Jurnal Esensi Infokom Vol 9 No. 1 Mei 2025
e-ISSN: 2828-6707
https://esensijournal.com/index.php/infokom
DOI: 10.55886/infokom.v9i1.949

Perbandingan Algoritma K-Means dan Fuzzy C-Means untuk Klasterisasi Produktivitas Kedelai di Pulau Jawa

Jefri Jaya[1], Teny Handhayani[2*]

[1,3]*Fakultas Teknologi Informasi, Universitas Tarumanagara*
Jl. S.Parman No 1.Jakarta Barat INDONESIA
[1]jefri.535210048@stu.untar.ac.id, [2*]tenyh@fti.untar.ac.id

Intisari— **Kedelai merupakan tanaman pangan yang memiliki nilai ekonomi yang tinggi dan merupakan salah satu bahan pangan yang sering di konsumsi oleh masyarakat Indonesia. Meskipun memiliki potensi besar dalam mendukung ketahanan pangan nasional, Indonesia masih menghadapi tantangan dalam memenuhi kebutuhan kedelai secara mandiri. Hal ini disebabkan oleh keterbatasan produksi lokal yang belum mampu mengimbangi tingginya permintaan masyarakat. Akibatnya, Indonesia masih bergantung pada impor kedelai dari luar negeri untuk memenuhi kebutuhan domestik. Penelitian ini bertujuan menghasilkan informasi berupa pengelompokkan wilayah, menganalisis pola tren pertumbuhan, dan menentukan model klasterisasi yang paling optimal dari K-Means dan Fuzzy C-Means dengan menggunakan data tanaman pangan Kedelai di Pulau Jawa. Data yang digunakan berupa *time series* tahunan dari tahun 2010 hingga 2022, yang diperoleh dari situs Basis Data Statistik Pertanian. Hasil evaluasi *silhouette* dari metode K-Means dan Fuzzy C-Means memiliki hasil yang serupa dalam menentukan klasterisasi data berdasarkan nilai *k* yang paling optimal berada di 0.4659 dengan jumlah klaster *k* = 2. Tiap klaster terbagi menjadi 2 klaster, klaster 0 memiliki 66 wilayah dengan hasil produktivitas rendah, klaster 1 memiliki 16 wilayah dengan hasil produktivitas tinggi.**
Kata kunci—Fuzzy C-Means, K-Means, Klasterisasi, Kedelai, Silhouette

***Abstract*— Soybeans are an agricultural commodity with high economic value and are one of the staple foods frequently consumed by the Indonesian population. Despite their significant potential to support national food security, Indonesia still faces challenges in meeting soybean demand independently. This is due to the limited local production, which cannot keep up with the increasing demand. As a result, Indonesia remains reliant on soybean imports from other countries to fulfill domestic needs. This study aims to produce information regarding the grouping of regions, analyze growth trend patterns, and determine the most optimal clustering model between K-Means and Fuzzy C-Means using soybean crop data in Java Island. The data utilized are annual time series data from 2010 to 2022, obtained from the official website of the Basis Data Statistik Pertania. The silhouette evaluation results for the K-Means and Fuzzy C-Means methods show similar outcomes in determining data clustering, with the most optimal k-value at 0.4659 and the number of clusters set to k = 2. Each klaster is divided into two groups: klaster 0 comprises 66 regions with low productivity, while klaster 1 consists of 16 regions with high productivity.**
Keywords— Fuzzy C-Means, K-Means, Clustering, Soybeans, Silhouette

I. PENDAHULUAN

Produksi kedelai di Indonesia menghadapi berbagai kendala, salah satunya adalah penurunan kualitas lahan akibat kurang penggunaan pupuk organik [1]. Kondisi ini berdampak pada rendahnya tingkat produktivitas kedelai di sejumlah daerah. Sebagai solusi, penggunaan pupuk organik disarankan untuk memulihkan kesuburan tanah dan meningkatkan hasil produksi [2]. Kedelai sendiri memiliki peran yang sangat vital sebagai bahan baku utama untuk berbagai produk pangan berbasis protein nabati sehingga menjadikannya komoditas yang strategis di Indonesia [3]. Produksi kedelai di Indonesia masih belum cukup untuk memenuhi kebutuhan konsumsi dalam negeri, sehingga negara Indonesia masih bergantung pada impor [4]. Namun, dalam dua dekade terakhir, produksi kedelai di Jawa Barat mengalami peningkatan yang cukup signifikan, meskipun tetap ada fluktuasi setiap tahunnya. Peningkatan tersebut didorong oleh dua faktor utama, yaitu meluasnya area panen dan meningkatnya produktivitas. Perluasan luas panen terjadi dalam jangka panjang melalui ekspansi sawah dan ladang, sementara dalam jangka pendek, peningkatan panen juga dipengaruhi oleh penurunan lahan yang mengalami gagal panen serta peningkatan Indeks Pertumbuhan [5]. Kedelai memiliki nilai ekonomi yang sangat tinggi dan merupakan salah satu bahan pangan yang paling banyak dikonsumsi oleh masyarakat Indonesia [6].

Metode yang digunakan dalam penelitian ini adalah K-Means dan Fuzzy C-Means yang merupakan metode klasterisasi. Tujuan akhir yang ingin dicapai dari penelitian ini adalah menghasilkan informasi yang komprehensif berupa pengelompokkan wilayah berdasarkan produktivitas kedelai, mengidentifikasi pola tren pertumbuhan produktivitas kedelai di Pulau Jawa selama rentang waktu 2010 hingga 2022, serta menentukan metode klasterisasi yang paling optimal di antara K-Means dan Fuzzy C-Means. Dengan pendekatan ini, penelitian ini diharapkan dapat memberikan kontribusi yang

signifikan terhadap pengembangan strategi peningkatan produktivitas kedelai di Pulau Jawa secara lebih terarah dan efektif.

II. REVIEW LITERATUR

A. Klasterisasi

Klasterisasi merupakan proses pengelompokan data atau objek ke dalam klaster berdasarkan kemiripan atribut yang dimiliki dalam setiap kelompok [7]. Klasterisasi yang efektif akan menghasilkan kelompok yang terdiri dari objek dengan tingkat kemiripan tinggi dalam satu klaster, tetapi memiliki perbedaan yang signifikan dengan objek di klaster lainnya [8]. Selain itu, klasterisasi berperan penting dalam berbagai permasalahan, seperti analisis pola musiman dan pengambilan keputusan [9].

