PERBANDINGAN METODE *FUZZY C-MEANS* DAN *K-MEDOIDS* DALAM PENGELOMPOKAN KABUPATEN/KOTA DI JAWA TENGAH BERDASARKAN POTENSI KOMODITAS UNGGULAN PERTANIAN

Fachrul Alam[1*], Emeylia Safitri[2]
Program Studi Statistika, Fakultas Sains dan Teknologi, Universitas Terbuka, Indonesia

**Penulis korespondensi: fachrulalam68@gmail.com*

ABSTRAK

Indonesia sebagai negara dengan populasi besar, tingginya konsumsi bahan pokok menjadikan ketahanan pangan nasional sangat vital. Dalam konteks ini, Jawa Tengah memegang peran strategis sebagai salah satu kontributor produksi komoditas unggulan pertanian. Penelitian ini bertujuan membandingkan kinerja antara metode *Fuzzy C-Means* (FCM) dan *K-Medoids* dalam melakukan pengelompokan 35 kabupaten/kota di Provinsi Jawa Tengah berdasarkan variabel produksi padi, buah, sayuran, telur, dan susu ke dalam beberapa kategori mulai dari produksi tinggi hingga rendah. Selain implementasi kedua algoritma tersebut, penelitian ini akan menyajikan pendekatan yang kuat dalam menentukan jumlah *cluster k* optimal dengan menggunakan metode *Elbow* yang kemudian divalidasi oleh *Silhouette Index*. Hasil analisis menunjukkan bahwa FCM memberikan kualitas pengelompokan terbaik, dengan *Silhouette Score* tertinggi (0,3848), dan menghasilkan tiga kelompok wilayah yang merepresentasikan pola produksi yang berbeda secara signifikan. Temuan ini menegaskan adanya spesialisasi komoditas antarwilayah serta menyediakan gambaran yang lebih jelas untuk perencanaan ketahanan pangan daerah.

Kata kunci: *Fuzzy C-Means, K-Medoids, Elbow, Silhouette Index,* Ketahanan Pangan.

1 PENDAHULUAN

Indonesia merupakan salah satu negara dengan jumlah penduduk terbesar di dunia. Berdasarkan hasil Sensus Penduduk tahun 2020 (SP2020) yang dilaksanakan pada bulan September, tercatat jumlah penduduk Indonesia mencapai 270,20 juta jiwa, atau meningkat sekitar 32,56 juta jiwa dibandingkan hasil Sensus Penduduk tahun 2010 (BPS, 2021). Besarnya populasi ini secara langsung berdampak pada tingginya tingkat konsumsi bahan pokok dalam negeri. Kebutuhan pangan yang terus meningkat ini harus dipenuhi, terutama dari hasil sektor pertanian. Indonesia selain dikenal sebagai negara maritim, juga merupakan negara agraria dengan tanah yang subur menjadikannya penghasil utama berbagai komoditas tani dan rempah yang melimpah. Berdasarkan penelitian yang dilakukan oleh Yulianti *et al*. (2023) tingkat konsumsi bahan pokok dalam hal ini beras dipengaruhi oleh jumlah penduduk. Menyadari urgensi ini, swasembada pangan ditetapkan sebagai salah satu program unggulan pemerinta. Program ini bertujuan memastikan pemenuhan kebutuhan pangan dalam negeri tetap terjaga, sehingga ketahanan pangan nasional dapat dipertahankan di tengah ketidakpastian ekonomi global. Dalam konteks ketahanan pangan nasional, peran masing-masing daerah sangatlah vital. Jawa Tengah merupakan salah satu provinsi dengan kontribusi pertanian yang signifikan di Indonesia. Berdasarkan data BPS tahun 2024, Jawa Tengah menempati posisi penting secara nasional, di antaranya sebagai penyumbang terbesar kedua untuk produksi padi dengan kontribusi sebesar (17%), telur (13%), buah dan sayuran (12%), dan terbesar ketiga untuk produksi susu (9%). Oleh karena itu, penting untuk memahami secara mendalam karakteristik produksi di setiap wilayah. Penelitian ini akan berfokus pada pengelompokan kabupaten dan kota di Provinsi Jawa Tengah berdasarkan variabel produksi padi, buah, sayuran, telur, dan susu ke dalam beberapa kategori mulai dari produksi tinggi hingga rendah. Secara umum,

pengelompokan ini dapat dilakukan secara kasat mata dengan membandingkan satu per satu hasil pertanian. Namun, pendekatan manual ini sering kali tidak memberikan hasil yang mendalam dan objektif. Untuk mendapatkan pengelompokan yang efektif dan memetakan segmentasi data secara terstruktur metode *data mining*, seperti *clustering* sangat diandalkan karena kemampuannya dalam mengidentifikasi pola tersembunyi dan menghasilkan kualitas pengelompokan yang terukur (Syam *et al*., 2024).

Untuk mencapai analisis pengelompokan yang lebih terstruktur dan berbasis data dengan menggunakan metode modern, seperti *Fuzzy C-Means* (FCM) dan *K-Medoids*. FCM adalah suatu metode yang sering digunakan dalam logika *fuzzy* untuk pengklasteran data yang ditentukan oleh derajat keanggotaan. FCM memperkenalkan konsep keanggotaan *fuzzy*, memungkinkan suatu objek data menjadi anggota parsial dari beberapa *cluster* secara bersamaan. Sedangkan *K-Medoids* biasa dikenal sebagai *Partitioning Around Medoids* (PAM) adalah algoritma *unsupervised learning* yang dirancang untuk mencari *medoid* (titik data paling sentral) pada setiap *cluster* (Kurmiati *et al*., 2021). Algoritma ini sering dianggap lebih mudah diimplementasikan daripada *K-Means* (Luo, 2022). Pengelompokan data dalam *K-Medoids* dilakukan melalui sistem partisi, di mana jarak antara data non *medoid* dengan data *medoid* dihitung. Untuk menentukan *cluster* terdekat dan mengalokasikan data ke sana, *K-Medoids* biasanya menggunakan perhitungan jarak, seperti *Euclidean Distance* (Sindi *et al*., 2020). Perbandingan kinerja keduanya sangat diperlukan untuk menentukan algoritma yang paling sesuai dan menghasilkan pemetaan potensi yang paling akurat.

Penelitian yang dilakukan oleh Jaya, et al. (2025) dalam studi mereka yang berjudul ”Perbandingan Algoritma *K-Means* dan *Fuzzy C-Means* untuk Klasterisasi Produktivitas Kedelai di Pulau Jawa” mengungkapkan bahwa metode *clustering* terbukti efektif dalam memetakan potensi pertanian di Indonesia. Penelitian tersebut berfokus pada klasterisasi data produktivitas kedelai dari tahun 2010 hingga 2022. Hasil evaluasi *silhouette* dari metode *K-Means* dan *Fuzzy C-Means* menunjukkan hasil yang serupa dan efektif dalam menentukan klasterisasi data, dengan nilai silhouette optimal berada di 0,4659 pada jumlah *cluster* $k = 2$. Temuan ini mendukung pemilihan metode *clustering* sebagai alat yang lebih baik untuk analisis pengelompokan, terutama dalam konteks pertanian dan penentuan potensi wilayah karena dapat memberikan hasil yang lebih akurat dan informatif dalam memahami pola pertumbuhan dan pembagian potensi produk secara objektif.

