

PERBANDINGAN FUZZY C-MEANS DAN K-MEANS PADA KLASTERISASI KETELA POHON

Sandy Permadi Sormin¹⁾ Teny Handhayani²⁾

^{1) 2)} Teknik Informatika Universitas Tarumanagara
Jalan Letjen S. Parman No. 1, Grogol, Jakarta Barat 11440, Indonesia.
email : ¹⁾ sandy.535230182@stu.untar.ac.id, ²⁾ tenyh@fti.untar.ac.id

ABSTRAK

Ketela pohon merupakan salah satu komoditas strategis dalam sektor pertanian Indonesia yang tersebar di berbagai wilayah dengan tingkat produktivitas yang bervariasi. Penelitian ini bertujuan untuk membandingkan performa dua algoritma klusterisasi, yaitu K-Means dan Fuzzy C-Means (FCM), dalam mengelompokkan wilayah produksi ketela pohon berdasarkan tiga variabel utama: luas panen, total produksi, dan produktivitas per hektar. Data yang digunakan merupakan data sekunder dari Badan Pusat Statistik (BPS) dan Kementerian Pertanian RI untuk periode 2010 hingga 2022. Sebelum dilakukan proses klusterisasi, data melalui tahap pra-pemrosesan seperti pembersihan data, normalisasi menggunakan metode min-max scaling, serta reduksi dimensi dengan Principal Component Analysis (PCA). Proses klusterisasi dilakukan dengan menetapkan jumlah kluster sebanyak tiga, menggunakan kedua algoritma tersebut. Evaluasi performa kluster dilakukan menggunakan metrik Silhouette Score, Davies-Bouldin Index, dan waktu komputasi. Hasil penelitian menunjukkan bahwa algoritma K-Means memiliki nilai Silhouette tertinggi sebesar 0,9341 dan Davies-Bouldin Index terendah sebesar 0,3224, serta waktu komputasi tercepat 0,0066 detik. Di sisi lain, FCM menunjukkan nilai Silhouette terbaik 0,9315 dan Davies-Bouldin Index 0,4896 dengan waktu komputasi minimum 0,0413 detik. Meskipun FCM menawarkan fleksibilitas dalam penanganan data yang bersifat ambigu, K-Means terbukti lebih unggul dalam hal efisiensi dan kualitas pemisahan kluster. Visualisasi hasil klusterisasi dalam bentuk diagram batang dan peta sebar wilayah memperkuat temuan tersebut. Dengan demikian, penelitian ini merekomendasikan penggunaan algoritma K-Means untuk pengelompokan wilayah produksi ketela pohon secara optimal dan efisien.

Key words

Davies-Bouldin Index, Fuzzy C-Means, K-Means, klusterisasi, ketela pohon

1. Pendahuluan

Pertanian merupakan sektor penting dalam pembangunan nasional Indonesia, yang berperan besar dalam ketahanan pangan, pengurangan kemiskinan, serta

penyediaan bahan baku industri. Salah satu komoditas pangan strategis adalah ketela pohon (*Manihot esculenta*), atau dikenal juga sebagai singkong. Ketela pohon memiliki nilai ekonomi yang tinggi karena dapat diolah menjadi berbagai produk turunan seperti tepung tapioka, bioetanol, dan pangan pokok masyarakat. Produksinya tersebar di berbagai provinsi dengan perbedaan signifikan dalam luas tanam, produktivitas, dan total produksi per wilayah [1].

Variasi tersebut dipengaruhi oleh berbagai faktor, seperti karakteristik kesuburan tanah, perbedaan iklim, teknik atau metode budidaya, dan tingkat adopsi teknologi pertanian. Untuk mengatasi kompleksitas tersebut, dibutuhkan pendekatan berbasis data yang dapat mengelompokkan wilayah-wilayah produksi ketela pohon secara sistematis. Salah satu pendekatan yang relevan adalah klusterisasi, yaitu metode analisis data yang mengelompokkan objek berdasarkan kemiripan karakteristiknya. Dengan klusterisasi, wilayah yang memiliki karakteristik produksi yang serupa dapat diidentifikasi untuk kepentingan perencanaan dan intervensi kebijakan pertanian yang lebih tepat sasaran [2].

Algoritma K-Means merupakan metode klusterisasi yang paling umum digunakan karena efisien dan mudah diterapkan. K-Means bekerja dengan membagi data ke dalam sejumlah kluster berdasarkan jarak terdekat ke pusat kluster. Namun, pendekatan ini bersifat kaku karena setiap objek hanya bisa masuk ke satu kluster, yang bisa menjadi kelemahan saat menangani data pertanian yang kompleks dan cenderung tumpang tindih [3].

Sebagai opsi lainnya, algoritma *Fuzzy C-Means* (FCM) menggunakan pendekatan *soft clustering* di mana setiap objek memiliki tingkat keanggotaan terhadap lebih dari satu kluster. Pendekatan ini dinilai lebih fleksibel dan akurat untuk mengelompokkan data yang memiliki karakteristik transisional, seperti wilayah dengan tingkat produktivitas ketela pohon yang beragam. Beberapa penelitian terdahulu menunjukkan bahwa FCM dapat menghasilkan kluster yang lebih representatif dalam konteks data pertanian [4].

Penelitian ini bertujuan untuk membandingkan performa algoritma K-Means dan Fuzzy C-Means dalam mengklusterkan data produksi ketela pohon di Indonesia. Data yang digunakan meliputi luas tanam, total produksi, dan produktivitas berdasarkan kabupaten/kota. Evaluasi performa kedua algoritma dilakukan menggunakan

Commented [H1]: Tambahkan spasi sebelum kutip, jadi wilayah [1], berlaku untuk semua kutipan di dokumen ini

metrik validitas kluster seperti *Silhouette Coefficient* dan *Davies-Bouldin Index*. Dengan hasil evaluasi ini, diharapkan dapat diketahui metode klusterisasi mana yang paling sesuai untuk diterapkan dalam pengelompokan wilayah produksi ketela pohon secara optimal [5].

Ruang lingkup penelitian ini terbatas pada data kuantitatif dari Badan Pusat Statistik dan Kementerian Pertanian Indonesia selama periode 2020–2024. Aspek sosial-ekonomi petani, infrastruktur, dan kebijakan lokal tidak dibahas secara langsung, namun hasil klusterisasi yang diperoleh dapat menjadi dasar awal dalam menyusun strategi pengembangan wilayah berbasis data untuk komoditas ketela pohon di Indonesia [6].

2. Metode

2.1. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode komputasi data berbasis klusterisasi untuk menganalisis pola produksi ketela pohon di Indonesia. Tujuan utama dari metode ini adalah untuk mengelompokkan wilayah-wilayah produksi ketela pohon berdasarkan kemiripan karakteristiknya, yaitu luas tanam, total produksi, dan produktivitas. Pemilihan algoritma klusterisasi K-Means dan Fuzzy C-Means (FCM) didasarkan pada efektivitas keduanya dalam menangani data multivariat dalam sektor pertanian, sebagaimana ditunjukkan dalam berbagai studi sebelumnya [7], [8].