B. K-Means

K-Means adalah salah satu metode pengelompokan yang digunakan untuk mengelompokkan titik data ke dalam sejumlah klaster yang telah ditentukan [10]. Metode ini salah satu yang melakukan pengelompokan data dengan sistem partisi dan berupaya mengelompokkan berbagai data yang ada ke dalam tiap kelompok, dimana data menjadi satu kelompok mempunyai karakteristik yang sama sedangkan karateristik yang berbeda akan dikelompokkan ke dalam kelompok yang lain [11] [12].

C. Fuzzy C-Means (FCM)

Fuzzy C-Means (FCM) adalah algoritma klasterisasi dengan logika fuzzy yang menetapkan titik data ke beberapa klaster dengan tingkat keanggotaan yang berbeda-beda, sehingga menempatkannya ke dalam satu kelompok [13]. Metode Fuzzy C-Means sendiri diperkenalkan oleh Jim Bezdek pada tahun 1981 [14]. FCM termasuk dalam kategori teknik pembelajaran tanpa pengawasan, sehingga sangat cocok untuk kumpulan data yang berisi *outlier* titik yang menyimpang secara signifikan dari data lainnya [15].

D. Elbow

Elbow adalah metode yang digunakan untuk menentukan jumlah klaster yang optimal dalam proses klasterisasi dengan cara menganalisis penurunan *Within Cluster Sum of Squares* (WCSS) seiring dengan bertambahnya jumlah klaster [16]. Dalam metode ini, dilakukan penghitungan dan pemodelan klasterisasi untuk berbagai jumlah klaster (k), lalu hasilnya dipetakan dalam bentuk grafik yang menunjukkan hubungan antara jumlah klaster dengan nilai WCSS [17]. Titik tertentu pada grafik, yang membentuk pola seperti siku, menunjukkan jumlah klaster yang ideal [18]. WCSS dapat dilakukan dengan persamaan (1).

$$SSE = \sum_{K=1}^{K} \sum x_i \ |x_i - \ c_k \ |^2 \qquad (1)$$

Keterangan:
K = klaster ke-c
x_i = jarak data objek ke-i
C_k = pusat klaster ke-i

E. Fuzzy Partition Coefficient (FPC)

Fuzzy Partition Coefficient (FPC) adalah metode untuk mengukur validitas klasterisasi dengan menilai tingkat tumpang tindih antar klaster [19]. Hasil FPC yang baik dapat ditentukan berdasarkan tingginya nilai FPC, yang menunjukkan bahwa pembagian data ke dalam setiap klaster terpisah dengan jelas dan efektif. Semakin tinggi nilai FPC, semakin sedikit tumpang tindih antara klaster yang ada, menandakan bahwa titik data lebih kuat keanggotannya pada klaster tertentu [20]. FPC dapat dilakukan dengan persamaan (2).

$$FPC = \left(\frac{1}{N}\right) \sum_{C=1}^{C} \sum_{i=1}^{N} u_{ci}^2 \qquad (2)$$

Keterangan:
N = jumlah data
C = jumlah klaster
u_{ci} = derajat keanggotaan

F. Silhouette

Silhouette merupakan salah satu metode evaluasi gabungan dari dua metode yaitu metode *cohesion* yang bertujuan untuk mengukur seberapa dekat objek dengan data objek lainnya dalam klaster yang sama dan metode *separation* yang mengukur seberapa jauh data objek dari klaster lain [21]. Selain itu *silhouette* digunakan untuk dapat melihat kualitas dan kekuatan pada setiap klaster dengan cara menghitung skor untuk setiap titik data berdasarkan jarak rata-rata ke titik lain dalam klaster yang sama dan jarak rata-rata ke titik di klaster tetangga terdekat [8]. Skor yang dihasilkan berkisar antara -1 hingga 1, dengan nilai yang mendekati 1 menunjukkan klaster yang terpisah dengan baik, nilai yang mendekati 0 menunjukkan klaster yang tumpang tindih, dan nilai negatif menunjukkan potensi kesalahan klasifikasi [22]. Indeks *Silhouette* keseluruhan diperoleh dengan merata-ratakan skor seluruh titik data, sehingga memberikan penilaian keseluruhan terhadap struktur pengelompokan. *Silhouette* dapat dilakukan dengan persamaan (3).

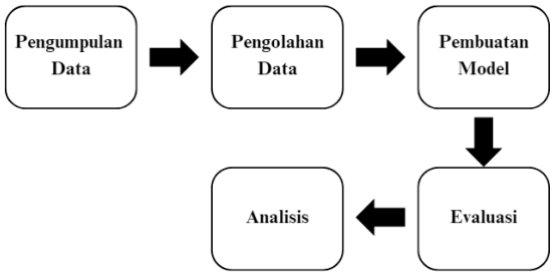
$$S(i) = \frac{b(i) - a(i)}{\max\left(a(i), b(i)\right)} \qquad (3)$$

Keterangan:
$S(i)$ = nilai silhouette
$a(i)$ = rata-rata jarak antara data *i* dengan setiap data lain yang tergabung dalam klaster yang sama
$b(i)$ = nilai terkecil dari rata-rata jarak antara data i dan data lain yang berada dalam klaster yang berbeda

Jurnal Esensi Infokom Vol 9 No. 1 Mei 2025
e-ISSN: 2828-6707
https://esensijournal.com/index.php/infokom
DOI: 10.55886/infokom.v9i1.949

III. METODOLOGI PENELITIAN

Melakukan penelitian perbandingan klasterisasi menggunakan metode K-Means dan Fuzzy C-Means terhadap data kedelai di Pulau Jawa, terdapat beberapa tahapan dalam melakukannya. Tahapan penelitian seperti berikut:



Gambar 1. Tahapan Penelitian

A. Pengumpulan Data

Sekumpulan data yang digunakan untuk kebutuhan penelitian ini menggunakan data tanaman pangan kedelai di Pulau Jawa yang terdiri dari lokasi, luas, produksi, dan produktivitas. Data yang digunakan adalah data *time series* tahunan dari tahun 2010 hingga 2022 dengan total 119 kabupaten atau kota yang diperoleh dari situs Basis Data Statistik Pertanian.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 119 entries, 0 to 118
Data columns (total 40 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   Lokasi              119 non-null    object
 1   Luas_2010           119 non-null    int64
 2   Luas_2011           119 non-null    int64
 3   Luas_2012           119 non-null    int64
 4   Luas_2013           119 non-null    int64
 5   Luas_2014           119 non-null    int64
 6   Luas_2015           119 non-null    int64
 7   Luas_2016           119 non-null    float64
 8   Luas_2017           119 non-null    float64
 9   Luas_2018           119 non-null    float64
 10  Luas_2019           119 non-null    float64
 11  Luas_2020           119 non-null    float64
 12  Luas_2021           119 non-null    float64
 13  Luas_2022           119 non-null    float64
 14  Produksi_2010       119 non-null    int64
 15  Produksi_2011       119 non-null    int64
 16  Produksi_2012       119 non-null    int64
 17  Produksi_2013       119 non-null    float64
 18  Produksi_2014       119 non-null    int64
 19  Produksi_2015       119 non-null    float64
...
 38  Produktivitas_2021  119 non-null    float64
 39  Produktivitas_2022  119 non-null    float64
dtypes: float64(29), int64(10), object(1)
memory usage: 37.3+ KB
```

Gambar 2. Variabel Data Kedelai di Pulau Jawa

B. Pengolahan Data

Data yang telah dikumpulkan kemudian diproses melalui beberapa tahapan. Pertama, data lokasi yang memiliki nilai 0 selama lebih dari 4 tahun akan dihapus dari dataset. Selanjutnya, untuk data yang memiliki nilai 0 yang kurang dari 4 tahun, nilai tersebut akan diisi menggunakan teknik interpolasi linear. Hasil data yang telah dibersihkan akan digunakan ke tahap selanjutnya berjumlah 82 kabupaten atau kota. Setelah proses ini selesai, data akan ditransformasi menggunakan *Min-Max Scaling* untuk mengubah nilai-nilai data ke dalam rentang antara 0 dan 1.

	0	1	2	3	4	5	6	7	8	9	...	29
0	0.055396	0.082634	0.086680	0.223937	0.375329	0.256372	0.133404	0.092198	1.000000	0.556210	...	0.038207
1	0.178455	0.184703	0.191431	0.083146	0.486576	0.312163	0.216737	0.106334	0.702910	0.465355	...	0.025716
2	0.003039	0.001024	0.002024	0.002972	0.009425	0.016644	0.026730	0.031909	0.303133	0.171531	...	0.042891
3	0.318548	0.268844	0.481652	0.388532	0.523899	0.505878	0.265262	0.232514	0.362865	0.869535	...	0.045861
4	0.035992	0.062374	0.042291	0.073231	0.106362	0.157439	0.204368	0.171014	0.466590	0.421274	...	0.036768
...	...	...	...	...	...	...	...	...	...	...	...	...
77	0.069252	0.161711	0.133461	0.312087	0.329853	0.344559	0.345966	0.342953	0.281273	0.436722	...	0.006858
78	0.590817	0.632331	0.759100	0.744466	0.865671	1.000000	1.000000	0.829311	0.505360	0.423975	...	0.016593
79	0.023140	0.014228	0.045751	0.020953	0.037736	0.058118	0.024142	0.033778	0.038470	0.021879	...	0.033401
80	0.277510	0.205389	0.248445	0.197192	0.192308	0.181300	0.126242	0.378151	0.582776	0.389730	...	0.038636
81	0.000028	0.000000	0.000552	0.000637	0.000808	0.000815	0.000158	0.000051	0.000100	0.000361	...	0.030982

82 rows × 39 columns

Gambar 3. Data Setelah di Normalisasi

C. Pembuatan Model