Berdasarkan latar belakang mengenai urgensi ketahanan pangan, besarnya populasi di Indonesia, dan peran vital Jawa Tengah sebagai sentra produksi pertanian, penelitian ini bertujuan membandingkan kinerja klasterisasi antara metode *Fuzzy C-Means* (FCM) dan *K-Medoids* untuk melakukan analisis pengelompokan kabupaten/kota di Jawa Tengah berdasarkan komoditas unggulan pertanian. Tujuan perbandingan ini adalah untuk memetakan potensi wilayah ke dalam kategori produksi dan menentukan algoritma mana yang paling sesuai untuk data tersebut. Selain implementasi kedua algoritma tersebut, penelitian ini akan menyajikan pendekatan yang kuat dalam menentukan banyak *cluster k* optimal dengan menggunakan metode *Elbow* yang kemudian divalidasi oleh *Silhouette Index*. Analisis ini dilakukan dengan bantuan *software Jupyter Notebook* menggunakan bahasa pemrograman *Python* untuk memproses dan menganalisis data pertanian tahun 2024. Hasilnya diharapkan memberikan kontribusi signifikan bagi pemerintah daerah dan agrobisnis dalam perumusan kebijakan pertanian yang lebih efisien dan tepat sasaran, sehingga mendukung program swasembada pangan.

2 METODE

2.1 Data dan Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari website resmi Badan Pusat Statistik (BPS) Provinsi Jawa Tengah. Pengumpulan data dilakukan melalui metode dokumentasi (studi literatur dan *web scraping*) terhadap publikasi resmi BPS. Data yang dikumpulkan adalah data produksi komoditas pertanian untuk periode tahun 2024.

Data tersebut mencakup variabel kuantitatif berupa produksi padi, buah, sayur, susu, dan telur dengan satuan ton untuk seluruh kabupaten/kota di Provinsi Jawa Tengah. Terdapat batasan data untuk variabel buah, sayur, susu, dan telur merupakan agregasi dari berbagai jenis yang diproduksi di wilayah tersebut. Misal variabel buah merupakan agregasi darin hasil produksi alpukat, jeruk, mangga, dan lain sebagainya. Analisis pengelompokan akan dilakukan berdasarkan data produksi pertanian yang mencakup 35 kabupaten/kota di Provinsi Jawa Tengah. Data produksi ini akan dianalisis untuk memahami pola potensi dan spesialisasi produksi setiap wilayah.

2.2 Analisis *Cluster*

Analisis *cluster* merupakan sebuah metode multivariat yang memiliki tujuan utama untuk mengklasifikasikan n objek menjadi kelompok-kelompok dan memastikan setiap kelompok yang terbentuk memiliki keseragaman internal yang tinggi, sementara perbedaan antar kelompok dimaksimalkan. Pengelompokan ini dilakukan berdasarkan pada kesamaan relatif dari karakteristik objek tersebut yang diukur melalui p variabel (Nuraini *et al*., 2024). Proses analisis *cluster* pada dasarnya melibatkan dua langkah utama (Härdle & Simar, 2019).

a) Penentuan ukuran kedekatan, setiap pasangan objek dievaluasi untuk menentukan tingkat kesamaan nilainya. Ukuran kesamaan ini berfungsi untuk mengukur seberapa dekat objek satu sama lain, semakin dekat nilainya, semakin besar kemungkinan objek tersebut berada dalam *cluster* yang sama.
b) Pemilihan algoritma pengelompokan, menggunakan ukuran kedekatan yang telah ditentukan, algoritma bertugas mengelompokkan objek sehingga kesamaan dalam kelompok (*intra-cluster*) tinggi dan perbedaan antar kelompok (*inter-cluster*) rendah.

Analisis *cluster* secara umum dibagi menjadi dua pendekatan utama (Johnson & Wichern, 2007). Pertama, metode hierarkis yang membentuk struktur pengelompokan secara bertingkat atau berjenjang, sering divisualisasikan menggunakan dendrogram. Metode ini memiliki dua jenis dasar, *agglomerative* (menggabungkan objek dari bawah ke atas) dan *divisive* (memisahkan data dari atas ke bawah). Kedua, metode nonhierarkis yang bertujuan langsung membagi data menjadi sejumlah *cluster* yang jumlahnya sudah ditetapkan di awal, tanpa membangun struktur bertingkat. Salah satu contoh populernya adalah *Fuzzy C-Means* (FCM) *Clustering*, sebuah variasi dari *K-Means*, yang memperkenalkan konsep keanggotaan fuzzy, memungkinkan suatu objek memiliki derajat keanggotaan terhadap lebih dari satu *cluster*. Selanjutnya, ada *K-Medoids* yang serupa dengan *K-Means* tetapi lebih kuat (*robust*) terhadap *outlier*. Ini karena *K-Medoids* menggunakan *medoid*, yaitu titik data aktual yang paling sentral, sebagai pusat klaster alih-alih menggunakan *mean* (rata-rata geometris) yang dapat sangat dipengaruhi oleh data pencilan.

2.3 Standardisasi Data

Standardisasi data merupakan tahapan esensial dalam analisis multivariat karena fungsinya mengubah data yang awalnya memiliki skala beragam menjadi format yang seragam (Han *et al*., 2012). Langkah ini sangat krusial karena jika terdapat perbedaan skala yang signifikan, variabel dengan nilai skala yang lebih besar cenderung akan mendominasi proses analisis yang pada akhirnya dapat menimbulkan bias dalam penentuan tingkat kesamaan antar sampel dan memengaruhi keakuratan hasil analisis. Penelitian ini menggunakan transformasi

Z-score untuk standardisasi data. Metode ini berfungsi mengukur jarak suatu nilai data dari rata-rata di dalam satuan deviasi standar (Grekousis, 2020). Hasil dari transformasi *Z-score* adalah data baru dengan rata-rata nol dan deviasi standar satu, yang secara efektif menunjukkan posisi relatif setiap nilai terhadap keseluruhan distribusi variabel tersebut (Ou *et al*., 2023). *Z-score* dapat dihitung dengan rumus berikut.

$$zscore = \frac{x_i - \bar{x}}{s} \quad (1)$$

dengan x_i merupakan nilai objek ke-i, $\bar{x}$ merupakan nilai rata-rata sampel, dan s merupakan standar deviasi.

2.4 Uji Korelasi

Menurut Sugiyono (2017) korelasi adalah koefisien yang menunjukkan kekuatan dan arah hubungan antara dua variabel atau lebih, serta digunakan untuk menguji hipotesis dalam penelitian asosiatif. Uji korelasi digunakan dalam analisis *cluster* untuk menentukan tingkat hubungan antar variabel yang dipakai (Everitt *et al*., 2011). Sangat penting untuk mengukur korelasi ini karena korelasi yang terlalu tinggi antar variabel dapat mengacaukan hasil pengelompokan. Hal ini terjadi karena variabel-variabel yang berkorelasi kuat dapat secara tidak wajar mendominasi proses pembentukan klaster. Nilai korelasi yang mendekati 1 menunjukkan hubungan yang kuat, sementara korelasi negatif menandakan hubungan berlawanan.

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{\{(n\sum x^2)(\sum x^2)\}\{(n\sum y^2)(\sum y^2)\}}} \quad (2)$$

Di mana,

r : koefisien korelasi
n : ukuran sampel
x : variabel pertama
y : variabel kedua.