Data yang digunakan merupakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) dan Kementerian Pertanian Republik Indonesia untuk periode tahun 2020 hingga 2024. Variabel yang dianalisis meliputi luas tanam (hektar), produksi total (ton), dan produktivitas (ton/hektar) per kabupaten/kota. Tahapan awal dalam pengolahan data adalah pra-pemrosesan, yang meliputi proses pembersihan data dari nilai kosong dan outlier, serta normalisasi menggunakan metode min-max scaling agar seluruh variabel berada dalam rentang yang sebanding. Proses ini penting dilakukan untuk memastikan tidak ada variabel yang mendominasi pembentukan kluster [9].

2.2. Algoritma Fuzzy C-Means

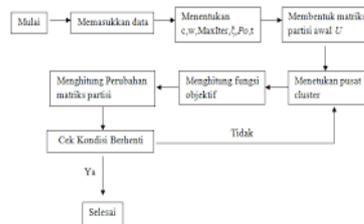
Fuzzy C-Means (FCM) merupakan algoritma klusterisasi berbasis partisi lunak yang memungkinkan setiap data memiliki hubungan atau keterkaitan terhadap lebih dari satu kluster melalui nilai keanggotaan. Menurut [10], metode ini bekerja dengan cara meminimalkan fungsi objektif berbobot untuk menentukan kedekatan suatu data terhadap pusat kluster, sehingga lebih adaptif terhadap data dengan intensitas dan distribusi yang kompleks. Hal yang serupa juga disampaikan oleh [11], yang menyatakan bahwa pendekatan fuzzy ini mengatasi keterbatasan klusterisasi tradisional dengan memberikan

fleksibilitas bagi data yang memiliki karakteristik ambigu atau tidak tersegmentasi secara tegas. FCM sangat sesuai digunakan dalam domain seperti pertanian dan data spasial karena kemampuannya dalam merepresentasikan keterkaitan yang tidak bersifat absolut antar kelompok data.

Fuzzy Dalam implementasinya, FCM bekerja dengan meminimalkan fungsi objektif berbobot, yang mengukur kedekatan antara setiap titik data dengan pusat kluster menggunakan jarak Euclidean. Nilai keanggotaan diperbarui secara iteratif hingga proses mencapai titik konvergensi. Pendekatan ini telah diadopsi dalam sejumlah penelitian terhadap data kompleks, seperti pemetaan penyebaran COVID-19 berbasis data deret waktu multivariat, di mana pola spasial dan temporal yang tumpang tindih berhasil diidentifikasi dengan baik menggunakan FCM [10].

Lebih lanjut, efektivitas FCM dalam menangani data tidak linear juga terlihat dalam studi [11], yang menunjukkan bahwa metode ini memberikan hasil klusterisasi wilayah yang lebih fleksibel dibanding pendekatan hirarkis seperti Agglomerative, khususnya dalam analisis data meteorologi Sumatra. Penelitian tersebut menekankan bahwa pada kondisi data yang tidak memiliki batasan yang tegas antar wilayah, pendekatan fuzzy mampu mencerminkan variasi lokal secara lebih akurat.

Dengan mempertimbangkan kemampuan tersebut, FCM dianggap sangat relevan untuk diterapkan dalam pengelompokan wilayah produksi ketela pohon. Hal ini karena pola produksi di sektor pertanian sering kali dipengaruhi oleh faktor-faktor yang bersifat kontinu dan tidak biner, seperti kondisi iklim, perbedaan teknologi budidaya, serta keragaman sumber daya di tingkat lokal.



Gambar 1. Konsep Model Fuzzy C-Means Clustering

Karakteristik Struktur yang dimiliki oleh algoritma Fuzzy C-Means (FCM) sangat cocok diterapkan dalam klasifikasi wilayah pertanian di Indonesia, termasuk pada komoditas ketela pohon. Di lapangan, karakteristik produksi di suatu daerah sering kali tidak sepenuhnya identik dengan satu kelompok tertentu. Sebuah wilayah bisa saja menunjukkan ciri produktivitas yang berada di antara dua kategori, misalnya antara kelompok dengan hasil tinggi dan kelompok dengan hasil sedang. Metode klusterisasi konvensional seperti K-Means akan langsung menetapkan wilayah tersebut ke dalam satu kluster saja, tanpa mempertimbangkan gradasi kondisinya. Berbeda

Commented [I12]: Tidak perlu spasi antar paragraf, berlaku untuk seluruh dokumen, bukan hanya yang bagian dikomen ini

Commented [I13]: Kutipan hanya dilakukan dengan nomor, contohnya yang lain-lain

dengan FCM yang memberikan fleksibilitas melalui nilai keanggotaan ganda, sehingga satu wilayah bisa dikaitkan dengan lebih dari satu kluster berdasarkan kedekatan karakteristiknya. Pendekatan ini telah terbukti efektif dalam menangani data spasial yang memiliki sifat transisional, sebagaimana ditunjukkan dalam penelitian [12] terhadap segmentasi citra tanah dan [13] dalam pemodelan fluktuasi air tanah yang tidak linier.

Secara matematis, FCM bertujuan untuk meminimalkan fungsi objektif yang mempertimbangkan jarak antara data dan pusat kluster yang ditimbang oleh derajat keanggotaan. Fungsi dan langkah-langkah tersebut dapat dituliskan sebagai berikut:

1. Menginputkan variabel atau data yang akan di cluster X, ke dalam bentuk matriks berukuran $n \times m$ (dimana n adalah jumlah sampel data dan m adalah atribut setiap data). $X_{ij} = \text{data ke } i \text{ (} i = 1,2,3 \dots, n \text{), atribut ke } j \text{ (} j = 1,2, \dots, m \text{)}$.
2. Menentukan parameter yang dibutuhkan, seperti Jumlah cluster (c), pangkat atau bobot (w), maksimum iterasi ($MaxIter$), error terkecil (ϵ), Fungsi objektif ($P_0 = 0$) dan iterasi awal ($t = 1$).
3. Membuat bilangan random $\mu_{ik} (i = 1,2, \dots, n \text{ dan } k = 1,2, \dots, c)$ yang akan dijadikan elemen-elemen pada matriks U. Kemudian, hitung jumlah setiap kolom dengan persamaan: $Q_i = \sum_k \mu_{ik} = 1$ (1)
Dengan $j = 1,2, \dots, n$ maka menghitung matriks random dapat menggunakan persamaan:
 $\mu_{ik} = \frac{u_{ik}}{Q_i}$ (2)
4. Menghitung pusat cluster ke- k dengan persamaan:
 $V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w \cdot X_{ij})}{\sum_{i=1}^n (\mu_{ik})^w}$ (3)
5. Menghitung fungsi objektif pada iterasi ke- t , P_t dengan persamaan :
 $P_t = \sum_{i=1}^n \sum_{k=1}^c (\mu_{ik})^w = 1$ ($\sum_{i=1}^n \sum_{k=1}^c (\mu_{ik})^w = 1$) (4)
6. Menghitung perubahan matriks dengan persamaan:
 $\mu_{ik} = \frac{[\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-\frac{1}{w-1}}}{\sum_{k=1}^c [\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-\frac{1}{w-1}}}$ (5)
Dimana nilai $i = 1,2, \dots, n$ dan nilai $k = 1,2, \dots, c$
7. Mengecek kondisi hingga berhenti dengan persamaan:
 $(P_t - (P_{t-1})) < \epsilon$ (6)
maka berhenti. Jika tidak $t = t + 1$ maka diulangi ke langkah 4.