Setelah melalui tahapan preprocessing data, langkah berikutnya adalah membangun model untuk melakukan proses klasterisasi pada data produktivitas kedelai di Pulau Jawa. Proses ini bertujuan untuk mengelompokkan data berdasarkan karakteristik tertentu, sehingga dapat memberikan wawasan yang lebih mendalam mengenai pola yang terdapat dalam data tersebut. Dalam penelitian ini, metode klasterisasi yang akan di uji coba untuk pemilihan hasil yang paling optimal adalah K-Means dan Fuzzy C-Means, yang masing-masing memiliki pendekatan dan karakteristik berbeda dalam pengelompokan data. Kedua metode ini akan dibandingkan untuk menentukan metode yang lebih efektif dalam melakukan klasterisasi data produktivitas kedelai, dengan mempertimbangkan keakuratan hasil pengelompokan, tingkat kemiripan dalam klaster, dan fleksibilitas masing-masing algoritma.

D. Evaluasi

Dengan menentukan model yang lebih optimal dari K-Means dan Fuzzy C-Means untuk penelitian ini, menggunakan model evaluasi Silhouette Score. Silhouette Score mengukur seberapa dekat objek dengan data objek lainnya dalam klaster yang sama dan metode separation yang mengukur seberapa jauh data objek dari klaster lain. Selain itu digunakan untuk dapat melihat kualitas dan kekuatan pada setiap klaster dengan cara menghitung skor untuk setiap titik data berdasarkan jarak rata-rata ke titik lain dalam klaster yang sama dan jarak rata-rata ke titik di klaster tetangga terdekat. Skor yang dihasilkan berkisar antara -1 hingga 1, dengan nilai yang mendekati 1 menunjukkan klaster yang terpisah dengan baik dan kompak, nilai yang mendekati 0 menunjukkan klaster yang tumpang

tindih, dan nilai negatif menunjukkan potensi kesalahan klasifikasi.

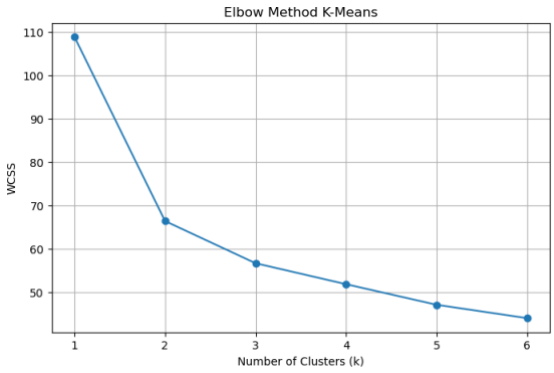
E. Analisis

Setelah evaluasi terhadap model K-Means dan Fuzzy C-Means dilakukan, langkah selanjutnya adalah menganalisis hasil untuk menentukan model yang paling optimal kemudian mengelompokkan wilayah berdasarkan data produktivitas kedelai dengan produktivitas rendah dan tinggi secara jelas, sehingga memberikan gambaran yang lebih terstruktur mengenai distribusi produktivitas di Pulau Jawa. Selanjutnya, analisis akan difokuskan pada perkembangan produktivitas kedelai dari tahun 2010 hingga 2022 dengan menggunakan data *time series*. Tujuannya adalah untuk mengidentifikasi pola tren, dan perubahan signifikan yang terjadi selama periode tersebut.

IV. HASIL DAN PEMBAHASAN

A. Hasil K-Means

Hasil elbow untuk mengamati WCSS *(Within Klaster Sum of Squares)* setiap klaster pada model K-Means, terlihat grafik penurunan WCSS yang cukup tajam dari klaster 1 ke klaster 2 yang diikuti dengan penurunan yang lebih bertahap. menuju klaster 3. Di luar klaster 3, laju penurunan menjadi dapat diabaikan, yang menunjukkan semakin berkurangnya keuntungan dalam hal kekompakan klaster. Pola ini merupakan karakteristik titik siku, dimana menambahkan lebih banyak klaster di luar titik ini tidak secara signifikan meningkatkan kemampuan model untuk mengelompokkan titik data serupa. Berdasarkan analisis ini, dapat disimpulkan bahwa jumlah klaster optimal untuk dataset ini adalah klaster 2.



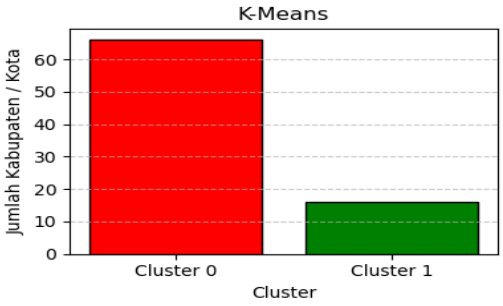
Gambar 4. Grafik Elbow

Hasil evaluasi klasterisasi data kedelai dengan menggunakan model K-Means dilakukan dengan nilai klaster k = {2, 3, 4, 5, 6, 7, 8, 9, 10, 11}. Dari hasil evaluasi *silhouette* pada model K-Means yang optimal berada di klaster 2 memiliki hasil evaluasi 0.4659. Hasil evaluasi *silhouette* K-Means dapat dilihat pada Tabel 1.

TABEL 1. Evaluasi *Silhouette* K-Means

Metode	Klasterisasi	*Silhouette*
K-Means	2	0.4659
	3	0.3603
	4	0.1159
	5	0.1479