2.5 Menentukan Banyak *Cluster*

Menentukan jumlah *cluster* optimal merupakan tahap krusial dalam analisis *cluster*, meskipun prosesnya tidak memiliki panduan tunggal yang baku. Penentuan ini wajib mempertimbangkan aspek kegunaan praktis, validitas statistik, dan dasar teoretis yang relevan (Ningsih *et al*., 2016). Untuk mengatasi ketidakpastian ini, penelitian saat ini akan menggunakan dua metode berbeda guna menemukan pendekatan pengelompokan terbaik yang hasilnya kemudian akan divalidasi. Terdapat beragam metode yang dapat diterapkan untuk menentukan banyak *cluster* yang optimal secara efektif. Metode-metode tersebut mencakup *Silhouette, Fuzzy Partition Coefficient* (FPC), *Elbow Method*, *Xie-Beni Index, Dunn Index,* dan *Gap Statistic*. Dalam penelitian ini akan menggunakan dua metode utama, yaitu menentukan banyak *cluster* dengan *Elbow Method* yang kemudian divalidasi oleh *Silhouette Index.*

a) *Elbow Method*

Dalam analisis *cluster* nonhierarkis, banyak *cluster* optimal *k* harus ditentukan terlebih dahulu, dalam hal ini menggunakan metode *Elbow*. Metode ini bekerja dengan memplot nilai *Sum of Squared Error* (SSE) terhadap berbagai banyak *cluster*, kemudian memilih titik ”tekukan” yang disebut titik *Elbow* sebagai banyak *cluster* terbaik. Titik *Elbow* ini dijadikan dasar karena menghasilkan pengelompokan yang paling efisien dan bermakna (Maori *et al*., 2023). Rumus SSE dapat dituliskan sebagai berikut.

$$SSE = \sum_{i=1}^{k} \sum_{x \epsilon c_i} \left| |x - \mu i| \right|^2 \quad (3)$$

Dengan k Adalah banyak *cluster*, c_i menyatakan *cluster* ke-*i*, x adalah titik data dalam *cluster*, μi adalah *centroid* untuk *cluster* ke-*i*.

b) *Silhouette Score*

Menurut Izzadin (2019) *Silhouette Score* digunakan untuk mengevaluasi kualitas hasil *clustering*. Indeks ini berfungsi untuk memperkirakan seberapa baik suatu objek penelitian terintegrasi dalam klaster tempat ia berada, atau dengan mengukur nilai rata-rata jarak antar *cluster*. Sebuah *cluster* dinilai semakin baik dan optimal apabila nilai *Silhouette Index* nya semakin mendekati 1. Untuk mengevaluasi hasil pengelompokan secara keseluruhan, dihitung rata-rata nilai *Silhouette* untuk semua *cluster* yang didefinisikan sebagai $sil(k)$ untuk banyak cluster k.

$$sil(c) = sil(k) \frac{1}{|k|} \sum_{i=1}^{k} sil(c_1) \quad (4)$$

Di mana $sil(k)$ merupakan nilai *Silhouette* semua *cluster*, $|k|$ merupakan banyaknya *cluster* k, serta $sil(c_i)$ merupakan rata-rata nilai *Silhouette*.

2.6 *Fuzzy C-Means Cluster*

Menurut Bezdek (1981) FCM adalah teknik *clustering* nonhierarkis yang menerapkan konsep logika *fuzzy*. Berbeda dari *clustering* konvensional, FCM memungkinkan setiap titik data memiliki derajat keanggotaan yang berbeda pada beberapa *cluster* secara simultan. Algoritma ini bekerja secara iteratif, prosesnya dimulai dengan penentuan pusat *cluster* (*centroid*) awal. Melalui iterasi, pusat *cluster* dan skor keanggotaan setiap titik data terus diperbarui hingga pusat *cluster* mencapai lokasi optimal. Tujuannya adalah menemukan pusat *cluster* yang paling akurat, yang dapat digunakan untuk membangun sistem inferensi *fuzzy* (Rahakbauw *et al*., 2017). Tahapan algoritma *Fuzzy C-Means* dapat dijelaskan sebagai berikut (Sanusi et al., 2019).

a) Input data dalam *cluster* berupa matriks berukuran $n \times m$, di mana n merupakan jumlah sampel data dan m merupakan banyak variabel atau karakteristik yang digunakan untuk mengukur setiap sampel data tersebut.
b) Menentukan beberapa parameter diantaranya banyak *cluster* (c), pangkat pembobotan (m), maksimum iterasi $(maxIter)$, *error* terkecil yang diharapkan (ε), fungsi objektif awal (p_0), dan iterasi awal $(t = 1)$. Parameter jumlah *cluster* (c) harus diisi dengan nilai yang lebih besar dari 1 dan lebih kecil dari jumlah data sampel $(1 < c < n)$. Parameter derajat pembobot (w) harus diisi dengan nilai lebih dari 1 $(w > 1)$. Syarat untuk parameter iterasi maksimum $(maxIter)$ adalah nilai lebih besar dari 1 $(maxIter > 1)$ dan parameter *error* $(\varepsilon > 0)$.
c) Menghitung jumlah setiap baris sesuai dengan Persamaan (5).

$$Q_i = \sum_{k=1}^{c} \mu_{ik} = 1 \quad (5)$$

Persamaan (5) menyakan bahwa Q_i adalah jumlah dari setiap kolom, di mana setiap kolom merupakan matriks acak dengan nilai 1. Bentuk matriks partisi awal, U, ditunjukkan oleh Persamaan (6).

$$U = \begin{bmatrix} \mu_{11}(x_1) & \mu_{12}(x_1) & \cdots & \mu_{1c}(x_1) \\ \mu_{21}(x_2) & \mu_{22}(x_2) & \cdots & \mu_{2c}(x_2) \\ \vdots & \vdots & \vdots & \vdots \\ \mu_{n1}(x_n) & \mu_{n2}(x_n) & \cdots & \mu_{nc}(x_n) \end{bmatrix} \quad (6)$$

Persamaan (6) menjelaskan bahwa matriks awal terbentuk dari setiap data yang akan digunakan dalam perhitungan. Jumlah *cluster* yang akan dibentuk digambarkan oleh $\mu_{11}(x_1)$ hingga $\mu_{1c}(x_1)$, sedangkan jumlah data yang akan dikelompokkan digambarkan oleh $\mu_{11}(x_1)$ hingga $\mu_{n1}(x_n)$.

d) Menghitung matriks V sebagai pusat *cluster* untuk setiap *cluster* sesuai dengan Persamaan (7).

$$V_{kj} = \frac{\sum_{i=1}^{n}\left((\mu_{ik})^w \times x_{ij}\right)}{\sum_{i=1}^{n}(\mu_{ik})^w} \quad (7)$$

V adalah pusat *cluster* yang ditentukan dengan menghitung rata-rata tertimbang dari data dalam klaster tersebut. Secara spesifik, pusat *cluster* diperoleh dengan menjumlahkan hasil perkalian antara bobot setiap data dengan pangkat keanggotaan *cluster* data tersebut, lalu hasil penjumlahan ini dibagi dengan total dari pangkat keanggotaan klaster yang telah dihitung.

e) Memperbarui derajat keanggotaan setiap data pada masing-masing *cluster* yang dijelaskan oleh persamaan (8).

$$\mu_{ik} = \frac{\left(\sum_{j=1}^{m}\left(X_{ij} - V_{kj}\right)^2\right)^{\frac{-1}{w-1}}}{\sum_{k=1}^{c}\left(\sum_{j=1}^{m}\left(X_{ij} - V_{kj}\right)^2\right)^{\frac{-1}{w-1}}} \quad (8)$$

μ_{ik} adalah hasil perhitungan untuk menentukan derajat keanggotaan. Nilai μ_{ik} diperoleh dengan menghitung setiap hasil perkalian antara nilai bobot dan pusat *cluster* menggunakan sistem perkalian matriks, kemudian dipangkatkan dengan $\frac{-1}{w-1}$. Seluruh nilai yang dihasilkan kemudian dibagi dengan jumlah baris setiap *cluster*.

f) Menghitung fungsi objektif pada iterasi ke-*t*, P_t ditunjukkan oleh Persamaan (9).

$$P_t = \sum_{i=1}^{n}\sum_{k=1}^{c}\left(\left[\sum_{j=1}^{m}\left(X_{ik} - V_{kj}\right)^2\right](\mu_{ik})^w\right) \quad (9)$$

Secara sederhana, pada persamaan (2.9) menjelaskan perhitungan fungsi objektif, di mana P_t adalah total dari perhitungan untuk masing-masing *cluster*.

g) Memeriksa kriteria pemberhentian, yaitu jika $[P_t - P_{t-1}] < \varepsilon$ atau $t > maxIter$ maka iterasi dihentikan. Jika syarat tersebut tidak terpenuhi, maka iterasi dilanjutkan dengan menambah nilai t menjadi $t = t + 1$ dan mengulangi langkah $ke - 4$.