Dalam proses klusterisasi, dibutuhkan metode evaluasi yang mampu memastikan hasil pengelompokan berjalan secara optimal. Salah satu indikator yang digunakan untuk menilai kualitas kluster adalah

Silhouette Coefficient, yang berfungsi untuk mengukur seberapa baik suatu data berada dalam kelompoknya dibandingkan dengan kelompok terdekat lainnya. Nilai ini akan memberikan gambaran apakah data tersebut memang sesuai dalam kluster yang dipilih atau justru lebih cocok berada di kluster lain. Untuk melakukan perhitungan tersebut, diperlukan ukuran jarak antar data yang akurat, dan salah satu pendekatan yang umum digunakan adalah *Euclidean Distance*. Jarak Euclidean sendiri didasarkan pada akar kuadrat dari selisih kuadrat antar atribut, yang memungkinkan penilaian kedekatan antar data dalam ruang multidimensi[14]. Berikutnya, perhitungan untuk mencari nilai *silhouette coefficient*.

$$S_i = \frac{(b_i - a_i)}{\max(a_i, b_i)} \quad (7)$$

Silhouette coefficient yang dilambangkan dengan S_i yang merupakan salah satu metode yang digunakan untuk mengevaluasi kualitas klusterisasi dengan cara mengukur seberapa baik suatu objek ditempatkan dalam kluster. Nilai ini dihitung berdasarkan dua komponen penting, yaitu a_i , yang merupakan rata-rata jarak antara objek i dengan seluruh objek lain dalam kluster yang sama, serta b_i , yaitu rata-rata jarak antara objek i dengan objek-objek dalam kluster terdekat yang berbeda. Nilai S_i memberikan gambaran seberapa cocok suatu objek berada dalam suatu kluster dibandingkan dengan kluster lainnya. Semakin tinggi nilai S_i (mendekati +1), maka semakin baik objek tersebut terkluster secara tepat; sebaliknya, nilai mendekati nol atau negatif menunjukkan objek mungkin kurang sesuai dengan kluster saat ini atau lebih cocok ditempatkan di kluster lain. Untuk menilai apakah nilai *silhouette coefficient* yang dihasilkan tergolong baik atau tidak, dapat mengacu pada pedoman interpretasi yang disusun oleh Kaufman dan Rousseeuw dalam bentuk tabel, yang berfungsi sebagai acuan standar dalam mengevaluasi hasil klusterisasi secara lebih objektif dan sistematis[14].

Tabel 1. Kaufman dan Rousseeuw *silhouette coefficient*

Nilai <i>silhouette coefficient</i>	Keterangan
$0,7 < silhouette\ coefficient < 1$	Baik
$0,5 < silhouette\ coefficient \leq 0,7$	Medium
$0,25 < silhouette\ coefficient \leq 0,5$	Buruk
$silhouette\ coefficient \leq 1$	Tidak ada

Tabel interpretasi nilai *silhouette coefficient* menurut Kaufman dan Rousseeuw memberikan gambaran penting dalam menilai kualitas hasil klusterisasi. Berdasarkan nilai yang diperoleh, suatu klusterisasi dikatakan memiliki hasil yang sangat baik apabila nilai *silhouette* berada antara 0,7 hingga 1, menandakan bahwa objek sangat sesuai berada di dalam kluster tersebut dan memiliki jarak yang jauh dari kluster lainnya. Nilai antara 0,5 hingga 0,7 mengindikasikan kualitas sedang atau cukup baik, sementara nilai 0,25 hingga 0,5 menunjukkan bahwa objek tidak sepenuhnya sesuai dengan kluster yang ditematinya. Jika nilainya di bawah 0,25, maka struktur kluster dianggap lemah atau bahkan tidak terbentuk

Commented [I14]: Yang betul adalah 'silhouette'

Banyak typo yang sama di dokumen ini

Commented [I15]: Font size jangan berbeda-beda, dan tidak perlu bold

dengan jelas. Penelitian oleh [15] menunjukkan bahwa evaluasi klusterisasi menggunakan *silhouette coefficient* mampu membedakan tingkat efektivitas penggunaan berbagai ukuran jarak dalam algoritma K-Means. Dalam penelitian tersebut, nilai *silhouette* yang tinggi berkorelasi langsung dengan kestabilan dan keakuratan hasil klusterisasi. Hal ini diperkuat oleh studi [16], yang menggunakan *silhouette score* untuk menentukan jumlah kluster optimal pada data pelanggan mal dan menunjukkan bahwa nilai tertinggi (sekitar 0,80) tercapai ketika struktur kluster terbentuk secara optimal. Berdasarkan temuan tersebut, nilai-nilai *silhouette* yang tinggi menjadi indikator penting dalam validasi kualitas pengelompokan data.

Dalam bidang pertanian di Indonesia, algoritma *Fuzzy C-Means* (FCM) telah terbukti efektif untuk mengelompokkan wilayah berdasarkan kesamaan karakteristik produksi. FCM mampu mengatasi variasi data yang tinggi akibat faktor iklim dan manusia, karena pendekatan keanggotaannya yang fleksibel dalam menentukan batas antar kluster. Penerapan algoritma ini terlihat pada penelitian yang dilakukan oleh [17] yang mengelompokkan data produksi tanaman perkebunan di Provinsi Jawa Timur dan menemukan bahwa nilai *silhouette coefficient* mencapai 0,6413, yang menunjukkan kualitas klusterisasi yang cukup baik. Sementara itu, [18] menggunakan FCM untuk mengelompokkan data hasil panen padi di Provinsi Bali, dan hasilnya menunjukkan tiga kelompok utama dengan nilai indeks validasi (FSI) sebesar 0,758, yang mencerminkan pemisahan kluster yang jelas meskipun data memiliki keragaman tinggi. Berdasarkan temuan tersebut, FCM tidak hanya memberikan hasil yang akurat dalam analisis spasial pertanian, tetapi juga dapat menjadi acuan dalam pengambilan keputusan untuk program intervensi dan kebijakan daerah.

2.3. Algoritma K-Means

Algoritma *K-Means* merupakan salah satu metode pengelompokan data yang cukup umum digunakan dalam bidang pertanian berbasis data, termasuk untuk mengklasifikasikan wilayah produksi komoditas seperti ketela pohon [19]. Metode ini bekerja dengan membagi data ke dalam sejumlah kelompok (kluster) berdasarkan kedekatan nilai terhadap titik pusat (*centroid*) yang mewakili masing-masing kluster. Proses klusterisasi diawali dengan menentukan jumlah kluster yang diinginkan, kemudian *centroid* awal ditentukan secara acak. Setelah itu, setiap data akan ditempatkan ke dalam kluster yang paling dekat secara matematis. Posisi *centroid* kemudian dihitung ulang berdasarkan rata-rata data pada masing-masing kluster, dan proses ini dilakukan secara iteratif hingga posisi *centroid* stabil atau perubahan tidak lagi signifikan.