	6	0.1535
	7	0.1086
	8	0.1124
	9	0.1156
	10	0.1278
	11	0.1171

Hasil klasterisasi dengan nilai k = 2 menunjukkan bahwa pada klaster 0 terdapat 66 kabupaten atau kota di Pulau Jawa, meliputi Kabupaten Bandung, Tasikmalaya, Ciamis, Kuningan, Cirebon, Majalengka, Sumedang, Indramayu, Subang, Purwakarta, Karawang, Cilacap, Banyumas, Purbalingga, Banjarnegara, Kebumen, Purworejo, Wonosobo, Boyolali, Klaten, Sukoharjo, Karanganyar, Sragen, Blora, Rembang, Pati, Kudus, Jepara, Demak, Semarang, Temanggung, Kendal, Batang, Pekalongan, Pemalang, Tegal, Brebes, Pandeglang, Lebak, Serang, Kulon Progo, Bantul, Sleman, Pacitan, Trenggalek, Tulungagung, Kediri, Malang, Lumajang, Bondowoso, Situbondo, Probolinggo, Sidoarjo, Mojokerto, Jombang, Madiun, Magetan, Tuban, Gresik, Bangkalan, Pamekasan, dan Sumenep, serta Kota Tasikmalaya, Banjar, Tangerang Selatan, dan Kediri. Sementara itu, klaster 1 terdiri dari 16 kabupaten, yaitu Sukabumi, Cianjur, Garut, Wonogiri, Grobogan, Gunung Kidul, Ponorogo, Blitar, Jember, Banyuwangi, Pasuruan, Nganjuk, Ngawi, Bojonegoro, Lamongan, dan Sampang.

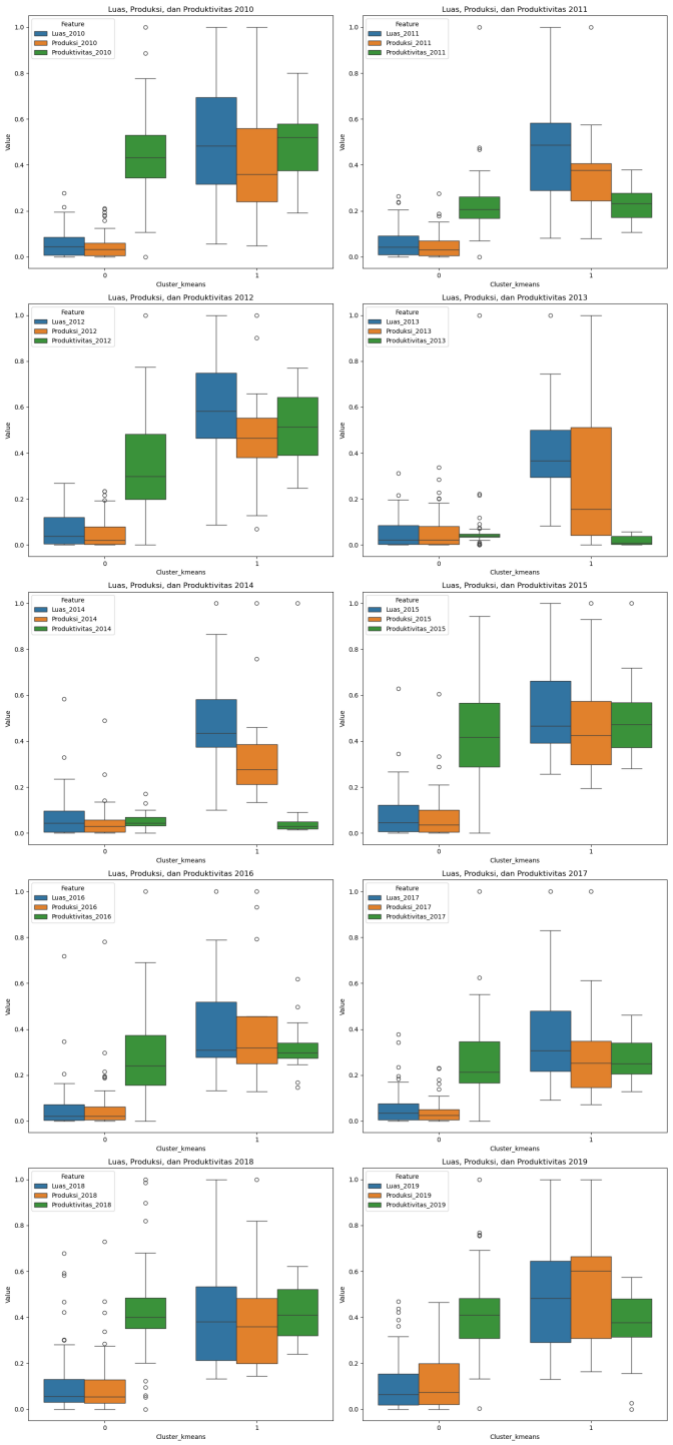


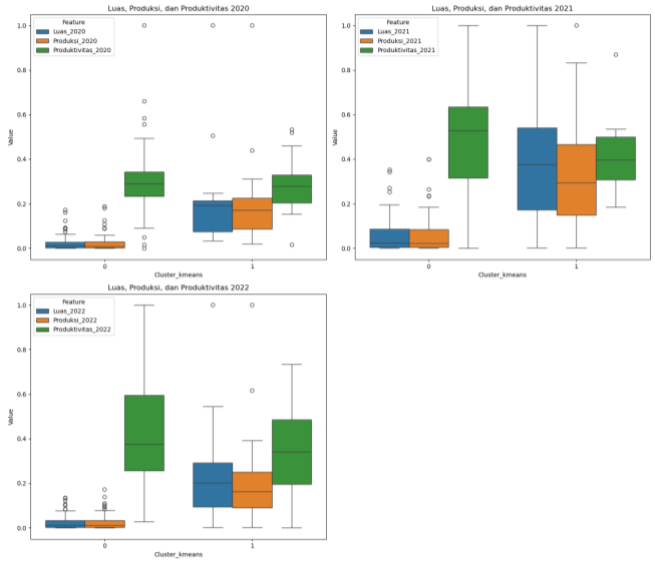
Gambar 5. Visualisasi Hasil Klasterisasi dengan K-Means

Setelah itu dilakukan perbandingan distribusi pada setiap fitur luas, produksi, dan produktivitas pada setiap tahunnya. Klaster 0 cenderung memiliki nilai distribusi luas, produksi dan produktivitas yang secara keseluruhan lebih rendah. Di sisi lain, klaster 1 memiliki nilai distribusi yang lebih tinggi, dengan distribusi yang lebih tersebar. Hal ini mengindikasikan bahwa data pada klaster ini memiliki variasi yang lebih besar dan mencakup rentang nilai yang lebih tinggi.

Dari tahun 2010 hingga 2022, terlihat perbedaan yang konsisten antara klaster 0 dan klaster 1 dalam produksi, dan produktivitas. Pada periode 2010 hingga 2013, perbedaan antar klaster masih belum terlalu signifikan, meskipun klaster 1

menunjukkan nilai yang lebih tinggi. Pada periode 2014 hingga 2017, perbedaan semakin jelas, dengan klaster 1 menunjukkan peningkatan signifikan pada produktivitas dan variasi yang lebih besar, sementara klaster 0 tetap stabil dengan nilai lebih rendah. Pada periode 2018 hingga 2020, fluktuasi produksi di klaster 1 semakin terlihat, tetapi klaster ini tetap memiliki nilai median yang lebih tinggi dibandingkan klaster 0. Di tahun 2021 hingga 2022, klaster 1 tetap mendominasi dengan memiliki produktivitas yang lebih tinggi, menunjukkan wilayah-wilayah tertentu yang terus berkembang lebih baik dibandingkan wilayah lain di klaster 0 yang cenderung stagnan.

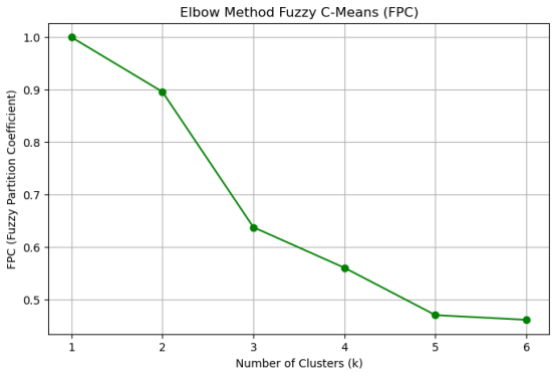




Gambar 6. Distribusi Data Kedelai dengan Metode K-Means

B. *Hasil Fuzzy C-Means*

Nilai menggunakan FPC dapat disimpulkan bahwa klaster 2 merupakan yang optimal dibandingkan klaster yang lain, dikarenakan memiliki FPC berada di angka 0.9, yang menunjukkan pemisahan antar klaster yang jelas dengan sedikit tumpang tindih. Pada klaster dengan nilai yang lebih rendah, seperti klaster 1 (FPC 1.0), meskipun partisi sangat jelas, model cenderung terlalu sederhana dan tidak dapat menangkap kompleksitas data secara optimal. Sebaliknya, pada klaster 3 dan klaster 4, terjadi penurunan signifikan pada nilai FPC, yang menandakan adanya tumpang tindih yang lebih besar antar klaster dan semakin menurunnya kualitas pemisahan data. Oleh karena itu, klaster 2 dengan nilai FPC 0.9 memberikan hasil yang cukup seimbang, antara kejelasan pemisahan data dan kemampuan untuk menangkap variasi dalam data, menjadikannya pilihan yang paling optimal untuk klastering dalam penggunaan model Fuzzy C-Means.