2.7 *K-Medoids*

K-Medoids juga dikenal sebagai *Partitioning Around Medoids* (PAM) adalah algoritma *unsupervised learning* yang dirancang untuk mencari *medoid* (titik data paling sentral) pada setiap *cluster* (Kurmiati *et al*., 2021). Algoritma ini sering dianggap lebih mudah diimplementasikan daripada *K-Means* (Luo, 2022). Pengelompokan data dalam *K-Medoids* dilakukan melalui sistem partisi, di mana jarak antara data non *medoid* dengan data *medoid*

dihitung. Untuk menentukan *cluster* terdekat dan mengalokasikan data ke sana, *K-Medoids* biasanya menggunakan perhitungan jarak, seperti *Euclidean Distance* (Sindi *et al*., 2020).

Langkah – langkah algoritma *K-Medoids* sebagai berikut (Rinaldi *et al*, 2019).

a) Inisiasi Data & Pusat *cluster*, siapkan data sampel dan tentukan banyak *cluster* k yang dibutuhkan.
b) Penentuan *Medoid* awal, tentukan *k medoid* awal secara acak dari data (atau tentukan nilai *centroid* awal secara acak).
c) Hitung jarak (*cost*) antara setiap objek data X dan *medoid* Y menggunakan persamaan *Euclidean Distance* pada persamaan (10). Setelah itu, setiap objek dialokasikan ke *cluster* yang memiliki *medoid* terdekat (jarak terpendek).

$$d(x,y)\sqrt{\sum_{i=1}^{n}\{X_i - Y_j\}^2} \qquad (10)$$

di mana $d(x,y)$ merupakan antara objek dengan objek *i* dan objek *j*, X_i merupakan nilai atribut pertama dari objek ke *i*, Y_j adalah nilai atribut pertama dari objek ke *j*, *n* jumlah atribut yang dipakai. Memilih secara acak objek data setiap masing-masing *cluster* sebagai kandidat *medoid* baru.
d) Pilih objek data secara acak pada setiap *cluster* sebagai kandidat medoid baru. Hitung total biaya (jarak) yang dihasilkan dari pengelompokan baru ini. Total biaya ini ditentukan dengan mengambil nilai jarak paling kecil pada setiap *cluster*.
e) Hitung perubahan biaya (selisih antara biaya lama dan biaya baru). Jika nilai perubahan ini kurang dari 0 (artinya biaya pengelompokan berkurang), maka *medoid* baru tersebut diterima. Langkah 2 hingga 4 diulangi hingga tidak terjadi perubahan *medoid* yang dihasilkan yang menunjukkan bahwa *cluster* beserta anggotanya telah optimal.

2.8 Pengolahan dan Analisis

Analisis data dalam penelitian ini merupakan perbandingan dua metode analisis pengelompokan antara metode *Fuzzy C-Means* dan *K-Medoids* yang kemudian diolah dengan *Python*. Adapun langkah-langkah analisis yang digunakan dalam penelitian ini adalah sebagai berikut.

a) Pengumpulan data produksi komoditas unggulan pertanian dari *website* BPS Jawa Tengah tahun 2024.
b) Melakukan standardisasi data.
c) Visualisasi data dan analisis korelasi.
d) Menentukan banyak *cluster* dan validasi.
e) Pengelompokan menggunakan perbandingan metode *Fuzzy C-Mean* dan *K-Medoids*.
f) Mendapatkan hasil perbandingan pengelompokan
g) Visualisasi hasil perbandingan pengelompokan.
h) Interpretasi hasil perbandingan pengelompokan.

3 HASIL DAN PEMBAHASAN

3.1 Analisis Deskriptif

Berdasarkan data komoditas pertanian di berbagai kabupaten dan kota di Jawa Tengah, terlihat jelas adanya spesialisasi geografis dalam distribusi produksi. Produksi Padi menunjukkan dominasi di Kabupaten Grobogan dengan volume tertinggi (735.187,08), sementara wilayah perkotaan seperti Kota Surakarta mencatatkan volume terendah (163,78). Untuk komoditas Buah, produksi terbesar secara signifikan terkonsentrasi di Kabupaten Banjarnegara (968.333,33). Disparitas paling signifikan teramati pada komoditas Sayuran, di mana Kabupaten Brebes muncul sebagai sentra produksi utama dengan volume yang jauh melampaui wilayah lain (571.459,20), yang kontras dengan produksi minimal di Kota

Magelang (22,95) dan Kota Surakarta (14,84). Sementara itu, produksi Susu didominasi oleh Kabupaten Boyolali (38.800,00), dan komoditas Telur menunjukkan konsentrasi tertinggi di Kabupaten Kendal (234.619,85). Secara umum, pola produksi menunjukkan bahwa output komoditas pertanian dan peternakan berskala besar sebagian besar terkonsentrasi di wilayah kabupaten, sedangkan wilayah kota secara konsisten mencatatkan nilai produksi terendah di hampir semua kategori, mengindikasikan pergeseran fokus ekonomi dan keterbatasan lahan untuk sektor agrikultur intensif.

Tabel 1. Data Komoditas Pertanian Jawa Tengah Tahun 2024

Kabupaten dan Kota	Padi	Buah	Sayuran	Susu	Telur
Cilacap	651.135,71	155.300,78	17.849,62	168,56	13.241,66
Banjarnegara	103.082,48	968.333,32	288.320,29	78,48	17.773,66
Boyolali	279.953,51	122.551,99	70.544,98	38.800,00	61.061,74
Grobogan	735.187,08	154.514,08	68.338,60	165,71	15.640,79
Kendal	175.081,41	73.955,89	51.420,69	12,40	234.619,85
…	…	…	…	…	…
Brebes	419.117,47	47.923,41	571.459,20	2,69	40.952,72
Kota Magelang	528,77	1.090,66	22,95	216,39	84,41
Kota Surakarta	163,78	422,98	14,84	10,00	16,35
Kota Salatiga	4.012,62	4.831,60	28,50	2.429,00	2.478,68
Kota Semarang	14.209,65	19.244,00	136,50	428,55	18.657,26

Ketiga komoditas agrikultur, yaitu Padi, Buah, dan Sayuran, merupakan pilar esensial dalam mendukung swasembada pangan regional. Padi jelas memimpin volume produksi keseluruhan, menjadikannya tulang punggung utama suplai makanan pokok. Kontribusi Sayuran sangat krusial, namun ditandai oleh konsentrasi produksi yang ekstrem pada sejumlah kecil wilayah yang sangat spesialis, suatu kondisi yang harus dicermati dari aspek kerentanan pasokan. Sementara itu, Buah melengkapi kebutuhan diversifikasi gizi. Secara kolektif, tingginya volume produksi dari ketiga sektor ini di berbagai wilayah kabupaten merupakan kunci utama untuk menjamin ketahanan pangan Jawa Tengah, dengan wilayah perkotaan memberikan kontribusi yang sangat minimal.