Keunggulan utama dari algoritma ini adalah efisiensinya dalam menangani data berukuran besar dan kemampuannya mengidentifikasi pola tersembunyi dalam data spasial pertanian. Dalam penelitian oleh [19], *K-Means* digunakan untuk mengelompokkan wilayah

produksi beras di Jawa Timur dan mampu menunjukkan segmentasi yang jelas antara wilayah berproduktivitas tinggi dan rendah. Sementara itu, studi oleh [17] menerapkan metode yang sama untuk menganalisis ketahanan pangan di 29 kabupaten/kota, yang dapat diterapkan pula untuk komoditas seperti ketela pohon, guna menentukan wilayah yang memerlukan intervensi khusus dari sisi produksi atau distribusi.

Algoritma *K-Means* dikenal sebagai salah satu metode klusterisasi yang paling banyak diterapkan dalam dunia analitik data dan pemetaan spasial karena keunggulannya dalam hal struktur algoritma yang sederhana serta efisiensi proses komputasi, terutama ketika dihadapkan pada dataset dalam skala besar. Metode ini bekerja dengan membagi data ke dalam beberapa kelompok berdasarkan tingkat kesamaan karakteristik antar data, sehingga pola atau struktur tersembunyi dalam kumpulan data dapat dikenali dengan lebih sistematis.

Dalam konteks pertanian, *K-Means* sering digunakan untuk segmentasi wilayah produksi berdasarkan parameter produktivitas, termasuk pada komoditas seperti ketela pohon. Sebagai contoh, penelitian oleh [9] menerapkan algoritma ini untuk mengelompokkan provinsi-provinsi di Indonesia menurut tingkat produktivitas ubi kayu, yang hasilnya dapat digunakan sebagai dasar dalam penyusunan kebijakan pertanian berbasis data. Selanjutnya, penelitian oleh [19] menunjukkan bahwa pengelompokan hasil panen ketela pohon di Kabupaten Trenggalek menggunakan algoritma *K-Means* mampu menciptakan kluster yang sesuai dengan kondisi geografis setempat. Tidak seperti algoritma *fuzzy* seperti *Fuzzy C-Means* (FCM) yang memungkinkan satu data menjadi bagian dari beberapa kluster secara bersamaan, *K-Means* menggunakan metode klusterisasi tegas (*hard clustering*), di mana setiap data hanya tergolong ke dalam satu kluster tertentu. Hal ini menjadikannya cocok untuk kebutuhan klasifikasi yang mengharuskan pembagian kelompok yang jelas dan tidak tumpang tindih.

Secara matematis, langkah-langkah dalam algoritma K-Means dapat dituliskan sebagai:

1. Menentukan jumlah *cluster*
2. Menentukan nilai *centroid* untuk awal iterasi, nilai awal *centroid* dilakukan secara acak. Sedangkan jika nilai *centroid* yang merupakan tahap dari iterasi, maka digunakan rumus berikut:

$$V_{ij} = \frac{1}{N} \sum_{k=0}^{N_i} X_{kj} \quad (8)$$

Dimana:

- V_{ij} adalah *centroid* atau rata-rata cluster ke- i untuk variabel ke- j .
- N adalah jumlah data yang menjadi anggota cluster ke- i .
- i, k adalah indeks dari cluster, j adalah indeks dari variabel.
- X_{kj} adalah nilai data ke- k yang ada di dalam *cluster* tersebut untuk variabel ke- j .

Commented [I16]: Perbaiki kata duplikat

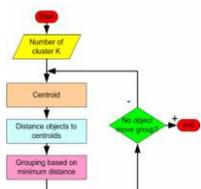
Commented [I17]: Yang betul adalah 'yang'

- Menghitung jarak antara titik *centroid* dengan titik tiap objek, untuk menghitung jarak tersebut dapat menggunakan *Ecludian Distance*, yaitu

$$De = \sqrt{(Xi - Si)^2 + (Yi - Ti)^2} \quad (9)$$

- De* adalah *Ecludian Distance*.
 - i* adalah banyaknya objek.
 - (X, Y) merupakan kordinat objek dan (S, T) merupakan *centroid*.
- Pengelompokan objek Untuk menentukan anggota cluster adalah dengan memperhitungkan jarak minimum objek. Nilai yang diperoleh dalam keanggotaan data pada distance matriks adalah 0 atau 1, dimana nilai 1 untuk data yang dialokasikan ke cluster dan nilai 0 untuk data yang dialokasikan ke cluster yang lain
 - Kembali ke tahap 2, lakukan perulangan hingga nilai *centroid* yang dihasilkan tetap dan anggota cluster tidak berpindah ke cluster lain

Berikut penggambaran algoritma k-means clustering menggunakan flowchart :



Gambar 2. K-Means Clustering

Gambar di atas menunjukkan diagram alir dari proses kerja algoritma *K-Means*, yang merupakan salah satu metode klasterisasi paling banyak digunakan dalam analisis data, termasuk dalam sektor pertanian seperti produksi ketela pohon. Proses dimulai dengan menentukan jumlah klaster yang diinginkan (nilai *K*), kemudian sistem akan menginisialisasi pusat klaster atau *centroid* secara acak. Setelah itu, setiap objek atau data akan dihitung jaraknya ke masing-masing *centroid*. Objek-objek tersebut kemudian dikelompokkan ke dalam klaster yang memiliki jarak terdekat dengannya. Setelah proses pengelompokan, *centroid* dihitung ulang berdasarkan rata-rata dari anggota klaster. Langkah ini dilakukan secara iterative atau berulang hingga tidak ada lagi data yang berpindah klaster, menandakan bahwa proses sudah selesai.

Dalam bidang analisis spasial pertanian berbasis data, algoritma *K-Means* dikenal luas sebagai salah satu teknik klasterisasi yang efektif untuk mengelompokkan wilayah berdasarkan parameter produksi seperti luas lahan, jumlah hasil panen, dan tingkat produktivitas[9]. Penggunaan metode ini sangat sesuai untuk komoditas seperti ketela pohon, di mana variabilitas produktivitas sangat dipengaruhi oleh faktor geografis, iklim, serta praktik budidaya petani setempat. Meskipun terkenal karena kesederhanaannya dan efisiensi komputasi, implementasi *K-Means* memerlukan pertimbangan teknis

yang cermat agar hasil klasterisasi dapat menggambarkan kondisi aktual di lapangan serta mendukung perumusan kebijakan pertanian secara tepat sasaran.

Salah satu tantangan berat dalam penerapan *K-Means* adalah penentuan jumlah klaster optimal (nilai *k*). Jika jumlah klaster terlalu sedikit, wilayah dengan karakteristik yang berbeda bisa tergabung dalam kelompok yang sama, sedangkan klaster yang terlalu banyak dapat menyebabkan hasil yang berlebihan dan menyulitkan dalam interpretasi. Dalam konteks ketela pohon, hal ini menjadi penting mengingat data produksi sering kali dipengaruhi oleh faktor yang kompleks dan bervariasi antarwilayah. Untuk mengatasi hal tersebut, metode evaluasi seperti *Elbow Method* dan *Silhouette Coefficient* kerap digunakan untuk menilai keakuratan dan kualitas pembentukan klaster. Evaluasi ini penting agar hasil pengelompokan benar-benar mencerminkan kondisi pertanian di lapangan dan dapat digunakan sebagai dasar dalam merancang kebijakan, seperti distribusi pupuk bersubsidi atau pengembangan wilayah sentra produksi.