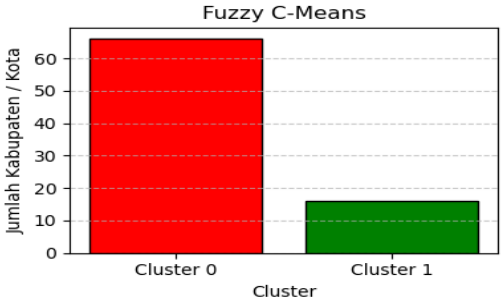
Gambar 7. Grafik Fuzzy Partition Coefficient

Hasil evaluasi model Fuzzy C-Means dilakukan dengan jumlah klaster $k = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}$. Di antara jumlah klaster tersebut, $k = 2$ muncul sebagai yang paling optimal dengan hasil 0,4659. hasil ini lebih tinggi dibandingkan klaster lainnya, yang menunjukkan bahwa titik data dalam klaster ini lebih berbeda dan terpisah dengan klaster lainnya. Sedangkan, hasil untuk klaster lainnya turun di bawah 0,21, menunjukkan pemisahan klaster yang lebih lemah dan kelompok yang kurang terdefinisi dengan baik. Hasil yang lebih rendah ini menyiratkan bahwa titik-titik data dalam klaster-klaster ini lebih tersebar atau lebih dekat dengan batas keputusan antar klaster, sehingga menyebabkan tingkat tumpang tindih yang lebih tinggi dan diferensiasi yang kurang jelas. Oleh karena itu, berdasarkan hasil evaluasi $k = 2$ dianggap sebagai pilihan paling optimal.

TABEL 2. Evaluasi *Silhouette* Fuzzy C-Means

Metode	Klaster	*Silhouette*
Fuzzy C-Means	2	0.4659
	3	0.2015
	4	0.1338
	5	0.1065
	6	0.1115
	7	0.0759
	8	0.0878
	9	0.0868
	10	0.0851
	11	0.0906

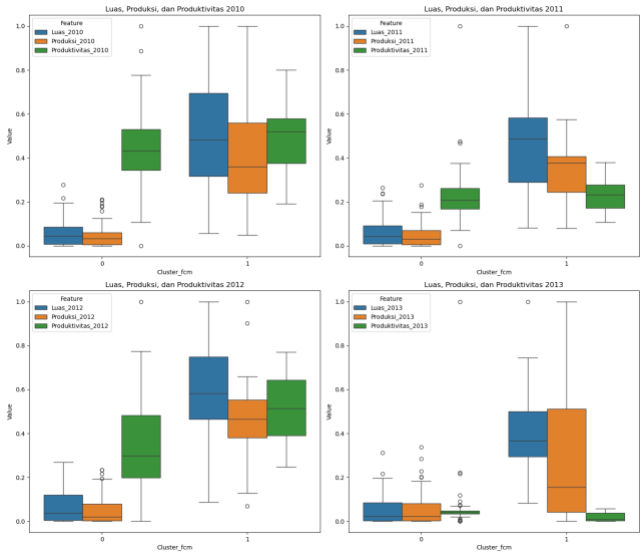
Pada model Fuzzy C-Means dengan nilai k yang paling optimal, yaitu $k = 2$, Hasil klasterisasi menghasilkan klaster 0 terdapat 66 kabupaten atau kota di Pulau Jawa, meliputi Kabupaten Bandung, Tasikmalaya, Ciamis, Kuningan, Cirebon, Majalengka, Sumedang, Indramayu, Subang, Purwakarta, Karawang, Cilacap, Banyumas, Purbalingga, Banjarnegara, Kebumen, Purworejo, Wonosobo, Boyolali, Klaten, Sukoharjo, Karanganyar, Sragen, Blora, Rembang, Pati, Kudus, Jepara, Demak, Semarang, Temanggung, Kendal, Batang, Pekalongan, Pemalang, Tegal, Brebes, Pandeglang, Lebak, Serang, Kulon Progo, Bantul, Sleman, Pacitan, Trenggalek, Tulungagung, Kediri, Malang, Lumajang, Bondowoso, Situbondo, Probolinggo, Sidoarjo, Mojokerto, Jombang, Madiun, Magetan, Tuban, Gresik, Bangkalan, Pamekasan, dan Sumenep, serta Kota Tasikmalaya, Banjar, Tangerang Selatan, dan Kediri. Sementara itu, klaster 1 terdiri dari 16 kabupaten, yaitu Sukabumi, Cianjur, Garut, Wonogiri, Grobogan, Gunung Kidul, Ponorogo, Blitar, Jember, Banyuwangi, Pasuruan, Nganjuk, Ngawi, Bojonegoro, Lamongan, dan Sampang.

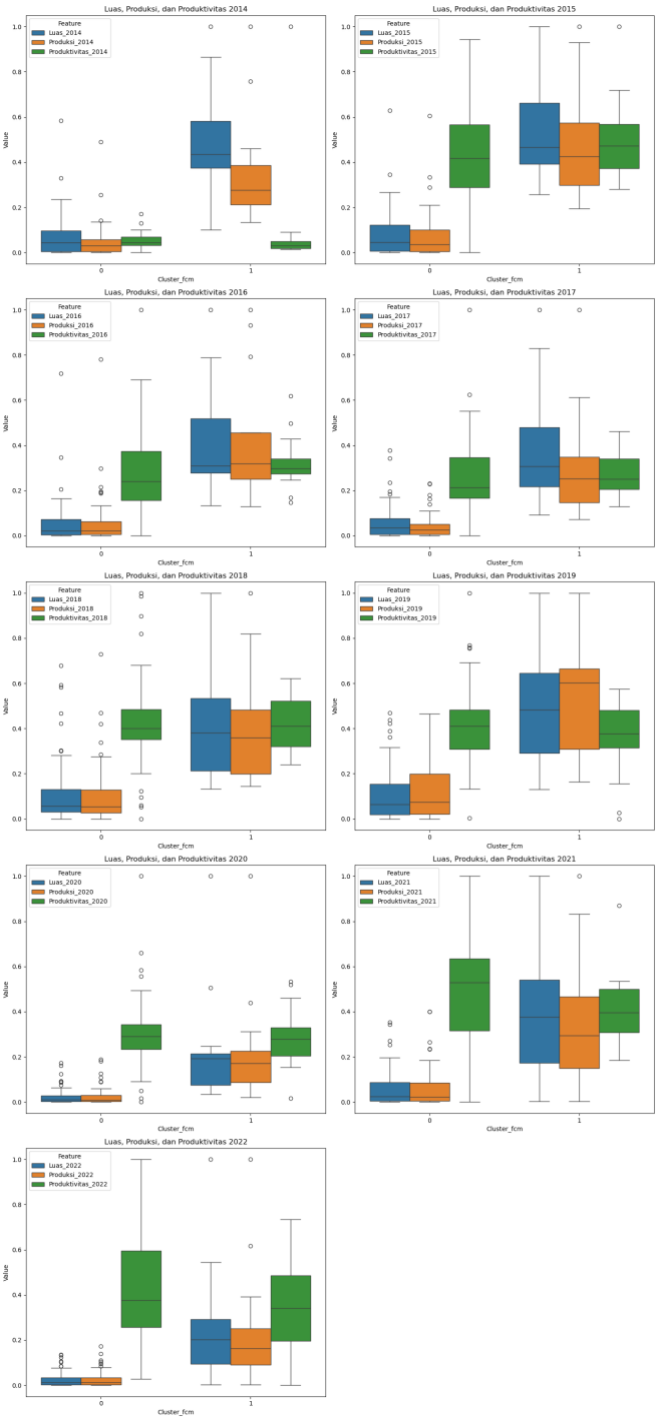


Gambar 8. Visualisasi Hasil Klasterisasi dengan Fuzzy C-Means

Berdasarkan distribusi dari data luas, produksi, dan produktivitas kedelai dari tahun 2010 hingga 2022, hasil klasterisasi menunjukkan adanya pola perubahan distribusi antar klaster setiap periode waktu tertentu. Pada tahun 2010 hingga 2014, terlihat bahwa distribusi nilai luas, produksi, dan produktivitas di klaster 0 lebih rendah dibandingkan klaster 1. Klaster 0 didominasi dengan nilai rata-rata lebih kecil, sedangkan klaster 1 memiliki rentang distribusi yang lebih luas, khususnya dalam hal produktivitas.

Perbedaan ini konsisten hingga tahun 2014. Pada tahun 2015 hingga 2018, terdapat perubahan dalam distribusi data, terutama dalam hal produktivitas. Rata-rata produktivitas di klaster 1 cenderung meningkat dan lebih stabil dibandingkan dengan periode sebelumnya, sedangkan klaster 0 tetap memiliki distribusi yang lebih kecil. Distribusi ini menandakan bahwa kabupaten atau kota dalam klaster 1 cenderung mempertahankan produktivitas yang lebih tinggi dibandingkan dengan klaster 0. Kemudian periode 2019 hingga 2022, distribusi nilai luas, produksi, dan produktivitas menunjukkan pola yang lebih dinamis. Perubahan signifikan terlihat