3.2 Standardisasi Data

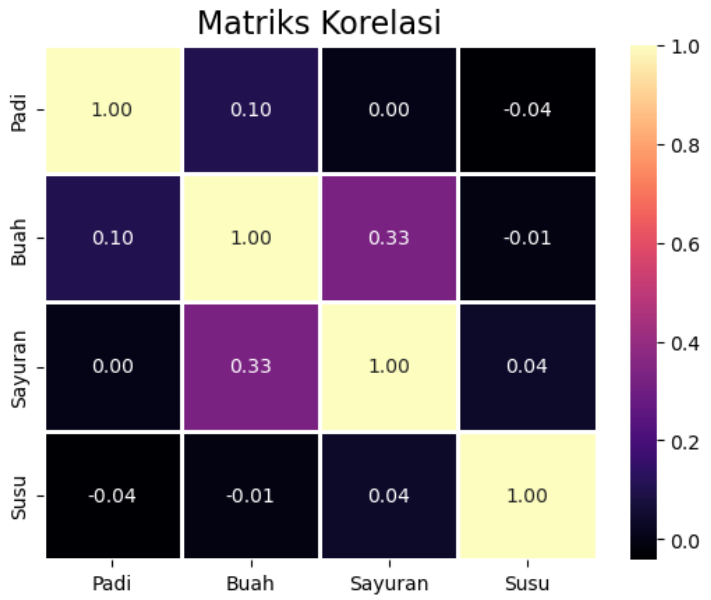
Standardisasi data dilakukan untuk menyelaraskan skala pengukuran antar variabel yang berbeda dalam kumpulan data, dengan mengaplikasikan metode *Z-score* melalui fungsi *StandardScaler* dari bahasa pemrograman *Python*, menghasilkan data terstandardisasi sebagaimana disajikan pada Tabel 2.

Tabel 2. Standardisasi Data

Kabupaten dan Kota	Padi	Buah	Sayuran	Susu	Telur
Cilacap	1,9361	0,2115	-0,4291	-0,2749	-0,2364
Klaten	0,2479	-0,3613	-0,5317	-0,0157	-0,1251
Kendal	-0,3850	-0,2919	-0,1393	-0,2963	4,2909
Brebes	0,8049	-0,4529	4,3511	-0,2977	0,3304
…	…	…	…	…	…
Kota Magelang	-1,2360	-0,7427	-0,5831	-0,2684	-0,5057
Kota Surakarta	-1,2378	-0,7468	-0,5831	-0,2967	-0,5070
Kota Salatiga	-1,2191	-0,7195	-0,5830	0,0349	-0,4570
Kota Semarang	-1,1693	-0,6304	-0,5821	-0,2393	0,1258

3.3 Analisis Korelasi

Analisis korelasi berfungsi untuk mengkuantifikasi dan mendeskripsikan derajat keterkaitan linear antar variabel dalam suatu set data. Tujuan utamanya adalah untuk mendeteksi pola kovariasi data dan menguji hipotesis mengenai interaksi antar variabel. Hasil analisis yang berupa koefisien, secara eksplisit menunjukkan arah hubungan (positif, negatif, atau nol) dan kekuatan asosiasi tersebut. Analisis korelasi sangat penting untuk mengukur korelasi ini karena korelasi yang terlalu tinggi antar variabel dapat mengacaukan hasil pengelompokan.



Gambar 1. Hasil Analisis Korelasi

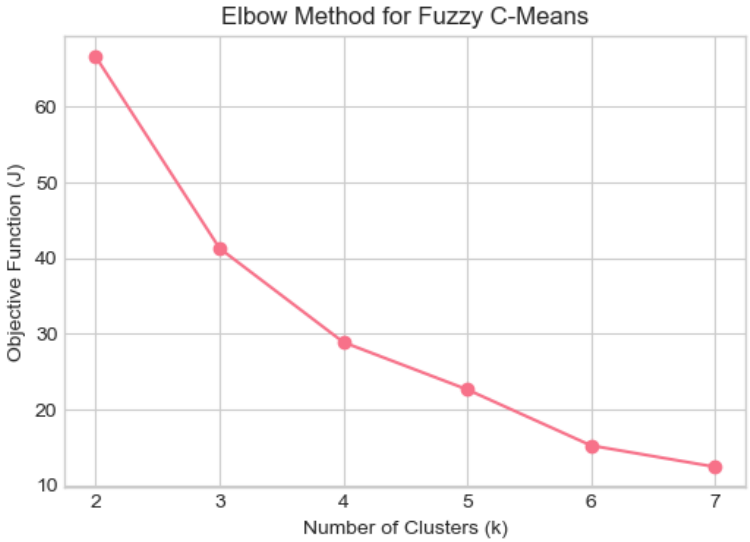
Berdasarkan hasil yang disajikan pada Gambar 1, ditemukan bahwa mayoritas variabel komoditas memiliki keterkaitan linear yang sangat lemah. Korelasi positif terkuat (0,33) teridentifikasi antara Buah dan Sayuran, yang mengindikasikan adanya kovariasi moderat, yaitu kecenderungan wilayah produsen kedua komoditas ini memiliki karakteristik agrikultur serupa. Sebaliknya, komoditas Padi menunjukkan asosiasi yang nyaris independen (koefisien 0,00 hingga 0,10) terhadap variabel lain. Secara keseluruhan, rendahnya koefisien korelasi ini menjadi indikator kuat bahwa data memiliki tingkat multikolinearitas yang minimal, sehingga sangat sesuai untuk dianalisis lebih lanjut menggunakan metode pengelompokan (*clustering*).

3.4 Menentukan Banyak *Cluster*

Dalam menentukan banyak *cluster* optimal metode *Elbow* digunakan untuk mengidentifikasi titik belok optimal yang mengindikasikan banyaknya *cluster* yang paling efisien, didasarkan pada minimasi inersia. Sebagai langkah validasi, *Silhouette Score* digunakan untuk mengevaluasi kualitas pengelompokan dengan mengukur seberapa baik setiap titik data sesuai dengan *cluster* nya dibandingkan dengan *cluster* lain, sehingga memberikan gambaran akurat mengenai kepadatan dan pemisahan antar *cluster*.

a) Metode *Elbow* untuk FCM

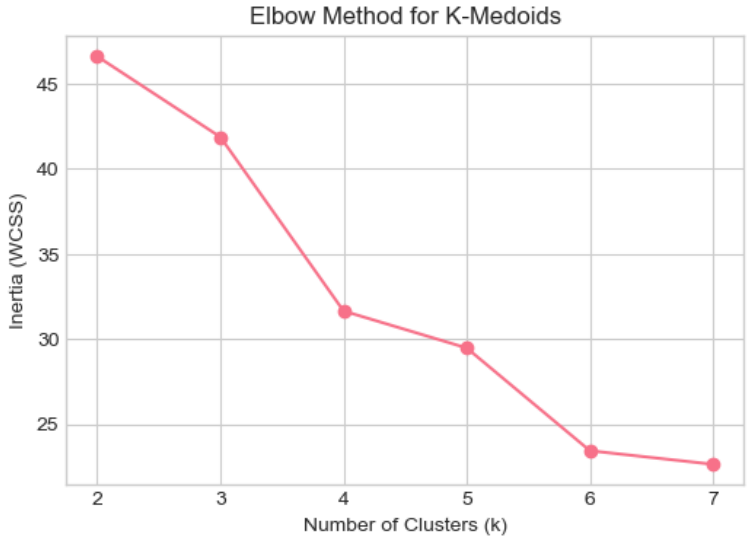
Berdasarkan metode *Elbow* pada Gambar 2 algoritma *Fuzzy C-Means*, titik siku terlihat pada $k = 3$. Hal ini menunjukkan bahwa jumlah cluster optimal untuk mengelompokkan kabupaten/kota berdasarkan indikator produksi komoditas unggulan pertanian adalah tiga *cluster*. Setelah $k = 3$, penurunan fungsi objektif tidak lagi signifikan, sehingga k = 3 memberikan keseimbangan terbaik antara akurasi dan kesederhanaan model.