Dalam aplikasi *K-Means* pada data pertanian seperti produksi ketela pohon, tantangan utama adalah kestabilan hasil klasterisasi akibat pemilihan *centroid* awal secara acak, yang bisa menyebabkan hasil berbeda setiap iterasi. Untuk mengatasi hal ini, sebagian peneliti menggunakan pendekatan *hybrid K-Means-PSO* (Particle Swarm Optimization). Salah satu studi yang sudah teruji adalah penelitian oleh [20], yang berhasil membuktikan bahwa penggabungan *PSO* dengan *K-Means* meningkatkan konsistensi dan akurasi pengelompokan data sayuran dibandingkan metode konvensional – terbukti dari meningkatnya nilai *Silhouette* dan menurunnya *quantization error*. Metode ini memberikan hasil klasterisasi yang lebih stabil, yang tentunya sangat bermanfaat untuk pengelompokan wilayah produksi ketela pohon berbasis data spasial.

Dengan segala keunggulan dan keterbatasannya, *K-Means* tetap menjadi pilihan utama dalam tugas-tugas klasterisasi, terutama ketika kecepatan dan skala data menjadi pertimbangan utama. Dalam studi ini, *K-Means* akan dibandingkan dengan *Fuzzy C-Means* untuk menentukan metode mana yang memberikan hasil klasterisasi yang paling relevan terhadap karakteristik data ketela pohon di Indonesia, berdasarkan parameter luas panen, produktivitas, dan hasil produksi.

3. Hasil Percobaan

Eksperimen Dalam eksperimen ini, data produksi ketela pohon diambil dari BPS dan Kementerian Pertanian untuk periode 2010–2022, meliputi tiga variabel utama: luas panen, total produksi, dan produktivitas per hektar—variabel yang mencerminkan kinerja kuantitatif dalam hortikultura. Sebelum klasterisasi, data diuji kualitasnya melalui tahapan prapemrosesan standar: pembersihan data (data cleaning), normalisasi agar variabel seimbang, dan reduksi dimensi menggunakan Principal Component Analysis (PCA)

Commented [I18]: Yang betul adalah 'antara'

Commented [I19]: Lebih baik gunakan bullet point saja untuk level setelah numbered list

Dan indentasi/letak tentu tidak boleh sama dengan numbered list-nya, letakkan ke kanan sedikit

Berlaku untuk seluruh dokumen, bukan hanya bagian ini saja

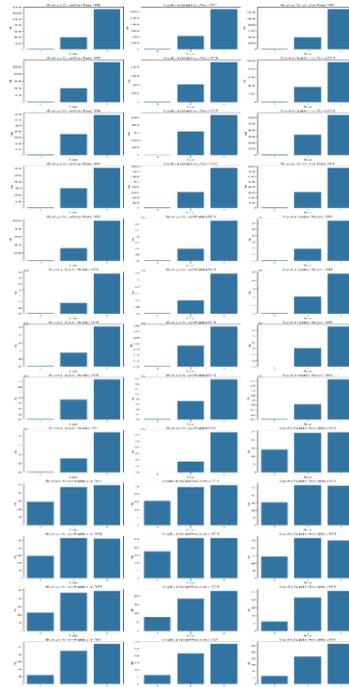
untuk meminimalkan redundansi antar variabel berkorelasi tinggi. Penelitian di Kabupaten Bojonegoro menggunakan PCA pada data pertanian menunjukkan bahwa pengurangan menjadi satu komponen utama dan tiga kluster menghasilkan nilai indeks Davies–Bouldin paling optimal, yakni 0,4072 [21]. Hal ini memperjelas bahwa penggunaan PCA membantu menyederhanakan data kompleks sambil tetap mempertahankan informasi penting dalam analisis ketela pohon.

Kemudian, dilakukan klusterisasi dengan metode K-Means dan Fuzzy C-Means. Jumlah kluster ditetapkan pada tiga ($k=3$), berdasarkan analisis internal menggunakan Elbow Method dan Silhouette Score, yang membuktikan kohesi dan separasi kluster terbaik dari data. Pendekatan ini telah digunakan pada analisis produksi padi di Jawa Barat, di mana metode Silhouette menunjukkan skor evaluasi kualitas kluster 0,39 untuk Elbow dan 0,27 untuk Silhouette – menyimpulkan bahwa metode Silhouette lebih unggul dalam memilih k optimal[22]. Untuk mengamati dinamika temporal, hasil kluster kemudian divisualisasi dalam diagram batang tahunan 2010–2022, menggambarkan tren setiap variabel dalam tiap kluster, sehingga mempermudah pengambilan keputusan berbasis wilayah dalam konteks ketela pohon.

Setelah jumlah kluster ditentukan, proses klusterisasi dilakukan dan hasilnya kemudian divisualisasikan dalam bentuk diagram batang untuk masing-masing tahun, mulai dari 2010 hingga 2022. Diagram ini menampilkan rata-rata nilai variabel luas panen, total produksi, dan produktivitas per hektar untuk setiap kluster, dengan setiap kluster diwakili oleh warna yang konsisten guna memudahkan analisis visual. Pendekatan visual seperti ini biasa digunakan dalam penelitian pengelompokan sektor pertanian di Indonesia; misalnya [23] menerapkan K-Means dan menampilkan hasil kluster sejenis melalui *Business Intelligence* untuk memetakan produktivitas sektor Perkebunan.

Visualisasi diagram batang tersebut sangat efektif dalam membandingkan tren antar tahun dan antar kluster wilayah produksi ketela pohon. Dengan cara ini, kita dapat mengidentifikasi daerah yang mengalami peningkatan produksi atau malah menurun, dan dapat mengaitkannya dengan faktor seperti intervensi program, penggunaan benih unggul, atau perubahan iklim. Hal ini menjadikan output klusterisasi tidak hanya bersifat eksploratif, tetapi juga merupakan alat strategis dalam perencanaan program pertanian berbasis data contohnya menentukan zona untuk pelatihan petani atau distribusi input unggul, sesuai kebutuhan tiap kluster.

Untuk memahami pola hasil klusterisasi secara lebih dalam, data divisualisasikan menggunakan diagram batang sebagai berikut:



Gambar 3. Hasil Variabel, Luas panen, Produksi, dan Produktivitas

Hasil Gambar di atas memperlihatkan hasil visualisasi klusterisasi data produksi ketela pohon dari tahun 2010 hingga 2022. Setiap baris menggambarkan klusterisasi berdasarkan tahun, sementara kolom-kolom menampilkan distribusi data untuk tiga variabel utama: luas panen, total produksi, dan produktivitas. Penyusunan grafik dalam format kronologis dan terpisah per variabel memudahkan pembaca dalam melakukan analisis perbandingan antar waktu maupun antar indikator produksi. Dari visualisasi ini, dapat dilihat pola distribusi data antar kluster yang cenderung stabil untuk beberapa tahun, namun terdapat juga pergeseran signifikan pada beberapa titik waktu. Misalnya, beberapa wilayah yang semula tergolong dalam kluster produktivitas rendah berpindah ke kluster yang lebih tinggi, mengindikasikan potensi keberhasilan program intervensi pertanian atau adanya perbaikan teknologi dan praktik budidaya. Di sisi lain, wilayah yang berpindah ke kluster dengan produktivitas lebih rendah dapat menjadi indikasi gangguan seperti gagal panen, perubahan iklim, atau penurunan kesuburan lahan. Visualisasi ini penting untuk mendukung pengambilan keputusan berbasis data dan menjadi dasar dalam evaluasi kebijakan pertanian di tingkat lokal dan nasional. Dengan demikian, hasil ini tidak hanya membandingkan performa K-Means dan FCM secara teknis, tetapi juga menunjukkan pola

Commented [I110]: Yang betul adalah ‘kluster’

produksi ketela pohon selama satu dekade. Evaluasi kuantitatif kluster akan dibahas lebih lanjut menggunakan metrik seperti Davies-Bouldin Index dan Silhouette Score.