pada distribusi luas lahan di klaster 1, yang memiliki variasi lebih besar dibandingkan klaster 0. Selain itu, produktivitas di klaster 1 tetap lebih tinggi, sementara klaster 0 menunjukkan peningkatan kecil namun konsisten dalam rata-rata produktivitasnya. Secara keseluruhan, hasil klastering dari tahun 2010 hingga 2022 menunjukkan bahwa kabupaten atau kota dalam klaster 1 cenderung memiliki produktivitas lebih tinggi dibandingkan klaster 0 di semua tahun. Berdasarkan hasil distribusi Fuzzy C-Means dengan menggunakan nilai k = 2, cenderung memiliki hasil yang serupa dengan model K-Means.





Gambar 9. Distribusi Data Kedelai dengan Metode Fuzzy C-Means

V. KESIMPULAN

Berdasarkan hasil penelitian klasterisasi data kedelai di Pulau Jawa dengan menggunakan metode K-Means dan Fuzzy C-Means, dilakukan uji coba dengan metode evaluasi *silhouette*. Kedua metode klastering ini menghasilkan hasil yang optimal pada klaster k = 2, kedua metode klasterisasi memiliki hasil *silhouette* yang sama diangka 0.4659. Metode K-Means di

evaluasi menggunakan Elbow Method, yang bertujuan untuk menentukan jumlah klaster optimal dengan melihat perubahan drastis dalam nilai *Within Cluster Sum of Squares* (WCSS). Hasil analisis menunjukkan bahwa titik elbow berada di angka 66, yang menandakan jumlah klaster optimal sebelum penurunan WCSS mulai melambat. Sementara itu, metode Fuzzy C-Means dievaluasi menggunakan *Fuzzy Partition Coefficient* (FPC), yang mengukur data secara unik diklasifikasikan ke dalam klaster. Nilai FPC yang diperoleh adalah 0.9, yang menunjukkan bahwa klaster yang terbentuk memiliki kejelasan dan pemisahan yang baik. Berdasarkan hasil evaluasi yang optimal, kedua metode ini terbukti dapat digunakan untuk mengelompokkan data tanaman pangan kedelai di Pulau Jawa dengan baik. Informasi yang diperoleh dari klasterisasi ini dapat dimanfaatkan dalam memberikan informasi terkait produktivitas dan distribusi kedelai di Pulau Jawa.

UCAPAN TERIMA KASIH

REFERENSI

[1] N. Marlina, I. S. Aminah, N. Amir, and R. Rosmiah, "Aplikasi Jenis Pupuk organik terhadap Kadar Hara NPK dan Produksi Kedelai (Glycine max (L.) Merril) pada Jarak Tanam yang Berbeda di Lahan Pasang Surut," *J. Lahan Suboptimal J. Suboptimal Lands*, vol. 8, no. 2, pp. 148–158, 2019, doi: 10.33230/jlso.8.2.2019.428.

[2] K. Penebel and K. Tabanan, "Keywords : pembinaan, pembuatan POC, limbah pertanian, dan limbah peternakan 28 Jurnal Sewaka Bhakti," vol. 7, pp. 28–37, 2021.

[3] A. R. Ruvananda and M. Taufiq, "Analisis faktor-faktor yang mempengaruhi impor beras di Indonesia," *Kinerja*, vol. 19, no. 2, pp. 195–204, 2022, doi: 10.30872/jkin.v19i2.10924.

[4] I. Khairunisa, "Pengaruh Produksi Kedelai, Harga Kedelai Impor, Dan Nilai Tukar Terhadap Impor Kedelai Indonesia Tahun 2011-2020," *Transekonomika Akuntansi, Bisnis dan Keuang.*, vol. 2, no. 6, pp. 57–70, 2022, doi: 10.55047/transekonomika.v2i6.266.

[5] J. Juswadi, P. Sumarna, and N. S. Mulyati, "Potensi Peningkatan Luas Panen, Produksi, Dan Produktivitas Kedelai Di Jawa Barat," *Paspalum J. Ilm. Pertan.*, vol. 9, no. 1, p. 86, 2021, doi: 10.35138/paspalum.v9i1.281.

[6] M. A. Perdana, I. R. Moeljani, and P. S. Djarwatiningsih, "Pengaruh masa simpan dan suhu simpan terhadap viabilitas dan vigor benih coating kedelai," *J. Agrium*, vol. 20, no. 1, pp. 1–7, 2023.

[7] A. Al Fahrozi, F. Insani, E. Budianita, and I. Afrianty, "Implementasi Algoritma K-Means dalam Menentukan Clustering pada Penilaian Kepuasan Pelanggan di Badan Pelatihan Kesehatan Pekanbaru," *Indones. J. Innov. Multidisipliner Res.*, vol. 1, no. 4, pp. 474–492, 2023, doi: 10.31004/ijim.v1i4.53.

