalam *cluster* ini mampu memanfaatkan lahan sempit dengan cukup efisien, tetapi hasil produksinya tetap terbatas.

- 6) *Cluster 6*: *Cluster* ini terdiri atas wilayah dengan luas panen sedang, produktivitas yang relatif tinggi, dan produksi total yang signifikan. Wilayah ini menunjukkan keseimbangan yang baik antara luas lahan, tingkat produktivitas, dan hasil produksi.

Dari hasil pengelompokan 6 *cluster*, *cluster 2* teridentifikasi sebagai *cluster* terbaik. Hal ini didukung oleh keunggulan *cluster 2* yang memiliki skala produksi besar dan luas panen yang signifikan, menunjukkan kualitas pengelompokan yang optimal dibandingkan dengan *cluster* lainnya.

## 5. KESIMPULAN

Banyaknya *cluster* optimal jika menggunakan algoritma *Fuzzy C-Means* adalah 6 *cluster*, dengan nilai *Sihouette Coefficient* yaitu 0,6902. Dari hasil pengelompokan dengan 6 *cluster*, menunjukkan variasi karakteristik wilayah berdasarkan luas panen, produktivitas, dan produksi. *Cluster 2* memiliki luas panen dan produksi tertinggi, sementara *cluster 4* mencatat produktivitas tertinggi meskipun luas panennya sangat kecil. *Cluster 1* dan *cluster 6* menunjukkan keseimbangan antara luas panen, produktivitas, dan produksi. Di sisi lain, *cluster 3* menghadapi tantangan terbesar dengan luas panen, produktivitas, dan produksi yang rendah. Hasil ini dapat menjadi dasar untuk prioritas pengembangan wilayah sesuai kebutuhan masing-masing *cluster*.

Saran dari penelitian ini yaitu perlu dilakukan evaluasi kontribusi masing-masing fitur terhadap pembentukan *cluster* karena pada fitur luas panen menunjukkan penurunan konsisten dari tahun 2020 hingga 2023, sementara fitur produksi mengalami penurunan dari tahun 2020 hingga 2022. Analisis korelasi antar fitur dapat dilakukan untuk memastikan bahwa setiap fitur memberikan informasi yang relevan tanpa mendominasi hasil clustering. Selain itu, disarankan untuk modifikasi fungsi jarak pada FCM, seperti *Manhattan* atau *Minkowski Distance*, untuk analisis lebih komprehensif.

## 6. DAFTAR PUSTAKA

- Aselnino, P., & Wijayanto, A. W. (2023). Analisis Perbandingan Metode Hierarchical dan Non-Hierarchical dalam Pembentukan Cluster Provinsi di Indonesia Berdasarkan Indikator Women Empowerment. *Indonesian Journal of Applied Statistics*, 6(1), 57.
- Badan Pusat Statistik. (2021). *Luas Panen dan Produksi Padi di Jawa Timur 2021*.
- Bappeda Jawa Timur. (2023, November 8). Gubernur Khofifah: Jatim Jadi Produsen Tertinggi Nasional 4 Tahun Berturut-turut. <https://bappeda.jatimprov.go.id/2023/11/08/gubernur-khofifah-jatim-jadi-produsen-padi-tertinggi-nasional-4-tahun-berturut-turut/>
- Dinas Pertanian dan Pangan. (2022). *Produktivitas Padi di KT. Tunas Harapan Banjararum Kalibawang Dengan Sistem Tanam Jajar Legowo*. <https://pertanian.kulonprogokab.go.id/detil/1182/produktivitas-padi-di-kt-tunas-harapan-banjararum-kalibawang-dengan-sistem-tanam-jajar-legowo#:~:text=Produktivitas%20padi%20adalah%20produksi%20padi,lahan%2C%20yaitu%20kuintal%20per%20hektar>
- Firmansyah, I., Fauziah, S., Ibrahim, N. H., Fauzi, F., & Utami, T. W. (2023). Clustering Untuk Menentukan Indeks Kesejahteraan Rakyat Di Provinsi Jawa Tengah 2022 Menggunakan Metode Fuzzy C-Means Info Artikel. *Journal of Data Insights*, 1(2), 81–91.
- Izonin, I., Tkachenko, R., Shakhovska, N., Ilchyshyn, B., & Singh, K. K. (2022). A Two-Step Data Normalization Approach for Improving Classification Accuracy in the Medical Diagnosis Domain. *Mathematics*, 10(11).
- Jumaidi, B. (2018). Peningkatan Hasil Evaluasi Clustering Davies-Bouldin Index. Tesis diterbitkan. Medan: Universitas Sumatera utara