Gambar 4. Pemetaan ketela pohon di Indonesia

Visualisasi di atas menunjukkan pemetaan produksi ketela pohon di Indonesia hasil dari klusterisasi menggunakan algoritma K-Means. Dalam peta ini, setiap kabupaten/kota dikelompokkan ke dalam tiga kluster: produksi rendah (merah), sedang (biru), dan tinggi (hijau) berdasarkan nilai rata-rata produksinya. Sebagian besar wilayah berada dalam kluster produksi rendah, terutama di kawasan timur Indonesia dan sebagian Pulau Jawa. Sementara itu, wilayah dengan tingkat produksi sedang dan tinggi hanya ditemukan di beberapa lokasi dengan potensi produksi yang lebih baik. Peta ini berguna untuk mengenali daerah yang perlu ditingkatkan produktivitasnya, sekaligus menjadi dasar pertimbangan dalam penentuan arah kebijakan sektor pertanian secara lebih terarah.

Untuk menilai kualitas hasil klusterisasi menggunakan algoritma Fuzzy C-Means, digunakan tiga metrik utama, yaitu *Silhouette Score*, *Davies-Bouldin Index*, dan waktu komputasi. Nilai *Silhouette* mencerminkan sejauh mana objek cocok dengan klusternya sendiri dibandingkan dengan kluster lain, sedangkan *Davies-Bouldin Index* mengukur seberapa baik kluster terpisah satu sama lain (nilai lebih rendah lebih baik). Waktu komputasi digunakan untuk mengetahui efisiensi proses algoritma. Tabel berikut menyajikan hasil rata-rata dari ketiga metrik tersebut pada beberapa pengujian:

Tabel 2. Hasil Eksperimen Algoritma Fuzzy C - Means

Jumlah Cluster	Fuzzy C - Means		
	Rata - Rata Silhouette	Rata - Rata Davies-Bouldin Index	Waktu Komputasi
2	0.9315	0.4896	0.0413
3	0.8161	0.7431	0.0461
4	0.7364	0.6926	0.0449
5	0.7110	0.7895	0.2351
6	0.6656	0.8478	0.0611
7	0.5628	1.1551	0.0816
8	0.5074	1.4081	0.0908
9	0.4818	1.5004	0.0818
10	0.4420	1.2576	0.0631

Tabel di atas menunjukkan performa algoritma Fuzzy C-Means berdasarkan tiga metrik evaluasi, yaitu *Rata-rata Silhouette Score*, *Davies-Bouldin Index*, dan *Waktu Komputasi*. Terlihat bahwa pada nilai *Silhouette* tertinggi yaitu 0,9315, didapatkan *Davies-Bouldin Index* paling rendah yaitu 0,4896 dengan waktu komputasi tercepat yaitu 0,0413 detik. Hal ini menunjukkan kualitas kluster yang sangat baik—kluster terpisah jelas dan kompak. Namun, seiring menurunnya nilai *Silhouette*, nilai *Davies-Bouldin* meningkat (misalnya pada nilai *Silhouette* 0,4420, *Davies-Bouldin* mencapai 1,2576), yang menandakan degradasi kualitas klusterisasi.

Waktu komputasi umumnya tetap stabil di bawah 0,1 detik, meskipun sempat meningkat pada satu pengujian hingga 0,2351 detik. Ini menunjukkan bahwa Fuzzy C-Means relatif efisien secara waktu, namun kualitas kluster sangat dipengaruhi oleh parameter atau karakteristik data. Secara keseluruhan, evaluasi ini menegaskan bahwa Fuzzy C-Means mampu menghasilkan kluster yang baik jika kondisi data dan parameter sesuai.

Sebagai perbandingan dengan Fuzzy C-Means, algoritma K-Means juga dievaluasi menggunakan metrik yang sama, yaitu *Rata-rata Silhouette Score*, *Davies-Bouldin Index*, dan *Waktu Komputasi*. Hasil evaluasi K-Means disajikan dalam tabel berikut untuk menunjukkan stabilitas dan kualitas performa kluster yang dihasilkan pada berbagai kondisi.

Tabel 3. Hasil Eksperimen Algoritma K - Means

Jumlah Cluster	K - Means		
	Rata - Rata Silhouette	Rata - Rata Davies-Bouldin Index	Waktu Komputasi
2	0.9341	0.4729	0.0066
3	0.9184	0.3224	0.0215
4	0.8355	0.5277	0.0524
5	0.7176	0.7685	0.0073
6	0.7065	0.7738	0.0082
7	0.7054	0.8287	0.0081
8	0.7302	0.7768	0.1853
9	0.6710	0.7779	0.0110
10	0.6667	0.7623	0.0110

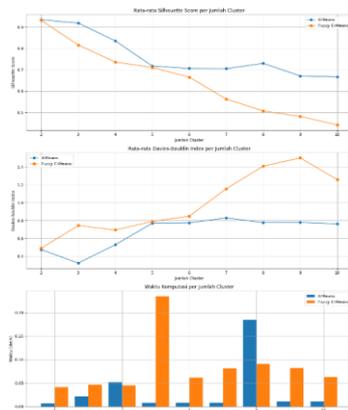
Tabel evaluasi untuk algoritma K-Means menampilkan performa model berdasarkan tiga metrik utama: *Rata-rata Silhouette Score*, *Davies-Bouldin Index* (DBI), dan *Waktu Komputasi*. *Silhouette Score* yang tinggi (mendekati 1) menunjukkan bahwa kluster yang terbentuk memiliki kohesi internal yang baik dan separasi yang jelas dengan kluster lain. Dalam tabel ini, nilai *silhouette* tertinggi tercatat sebesar 0.9341, mengindikasikan pemisahan kluster yang sangat optimal. Sementara itu, nilai *Davies-Bouldin Index* yang lebih rendah menunjukkan kualitas klusterisasi yang lebih baik. Pada hasil ini, nilai DBI terendah mencapai

Commented [H11]: Tabel 1 sudah ada, ini seharusnya Tabel 2

0.3224, yang menunjukkan pemisahan antar kluster yang efisien dengan bentuk kluster yang kompak.

Dari segi waktu komputasi, K-Means menunjukkan efisiensi yang sangat tinggi, dengan sebagian besar eksekusi berlangsung di bawah 0.02 detik. Hanya satu percobaan yang menunjukkan waktu lebih lama (0.1853 detik), kemungkinan disebabkan oleh variasi data awal atau posisi centroid awal yang kurang optimal. Secara keseluruhan, hasil tabel menunjukkan bahwa K-Means memiliki kinerja yang konsisten dalam membentuk kluster yang baik dengan waktu pemrosesan yang cepat, menjadikannya metode yang efisien untuk pengelompokan data produksi ketela pohon.