[8] D. A. I. C. Dewi and D. A. K. Pramita, "Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali," *Matrix J. Manaj. Teknol. dan Inform.*, vol. 9, no. 3, pp. 102–109, 2019, doi: 10.31940/matrix.v9i3.1662.

[9] S. Paembonan and H. Abduh, "Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat," *PENA Tek. J. Ilm. Ilmu-Ilmu Tek.*, vol. 6, no. 2, p. 48, 2021, doi: 10.51557/pt_jiit.v6i2.659.

[10] K. P. Sinaga and M. S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.

[11] A. A. Aldino, D. Darwis, A. T. Prastowo, and C. Sujana, "Implementation of K-Means Algorithm for Clustering Corn Planting Feasibility Area in South Lampung Regency," *J. Phys. Conf. Ser.*, vol. 1751, no. 1, 2021, doi: 10.1088/1742-6596/1751/1/012038.

[12] F. Marisa *et al.*, "Digitasi Produktivitas Panen Padi Berbasis K-Means Clustering," *SMARTICS J.*, vol. 7, no. 1, pp. 21–26, 2021.

[13] M. Ula, G. Perdinanta, R. Hidayat, and I. Sahputra, "Analyze the Clustering and Predicting Results of Palm Oil Production in Aceh Utara," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 17, no. 2, pp. 195–206, 2023, doi: 10.22146/ijccs.83195.

[14] T. Yulianto, A. F. Rahmah, F. Faisol, and R. Amalia, "Clustering Daerah Bencana Alam di Indonesia Menggunakan Metode Fuzzy C-Means," *Unisda J. Math. Comput. Sci.*, vol. 9, no. 2, pp. 29–39, 2023, doi: 10.52166/ujmc.v9i2.4776.

[15] A. R. Said, D. Arifianto, and H. A. Al Faruq, "Pengelompokan Kecamatan Di Kabupaten Jember Berdasarkan Tanaman Pangan Dengan Algoritma Fuzzy C-Means Dan Metode Elbow," *J. Smart Teknol.*, vol. 2, no. 1, pp. 1–12, 2020.

[16] R. Yuliana Sari, H. Oktavianto, and H. Wahyu Sulistyo, "Algoritma K-Means Dengan Metode Elbow Untuk Mengelompokkan Kabupaten/Kota Di Jawa Tengah Berdasarkan Komponen Pembentuk Indeks Pembangunan Manusia K-Means Algorithm With Elbow Method To Grouping District/City in Central Java Based on Components of Human D," *J. Smart Teknol.*, vol. 3, no. 2, pp. 2774–1702, 2022, [Online]. Available: http://jurnal.unmuhjember.ac.id/index.php/JST.

[17] N. A. Maori and E. Evanita, "Metode Elbow dalam Optimasi Jumlah Cluster pada K-Means Clustering," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 14, no. 2, pp. 277–288, 2023, doi: 10.24176/simet.v14i2.9630.

[18] N. T. Hartanti, "Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional," *J. Nas. Teknol. dan Sist. Inf.*, vol. 6, no. 2,

pp. 82–89, 2020, doi: 10.25077/teknosi.v6i2.2020.82-89.

[19] A. C. Putra and K. D. Hartomo, "Optimalisasi Penyaluran Bantuan Pemerintah Untuk UMKM Menggunakan Metode Fuzzy C-Means," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 1, no. 10, pp. 474–482, 2021, [Online]. Available: https://jurnal.iaii.or.id/index.php/RESTI/article/view/2980/423.

[20] I. G. Harsemadi, D. P. Agustino, and I. G. B. A. Budaya, "Klasterisasi Pelanggan Tenant Inkubator Bisnis STIKOM Bali Untuk Strategi Manajemen Relasi Dengan Menggunakan Fuzzy C-Means," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 4, no. 4, pp. 232–243, 2023, doi: 10.35746/jtim.v4i4.293.

[21] M. Sholeh and K. Aeni, "Perbandingan Evaluasi Metode Davies Bouldin, Elbow dan Silhouette pada Model Clustering dengan Menggunakan Algoritma K-Means," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 8, no. 1, p. 56, 2023, doi: 10.30998/string.v8i1.16388.

[22] R. F. Utari, F. Insani, S. Agustian, and L. Afriyanti, "Pengelompokan Data Pendistribusian Listrik Menggunakan Algoritma Mean Shift," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 3, pp. 1015–1023, 2024, doi: 10.57152/malcom.v4i3.1428.