Gambar 2. *Elbow Method* untuk FCM

b) Metode *Elbow* untuk *K-Medoids*

Berbeda dengan hasil *Fuzzy C-Means* berdasarkan hasil metode *Elbow* pada algoritma *K-Medoids pada Gambar 3*, terlihat bahwa titik siku berada pada $k = 4$. Hal ini menunjukkan bahwa jumlah *cluster* optimal adalah empat. Penurunan nilai setelah $k = 4$ tidak signifikan, sehingga pembagian data ke dalam empat kelompok sudah cukup efisien dalam menggambarkan variasi karakteristik produksi pangan antar kabupaten/kota.



Gambar 3. *Elbow Method* untuk *K-Medoids*

c) Evaluasi dengan *Silhouette Score*

Berdasarkan hasil metode *Elbow* dan evaluasi dengan koefisien *Silhouette* yang terlihat pada Tabel 3, diperoleh bahwa jumlah *cluster* optimal untuk *Fuzzy C-Means* adalah tiga ($k = 3$) dan untuk *K-Medoids* adalah empat ($k = 4$). Hasil ini menunjukkan bahwa kedua metode evaluasi memberikan hasil yang sejalan, di mana FCM dengan tiga *cluster* dan *K-Medoids* dengan empat *cluster* menghasilkan pengelompokan yang paling efisien serta mampu merepresentasikan struktur alami data dengan baik.

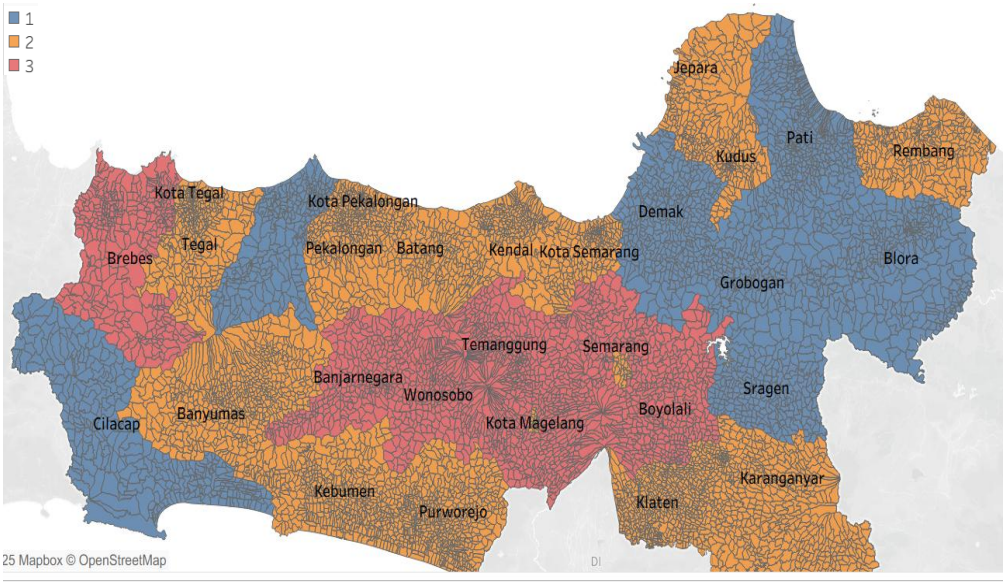
Tabel 3. *Average Silhouette Score*

Banyak *k*	FCM	*K-Medoids*
2	0,3649	0,1836
3	**0,3848**	0,1415
4	0,2898	**0,3109**
5	0,2885	0,2238
6	0,3608	0,1957
7	0,3664	0,1876

3.5 Perbandingan Analisis *Cluster*

a) Analisis *Cluster* dengan FCM

Algoritma ini bekerja secara iteratif, prosesnya dimulai dengan penentuan pusat *cluster* (*centroid*) awal. Melalui iterasi, pusat *cluster* dan skor keanggotaan setiap titik data terus diperbarui hingga pusat *cluster* mencapai lokasi optimal. Tujuannya adalah menemukan pusat *cluster* yang paling akurat, yang dapat digunakan untuk membangun sistem inferensi *fuzzy* (Rahakbauw *et al*., 2017). Pengelompokan produksi pertanian di Jawa Tengah dilakukan pada variabel, yaitu Padi, Buah, Sayuran, Susu, dan Telur. Berikut hasil visualisasi pengelompokan kabupaten/kota di Jawa Tengah berdasarkan data produksi komoditas unggulan pertanian tahun 2024.



Gambar 4. Hasil *Cluster* untuk FCM

Berdasarkan Gambar 4, hasil pengelompokan menggunakan metode *Fuzzy C-Means* (FCM) dengan jumlah cluster optimal sebanyak tiga *cluster* (k = 3) menunjukkan adanya perbedaan karakteristik antar kelompok kabupaten/kota di Jawa Tengah. Pada *Cluster* 1 terdapat tujuh kabupaten, yaitu Cilacap, Sragen, Grobogan, Blora, Pati, Demak, dan Pemalang. Selanjutnya, *Cluster* 2 terdiri atas dua puluh satu kabupaten/kota, yaitu Banyumas, Purbalingga, Kebumen, Purworejo, Klaten, Sukoharjo, Wonogiri, Karanganyar, Rembang, Kudus, Jepara, Kendal, Batang, Pekalongan, Tegal, Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Semarang, Kota Pekalongan, dan Kota Tegal. Sementara itu, *Cluster* 3 terdiri atas tujuh kabupaten, yaitu Banjarnegara, Wonosobo, Magelang, Boyolali, Semarang, Temanggung, dan Brebes. Dengan demikian, hasil pengelompokan ini menunjukkan adanya variasi distribusi

wilayah pada tiap *cluster* yang mencerminkan perbedaan tingkat produksi komoditas pangan antar daerah.

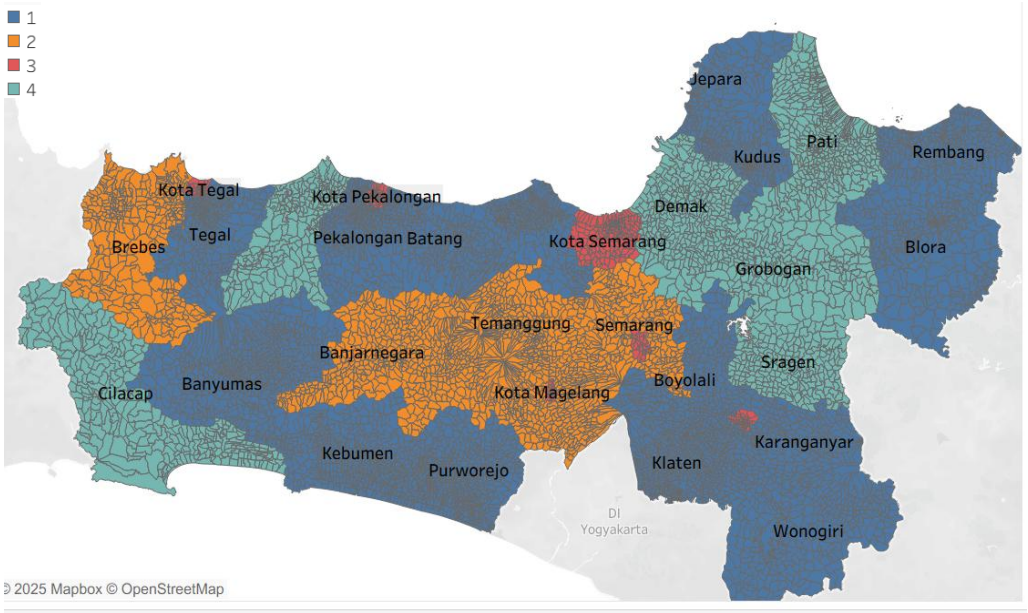
Tabel 4. Rata – Rata 3 *Cluster* FCM

Variabel	*Cluster* 1	*Cluster* 2	*Cluster* 3
Padi	583.872,20	170.724,79	174.138,72
Buah	169.283,77	67.865,88	232.742,12
Sayuran	35.944,30	17.238,86	250.077,82
Susu	143,95	536,13	9.120,98
Telur	11.430,18	27.621,77	34.575,66