Untuk memperjelas perbandingan antara algoritma K-Means dan Fuzzy C-Means (FCM), dilakukan visualisasi performa berdasarkan tiga metrik evaluasi utama, yaitu nilai rata-rata Silhouette Score, Davies-Bouldin Index, dan waktu komputasi. Grafik berikut menunjukkan bagaimana masing-masing algoritma berkinerja seiring dengan bertambahnya jumlah kluster dari 2 hingga 10.



Gambar 5. Perbandingan dua algoritma

Pada gambar diatas terdapat tiga grafik. Grafik pertama memperlihatkan nilai rata-rata Silhouette Score untuk masing-masing jumlah kluster. Terlihat bahwa algoritma K-Means secara konsisten menghasilkan skor lebih tinggi dibandingkan Fuzzy C-Means (FCM), terutama pada jumlah kluster 2 hingga 5. Hal ini menunjukkan bahwa struktur kluster hasil K-Means memiliki kohesi internal yang baik dan pemisahan antar kluster yang jelas, sehingga lebih representatif terhadap pola dalam data.

Grafik kedua menggambarkan nilai rata-rata Davies-Bouldin Index. Nilai yang lebih rendah menunjukkan kluster yang lebih baik. K-Means cenderung menghasilkan indeks yang lebih rendah daripada FCM di sebagian besar jumlah kluster, menandakan bahwa kluster yang dibentuk lebih kompak dan terpisah dengan baik. Sebaliknya, nilai indeks FCM meningkat seiring

bertambahnya jumlah kluster, mengindikasikan penurunan kualitas kluster yang terbentuk. Pada grafik ketiga ditampilkan perbandingan waktu komputasi antara kedua metode. FCM menunjukkan kebutuhan waktu yang lebih tinggi dibandingkan K-Means pada hampir seluruh jumlah kluster, terutama pada kluster ke-6 dan ke-8 yang mengalami lonjakan signifikan. Hal ini menandakan bahwa secara komputasional, FCM lebih kompleks dan memerlukan sumber daya yang lebih besar. Berdasarkan ketiga grafik ini, dapat disimpulkan bahwa K-Means unggul dalam hal kualitas kluster dan efisiensi waktu dalam penerapan klusterisasi produksi ketela pohon di Indonesia. Secara keseluruhan, dari ketiga grafik ini dapat disimpulkan bahwa K-Means lebih unggul dari segi kualitas kluster dan efisiensi waktu dibandingkan Fuzzy C-Means dalam kasus klusterisasi produksi ketela pohon di Indonesia.

4. Kesimpulan

Dalam era pertanian modern, pengelolaan data menjadi komponen penting dalam perencanaan pembangunan berbasis wilayah. Salah satu pendekatan analitis yang banyak digunakan untuk memahami pola distribusi hasil pertanian adalah klusterisasi. Dengan teknik ini, wilayah-wilayah yang memiliki karakteristik serupa dapat dikelompokkan ke dalam kluster yang homogen, sehingga memudahkan penyusunan strategi peningkatan produksi secara lebih terarah dan efisien. Sejumlah studi di Indonesia telah membuktikan efektivitas teknik klusterisasi dalam sektor pertanian. Misalnya, Cahaya et al. (2020), menerapkan metode K-Means untuk pengelompokan produktivitas tanaman padi dan berhasil mengidentifikasi zona prioritas untuk peningkatan produksi[24]. Studi lainnya oleh Nurhidayati, N., & Marzuki, I. (2020) menggunakan Fuzzy C-Means untuk memetakan wilayah berdasarkan hasil panen jagung, yang kemudian dijadikan dasar dalam evaluasi program ketahanan pangan daerah[25].

Penelitian ini membandingkan metode K-Means dan Fuzzy C-Means dalam mengelompokkan wilayah produksi ketela pohon berdasarkan luas panen, produksi, dan produktivitas. Evaluasi dilakukan menggunakan metrik Silhouette Score, Davies-Bouldin Index, dan waktu komputasi. Hasil visualisasi dan analisis menunjukkan perbedaan performa antara kedua metode. Berikut adalah poin-poin kesimpulan dari penelitian ini:

- Algoritma K-Means dan Fuzzy C-Means mampu mengelompokkan wilayah produksi ketela pohon berdasarkan variabel luas panen, total produksi, dan produktivitas. Proses klusterisasi menghasilkan 3 kluster utama (rendah, sedang, tinggi), yang divisualisasikan melalui diagram batang dan grafik metrik evaluasi, memudahkan analisis tren antar tahun.

- Evaluasi kualitas klusterisasi menunjukkan K-Means lebih unggul dibandingkan FCM. Hal ini dibuktikan dengan nilai rata-rata Silhouette Score yang lebih tinggi pada K-Means, yaitu mencapai 0.9341 pada kluster 2, dibandingkan FCM yang hanya mencapai 0.9315. Selain itu, nilai *Davies-Bouldin Index* terendah pada K-Means sebesar 0.3224 (kluster 3), jauh lebih kecil dibandingkan FCM yang terbaik hanya 0.4896 (kluster 2), menunjukkan kluster K-Means lebih kohesif dan terpisah dengan baik.
- Dari sisi efisiensi waktu komputasi, algoritma K-Means jauh lebih cepat. Waktu tercepat yang dicapai K-Means adalah 0.0066 detik pada kluster 2, sedangkan FCM paling cepat 0.0413 detik artinya K-Means dapat memproses data hingga 6 kali lebih cepat. Ini penting dalam konteks pengolahan data produksi berskala besar.
- Kekurangan penelitian meliputi keterbatasan pada jumlah kluster yang ditentukan manual, serta belum mempertimbangkan faktor-faktor eksternal seperti curah hujan, jenis varietas, atau dukungan kebijakan daerah.
- Pengembangan selanjutnya dapat mengintegrasikan algoritma optimasi seperti PSO (*Particle Swarm Optimization*) untuk menentukan jumlah kluster secara dinamis, serta memperkaya variabel masukan agar hasil kluster lebih representatif terhadap kondisi pertanian ketela pohon secara menyeluruh.