- Kementerian Pertanian. (2022). *Pemanfaatan Satelit Untuk Pertanian: Pengukuran Luas Panen, Produksi, dan Produktivitas Padi Dengan Siscrop 2.0*. <https://tanamanpangan.pertanian.go.id/detil-konten/iptek/40>
- Kusnaldi, M. R., Gulo, T., & Aripin, S. (2022). Penerapan Normalisasi Data Dalam Mengelompokkan Data Mahasiswa Dengan Menggunakan Metode K-Means Untuk Menentukan Prioritas Bantuan Uang Kuliah Tunggal. *Journal of Computer System and Informatics (JoSYC)*, 3(4), 330–338.
- Mafor, K. I. (2015). *Analisis Faktor Produksi Padi Sawah di Desa Tompasobaru Dua Kecamatan Tompasobaru*. 6(2), 1–9.
- Nainggolan, P., Chaireni, R., Agustanto, D., & Wahyu, R. A. (2020). Ketahanan Pangan Berkelanjutan. *Jurnal Kependudukan Dan Pembangunan Lingkungan*, 2, 23–32.
- Paembonan, S., Abduh, H., & Kunci, K. (2021). Penerapan Metode Silhouette Coefficient Untuk Evaluasi Clustering Obat Clustering; K-means; Silhouette coefficient. *Jurnal Ilmiah Ilmu-Ilmu Teknik*, 6(2), 48–54.
- Paipan, S., & Abrar, M. (2020). Analisis Kondisi Ketergantungan Impor Beras di Indonesia. *Jurnal Perspektif Ekonomi Darussalam*, 6(2), 212.
- Putri, R. K., & Fahira, A. (2021). Observasi Faktor Pendorong Produksi Padi (Studi Kasus Kecamatan Tambakdahan, Subang). *Jurnal Riset Ilmu Ekonomi*, 1(3), 131–140. [www.jrie.feb.unpas.ac.id](http://www.jrie.feb.unpas.ac.id)
- R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org>
- Rahakbauw, D. L., Ilwaru, V. Y. I., & Hahury, M. H. (2017). *Implementasi Fuzzy C-Means Clustering Dalam Penentuan Beasiswa* (Vol. 11).
- Rengganis, A. P., Arifianto, D., & Oktavianto, H. (2021). Algoritma Partitioning Around Medoids (PAM) dalam Mengelompokkan Kecamatan di Kabupaten Jember Berdasarkan Produktivitas Tanaman Padi Partitioning Around Medoids (PAM) Algorithm in Classifying Districts in Jember Regency Based on Rice Productivity. *Jurnal Smart Teknologi*, 1(1), 100–102.
- Rizal, A., Novitasari, D. C. R., & Hafiyusholeh, Moh. (2022). Pengelompokan Karyawan Berdasarkan Kesalehan Menggunakan Perbandingan Fuzzy C-Means, K-Means, dan Probabilistic Distance Clustering. *Jurnal Fourier*, 11(2), 69–77.
- Sanusi, W., Zaky, A., & Besse, N. A. (2019). Analisis Fuzzy C-Means dan Penerapannya Dalam Pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan Berdasarkan Faktor-faktor Penyebab Gizi Buruk. In *Journal of Mathematics, Computations, and Statistics* (Vol. 2, Issue 1). <http://www.ojs.unm.ac.id/jmathcos>
- Ulinuha, N. (2020). Provincial Clustering in Indonesia Based on Plantation Production Using Fuzzy C-Means. *Jurnal Ilmiah Teknologi Dan Informasi*, 9(1), 8–12.
- Wijayanto, S., & Fathoni, M. Y. (2021). Pengelompokan Produktivitas Tanaman Padi di Jawa Tengah Menggunakan Metode Clustering K-Means. *Jurnal JUPITER*, 13(2), 212–219.