Hasil pengelompokan *Fuzzy C-Means* (FCM) terlihat pada Tabel 4 mengidentifikasi tiga kelompok wilayah dengan karakteristik produksi yang berbeda secara signifikan. *Cluster* 1 dicirikan sebagai sentra utama produksi Padi, dengan rata-rata Padi mencapai nilai tertinggi di antara semua *cluster*, menunjukkan spesialisasi wilayah ini pada komoditas pangan pokok. Berbeda dengan itu, *Cluster* 3 mewakili wilayah dengan produksi yang sangat terdiversifikasi, ditandai oleh rata-rata tertinggi pada empat komoditas lain (Buah, Sayuran, Susu, dan Telur) mengindikasikan bahwa wilayah ini unggul dalam sektor hortikultura dan peternakan. Sementara itu, *Cluster* 2 merepresentasikan wilayah dengan output pertanian yang paling rendah secara umum di sebagian besar komoditas, mencerminkan adanya keterbatasan sumber daya lahan atau pergeseran fokus ekonomi di wilayah tersebut.

b) Analisis *Cluster* dengan *K-Medoids*

Pengelompokan data dalam *K-Medoids* dilakukan melalui sistem partisi, di mana jarak antara data non *medoid* dengan data *medoid* dihitung. Untuk menentukan *cluster* terdekat dan mengalokasikan data ke sana, *K-Medoids* biasanya menggunakan perhitungan jarak, seperti *Euclidean Distance* (Sindi *et al*., 2020). Berikut hasil visualisasi pengelompokan kabupaten/kota di Jawa Tengah berdasarkan data produksi komoditas unggulan pertanian tahun 2024.



Gambar 6. Hasil *Cluster* untuk *K-Medoids*

Berdasarkan Gambar 6 *Cluster* 1 merupakan kelompok terbesar, terdapat tujuh belas daerah, yaitu Banyumas, Purbalingga, Kebumen, Purworejo, Boyolali, Klaten, Sukoharjo, Wonogiri, Karanganyar, Blora, Rembang, Kudus, Jepara, Kendal, Batang, Pekalongan, dan Tegal. *Cluster* 2 terdiri atas enam kabupaten yang mayoritas merupakan wilayah dataran tinggi Banjarnegara, Wonosobo, Magelang, Semarang, Temanggung, dan Brebes. *Cluster 3* secara konsisten mengelompokkan enam wilayah perkotaan Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Semarang, Kota Pekalongan, dan Kota Tegal. Sementara itu, *Cluster 4* terdiri atas enam kabupaten, yang mencakup wilayah produksi padi strategis dan pesisir, yaitu Cilacap, Sragen, Grobogan, Pati, Demak, dan Pemalang. Dengan demikian, hasil pengelompokan ini menunjukkan adanya variasi distribusi wilayah yang mencerminkan diferensiasi spesialisasi dan tingkat produksi antar daerah, dengan *Cluster 3* secara jelas memisahkan wilayah metropolitan.

Tabel 5. Rata – Rata 4 *Cluster K-Medoids*

Variabel	***Cluster* 1**	***Cluster* 2**	***Cluster* 3**	***Cluster* 4**
Padi	250.489,62	156.502,92	5.210,22	610.0449,10
Buah	92.513,62	251.107,14	4.861,69	188.470,05
Sayuran	25.951,77	279.999,96	451,99	40.046,50
Susu	2.752,31	4.174,47	582,94	129,92
Telur	36.879,01	30.161,31	3.734,52	11.963,30

Hasil pengelompokan *K-Medoids* terlihat pada Gambar 7 dengan $k = 4$ mengidentifikasi empat tipologi wilayah yang berbeda secara signifikan. *Cluster* 4 dicirikan sebagai sentra utama produksi Padi, dengan rata-rata Padi tertinggi yang menunjukkan spesialisasi pada pangan pokok. *Cluster* 2 mewakili wilayah dengan diversifikasi dan intensitas tinggi pada sektor hortikultura, ditandai oleh rata-rata tertinggi pada Buah dan Sayuran. *Cluster* 1 menunjukkan spesialisasi pada sektor peternakan unggas dengan rata-rata Telur tertinggi di samping produksi pertanian menengah. Sementara itu, *Cluster* 3 merepresentasikan wilayah dengan output pertanian yang paling rendah secara umum di hampir semua komoditas, mencerminkan keterbatasan sumber daya lahan atau sifat metropolitan dari wilayah anggotanya.

4 KESIMPULAN

Analisis pengelompokan dilakukan terhadap data produksi lima komoditas unggulan pertanian dari 35 kabupaten dan kota di Jawa Tengah, yang menunjukkan adanya spesialisasi geografis yang kuat ditandai oleh kontribusi tinggi Padi di Grobogan, Buah di Banjarnegara, Sayuran di Brebes, Susu di Boyolali, dan Telur di Kendal sementara wilayah perkotaan mencatat ouput rendah. Pemeriksaan korelasi lebih lanjut mengkonfirmasi bahwa mayoritas variabel komoditas memiliki keterkaitan linear yang sangat lemah mengindikasikan tingkat multikolinearitas yang minimal. Dalam penentuan banyak *cluster* optimal (k), digunakan metode *Elbow* yang divalidasi oleh *Silhouette Score*. Hasil evaluasi ini menunjukkan bahwa FCM optimal pada $k = 3$ dengan *Silhouette Score* tertinggi (0,3848), sementara *K-Medoids* optimal pada $k = 4$ dengan nilai tertinggi (0,3109). Karena FCM mencapai nilai *Silhouette* yang lebih tinggi, metode ini dianggap paling efisien dan menghasilkan pengelompokan yang terbaik untuk merepresentasikan struktur alami data yang diamati. Hasil pengelompokan *Fuzzy C-Means* (FCM) dengan ($k = 3$) membagi wilayah di Jawa Tengah menjadi tiga kelompok produksi yang berbeda secara signifikan. *Cluster* 1 (7 kabupaten) dicirikan sebagai sentra utama produksi Padi, menunjukkan spesialisasi pada komoditas pangan pokok. Berbeda dengan itu, *Cluster 3* (7 kabupaten) mewakili wilayah dengan produksi yang sangat terdiversifikasi, ditandai oleh rata-rata tertinggi pada komoditas hortikultura dan peternakan (Buah, Sayuran, Susu, dan Telur). Sementara itu, *Cluster 2* (21 kabupaten/kota) merupakan kelompok terbesar

yang merepresentasikan wilayah dengan output pertanian yang paling rendah secara umum di sebagian besar komoditas. Hasil pengelompokan *K-Medoids* ($k = 4$) mengidentifikasi empat kelompok wilayah yang berbeda secara signifikan berdasarkan produksi komoditas. *Cluster* 4 (6 kabupaten) dicirikan sebagai sentra utama produksi Padi, sementara *Cluster* 2 (6 kabupaten) mewakili wilayah dengan diversifikasi dan intensitas tinggi pada sektor hortikultura (Buah dan Sayuran tertinggi). *Cluster* 1 (17 kabupaten) adalah kelompok terbesar yang menunjukkan spesialisasi pada sektor peternakan unggas (Telur tertinggi). Kemudian yang paling menonjol, *Cluster* 3 (6 kota) secara konsisten mengelompokkan wilayah metropolitan (seperti Kota Surakarta dan Kota Magelang) yang memiliki output pertanian yang paling rendah secara umum di hampir semua komoditas, mencerminkan adanya keterbatasan sumber daya lahan.