REFERENSI

- [1] Badan Pusat Statistik, "Statistik Hortikultura Indonesia 2023," BPS, Jakarta, 2023.
- [2] A. N. Hayuningtyas and S. Darwati, "Analisis Kluster Produksi Komoditas Pertanian Unggulan di Indonesia," *J. Mat. dan Sains*, vol. 25, no. 1, pp. 27–36, 2020. [Online]. Available: <https://ejournal.upi.edu/index.php/joms/article/view/24389>
- [3] I. Nur Aulia and A. Intan, "Analisis Kinerja Algoritma K-Means dalam Klusterisasi Wilayah Produksi Pertanian," *J. Ilm. Inform.*, vol. 9, no. 2, pp. 115–122, 2023. [Online]. Available: <https://ejournal.mercubuana-yogyaya.ac.id/index.php/ilmiah-informatika/article/view/539>
- [4] Miftakhul Kharis, R., et al. (2023). Implementasi Algoritma Fuzzy C-Means untuk Klusterisasi Hasil Panen di Provinsi Bali. *Variance: Journal of Statistics and Its Applications*, 5(1), 13–24.
- [5] U. W. Latifah, Sugiyarto, and Suparman, "Comparison of FCM and K-Means Algorithms for Clustering Human Development Index Data," *Desimal J. Mat.*, vol. 5, no. 2, pp. 165–172, 2022. [Online]. Available: <https://jurnal.usk.ac.id/DEK/article/view/23795>
- [6] S. Rohmatullah, D. Purnama, and I. Wulandari, "Pemetaan Wilayah Potensi Produksi Ketela Pohon Berbasis Data Statistik," *J. Agribisnis Indones.*, vol. 8, no. 3, pp. 183–192, 2020. [Online]. Available: <https://journal.ipb.ac.id/index.php/jagbi/article/view/34825>
- [7] Yosia, A., Ishak, R., & Nugroho, S. (2025). Advanced Clustering Approach for Mapping Regions of Paddy Productivity in Indonesia Using Intelligent K-Means. *IEEE International Conference on Data Science and Its Applications (ICoDSA)*. doi: 10.1109/ICoDSA.2025.10932942
- [8] Handhayani, T., & Rusdi, R. (2023). K-Means Using Dynamic Time Warping For Clustering Cities in Java Island According to Meteorological Conditions. *IEEE International Conference on Informatics and Computational Sciences (ICICoS)*. doi: 10.1109/ICICoS.2023.10381899
- [9] Wulandari, R., Pratama, I., & Santoso, B. (2024). Clustering Data Meteorologi Wilayah Indonesia Timur Dengan Metode K-Means dan Fuzzy C-Means. *Jurnal Teknik Informatika (INTI)*, 11(1). Retrieved from <https://ejournal.nusamandiri.ac.id/index.php/inti/article/view/5039>
- [10] Fauzi, A., Prasetyo, R., Susanti, D., & Wahyudi, A. (2023). An Intelligent Clustering Approach for Analyzing a Multivariate Time Series Dataset: Case Study COVID-19 Outbreak in Indonesia. *2023 International Conference on Computer, Control, Informatics and Its Applications (IC3INA)*. doi: 10.1109/IC3INA57724.2023.10367007
- [11] Saputra, M., Putra, R., & Mulyadi, D. (2024). An Analysis of Meteorological Data in Sumatra and Nearby Using Agglomerative Clustering. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 8(1), 141–148. doi: 10.29207/resti.v8i1.5663
- [12] Zeng, S., Wu, Y., Wang, S., & He, P. (2021). Adaptive Scale Weighted Fuzzy C-Means Clustering for the Segmentation of Purple Soil Color Image. *Journal of Intelligent & Fuzzy Systems*, 40(6), 11615–11626. doi: 10.3233/JIFS-202401
- [13] Jafari, M. M., Noroozian, A., Salajegheh, A., & Ebrahimi, K. (2021). Application of a Novel Hybrid Wavelet-ANFIS/Fuzzy C-Means Clustering Model to Predict Groundwater Fluctuations. *Atmosphere*, 12(1), 9. doi: 10.3390/atmos12010009
- [14] Aisah, S. N., Nurcahyani, A., & Rini, D. C. (2022). Implementasi Fuzzy C-Means Clustering (Fcm) Pada Pemetaan Daerah Potensi Transmigrasi Di Jawa Timur. *Jurnal Teknik Informatika UNIKA Santo Thomas*, 07, 33–40. doi: 10.54367/jtiust.v7i1.1841
- [15] Hidayati, R., Rosmansyah, Y., & Lazuardi, R. A. (2021). Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering. *Techno.COM: Jurnal Teknologi Informasi Dan Komunikasi*, 20(2), 186–197. doi: 10.33633/tc.v20i2.4556
- [16] Mulyani, H., Sari, R. N., & Nugraha, R. (2023). Optimization of K Value in Clustering Using Silhouette Score (Case Study: Mall Customers Data). *Journal of Information Technology and Its Utilization*, 6(2), 116–123. doi: 10.56873/jitu.6.2.5243
- [17] Zahra, A. L., Rahayu, S. T., & Ardini, A. F. (2024). Implementasi Clustering Algoritma K-Means Pada Produksi Beras di Provinsi Jawa Timur Tahun 2022. *Journal of Computer and Information Systems Ampera*, 5(3). doi: 10.51519/journalcisa.v5i3.485
- [18] Tuslaela, R., & Supendar, S. (2024). Implementation of K-Means Clustering in Food Security by Regency in East Java Province in 2022. *Sinkron: Jurnal Dan Penelitian Teknik Informatika*, 9(1). doi: 10.33395/sinkron.v9i1.13169

- [19] Hayuningtyas, R. Y., & Darwati, I. (2024). Clustering Hasil Panen Ubi Kayu Menggunakan Algoritma K-Means. *Jurnal Teknik Informasi Dan Komputer (Tekinkom)*, 7(1), 25–32. doi: 10.37600/tekinkom.v7i1.1327
- [20] D. Budiman and I. Winarno, “Optimalisasi K-Means Berbasis Particle Swarm Optimization untuk Hasil Produksi Tanaman Sayuran di Indonesia,” *Ilk. (Ilmu Komputer)*, vol. 11, no. 1, 2022, doi: 10.37034/ilkom.v11i1.6646.
- [21] D. Hedyati and I. M. Suartana, “Penerapan Principal Component Analysis (PCA) Untuk Reduksi Dimensi Pada Proses Clustering Data Produksi Pertanian di Kabupaten Bojonegoro,” *J. Inf. Eng. Educ. Technol.*, vol. 5, no. 2, pp. 49–59, 2021, doi: 10.26740/jieet.v5n2.p49-54.
- [22] P. Vania and B. N. Sari, “Perbandingan Metode Elbow dan Silhouette untuk Penentuan Jumlah Klaster yang Optimal pada Clustering Produksi Padi di Jawa Barat,” *J. Ilm. Wahana Pendidik.*, vol. 9, no. 21, pp. 547–558, 2023, doi: 10.5281/zenodo.10081332.
- [23] H. Pratiwi and A. P. W. Wibowo, “Implementasi Algoritma K-Means untuk Mengklaster Kelompok Sektor Perkebunan di Indonesia,” *JOISIE J. Inf. Syst. Informatics Eng.*, vol. 6, no. 1, 2020, doi: 10.35145/joisie.v6i1.2252.
- [24] D. Tri Cahaya, D. Puspita, and R. Syahri, “Penerapan Metode K-Means Clustering Untuk Pengelompokan Potensi Padi Di Kota Pagar Alam,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 2187–2193, 2024, doi: 10.36040/jati.v8i2.9432.
- [25] Nurhidayati and I. Marzuki, “Deteksi Otomatis Penyakit Daun Jagung Menggunakan Teknik Klasterisasi Data dan Operasi Morfologi,” *J. Energy (Jurnal Ilm. Ilmu-ilmu Tek.*, vol. 10, no. 1, pp. 25–32, 2020.

Teny Handhayani, memperoleh gelar Sarjana Komputer (S.Kom.) dari Institut Pertanian Bogor (IPB) dan Magister Ilmu Komputer (M.Kom.) dari Universitas Indonesia (UI). Kemudian beliau memperoleh gelar Ph.D. dalam bidang Ilmu Komputer dari University of York, Inggris. Saat ini beliau aktif sebagai Staf Pengajar pada program studi Teknik Informatika, Universitas Tarumanagara.