DAFTAR PUSTAKA

Bardadi, A., Warsito, B., Surarso, B., & Sugiharto, W. H. (2025). *Interval Type-2 Fuzzy C-Means (IT2FCM) for Robust Forecasting of Carbon Monoxide Concentrations in Urban Air Quality Monitoring*. *Mathematical Modelling of Engineering Problems*, *12*(6), 2021–2030. https://doi.org/10.18280/mmep.120617.

Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. Springer Science & Business Media.

bps.go.id. (2023). Hasil Sensus Penduduk (SP2020) pada September 2020 mencatat jumlah penduduk sebesar 270,20 juta jiwa. https://www.bps.go.id/id/pressrelease/2021.

Dwitiyanti, N., Kumala, S. A., & Handayani, S. D. (2024). *Comparative Study of Earthquake Clustering in Indonesia Using K-Medoids, K-Means, DBSCAN, Fuzzy C-Means and K-AP Algorithms*. *Jurnal RESTI*, *8*(6), 768–778. https://doi.org/10.29207/resti.v8i6.5514.

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis (5th ed)*. Wiley.

Fuad, M., Rochman, E. M. S., & Rachmad, A. (2022). *Salt Commodity Data Clustering Using Fuzzy C-Means. Journal of Physics: Conference Series, 2406 (1)*. https://doi.org/10.1088/1742-6596/2406/1/012025.

Grekousis, G. (2020). *Spatial Analysis Methods and Practice*. https://doi.org/https://doi.org/10.1017/9781108614528.

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques (3rd ed)*. Morgan Kaufmann.

Handhayani, T., & Lewenusa, I. (2024). *An Analysis of Meteorological Data in Sumatra and Nearby using Agglomerative Clustering*. *Jurnal RESTI*, *8*(2), 234–241. https://doi.org/10.29207/resti.v8i2.5663.

Härdle, W. K., & Simar, L. (2019). *Applied Multivariate Statistical Analysis. In Applied Multivariate Statistical Analysis*. Springer International Publishing. https://doi.org/10.1007/978-3-030-26006-4/COVER.

Izzadin, F. M. 2019. Optimasi Jumlah *Cluster K-Means* Dengan Metode *Elbow* dan *Silhouette* Untuk Pengelompokan Luas Panen Palawija Kabupaten Magelang Pada Tahun 2017. Universitas Islam Indonesia : Yogyakarta.

Jaya, J., & Handhayani, T. (2025). Perbandingan Algoritma *K-Means* dan *Fuzzy C-Means* untuk Klasterisasi Produktivitas Kedelai di Pulau Jawa. *Jurnal Esensi Infokom : Jurnal Esensi Sistem Informasi Dan Sistem Komputer*, *9*(1), 11–18. https://doi.org/10.55886/infokom.v9i1.949.

Johnson, R. A., & Wichern, D. D. (2007). *Applied Multivariate Statistical Analysis*. PearsonEducation.

Kurmiati, D., Fauzi, M., Z., Ripangi, A. Falegas, and Indria. (2021). *Clustering of Earthquake Prone Areas in Indonesia Using K-Medoids Algorithm*. J. Mach. Learn. Comput. Sci. , vol. 1, no. 1, pp. 47 57.

Kurnia, A. (2023). Perbandingan Algoritma *K-Means* dan *Fuzzy C-Means* Untuk *Clustering* Puskesmas Berdasarkan Gizi Balita Surabaya. *Jurnal PROCESSOR*, *18*(1). https://doi.org/10.33998/processor.2023.18.1.696.

Luo, X. (2020). *Digital Art Design Effectiveness Model System Based on K-Medoids Algorithm*. Adv. Multimed., vol. 2022 doi: 10.1155/2022/2901815.

Maori, N., A., and Evanita, E. (2023). Metode *elbow* dalam optimasi jumlah *cluster* pada *k-means clustering*. Jurnal Simetris, vol. 14, no. 2, pp. 1–11.

Ningsih, S., Wahyuningsih, S., & Nasution, Y. N. (2016). Perbandingan Kinerja Metode *Complete Linkage* dan *Average Linkage* dalam Menentukan Hasil Analisis *Cluster* (Studi Kasus: Produksi Palawija Provinsi Kalimantan Timur 2014/2015). In Prosiding Seminar Sains dan Teknologi FMIPA Unmul (Vol. 1, Issue 1).

Nur'aini, R., Widiharih, T., & Saputra, B. A. (2024). Pengelompokan Kabupaten/Kota Di Provinsi Jawa Tengah Berdasarkan Indikator Kesehatan Bayi dan Balita Menggunakan Algoritma *Fuzzy C-Means* dan *K-Medoids*. *Jurnal Gaussian*, *13*(1), 189–198. https://doi.org/10.14710/j.gauss.13.1.189-198.

Rahakbauw, D. L., Ilwaru, V. Y., & Hahury, M. , H. ,. (2017). Implementasi *Fuzzy C-Means Clustering* dalam Penentuan Beasiswa. Jurnal Ilmiah Matematika Dan Terapan, 11, 1–11.

Rinaldi, A. R., Surlanto, L., Sudrajat, D., & Kurnia, D. A. (2019). Analisa Tingkat Kepuasan Mahasiswa Terhadap Layanan Pembelajaran Menggunakan *K-Means* dan Algoritma Genetika. Jurnal ICT: Information Communication & Technology 18 (1): 60-64.

Sanusi, W., Zaky, A., Besse, D., & Afni, N. (2019). Analisis *Fuzzy C-Means* dan Penerapannya dalam Pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan Berdasarkan Faktor-faktor Penyebab Gizi Buruk. Journal of Mathematics, Computations, and Statistics, 2(1), 47–54. http://www.ojs.unm.ac.id/jmathcos

Sindi, S., Ningse, W., R., O., Sihombing, I., A., F. I. R.H.Zer, and Hartama, D. (2020). Analisis Algoritma *K-Medoids Clustering* Dalam Pengelompokan Penyebaran Covid-19 Di Indonesia. J. Teknol. Inf., vol. 4, no. 1, pp. 166–173 doi: 10.36294/jurti.v4i1.1296.

Sugiyono. 2017. Metode Penelitian Kuantitatif, Kualitatif, dan R&D. Bandung: Alfabeta.

Syam, S., Tokoro, Y., Judijanto, L., Garonga, M., Sinaga, F. M., Umar, N., Handika, I. P. S., Lin, J. N., Apriyanto, & Sitanggang, A. T. (2023). *Data mining*: Teori dan penerapannya dalam berbagai bidang. PT. Sonpedia Publishing Indonesia.

Ou, G., Zhu, Z., Dong, B., & E, Weinan. (2023). *Introduction to Data Science*. World Scientific Publishing Company.