

I. PENDAHULUAN

1.1 Latar Belakang

Pangan merupakan kebutuhan manusia paling pokok, karena seiring dengan pertumbuhan penduduk yang semakin meningkat, maka membuat kebutuhan pangan juga ikut meningkat. Oleh karena itu, perlu perhatian yang serius terutama dalam pemenuhannya yaitu melalui sektor pertanian. Salah satu komoditas utama dalam sektor pertanian adalah padi. Melihat begitu pentingnya sektor pertanian khususnya padi tersebut, maka potensi padi yang dimiliki setiap wilayah perlu diperhatikan karena tidak semuanya memiliki kemampuan produksi yang sama, sehingga perlu perencanaan agar dapat dimanfaatkan secara optimal untuk memenuhi kebutuhan pangan yang terus bertambah (Supranto, 2022).

Pada tahun 2024, Indonesia terdiri dari 38 provinsi yang memiliki karakteristik potensi padi yang dipengaruhi oleh beberapa faktor seperti luas panen padi, produktivitas padi, produksi padi, serta produksi beras. Provinsi Jawa Timur tercatat memiliki luas panen padi yang tinggi yakni 1,61 juta hektar. Kemudian Provinsi Jawa Tengah dan Jawa Barat berada di urutan kedua dan ketiga dengan luas panen padi masing-masing 1,55 juta hektare dan 1,51 juta hektar. Ketiganya menempati posisi teratas baik dari sisi produktivitas padi, produksi padi dan produksi beras (BPS, 2024). Namun, di sisi lain terdapat beberapa provinsi dengan luas panen rendah seperti Provinsi Papua Pegunungan yang hanya memiliki luas panen 9,66 hektare, lalu Provinsi Kepulauan Riau sebesar 113,03 hektare dan Provinsi Papua Barat Daya 363,87 hektare (BPS, 2024).

Kondisi tersebut menunjukkan adanya perbedaan dan penyebaran *outlier* pada data potensi padi tahun 2024. Beberapa provinsi seperti Jawa Timur, Jawa Tengah, dan Jawa Barat memiliki luas panen dan produksi sangat tinggi, sementara provinsi lain seperti Papua Pegunungan, Kepulauan Riau, dan Papua Barat Daya justru berada pada posisi paling rendah. Kondisi ini memperlihatkan bahwa data padi tidak merata dan cenderung timpang. Oleh karena itu, diperlukan analisis lebih lanjut untuk mengelompokkan wilayah-wilayah berdasarkan karakteristik tiap provinsi agar provinsi-provinsi dengan karakteristik serupa dapat dikelompokkan dalam satu *cluster*, sementara provinsi dengan karakteristik berbeda ditempatkan pada *cluster* lain. Pengelompokan ini berguna melihat potensi di Indonesia berdasarkan wilayah (provinsi) (Supranto, 2022).

Pengelompokan provinsi berdasarkan potensi padi tersebut dapat menggunakan konsep dalam ilmu matematika yaitu analisis *cluster*, dimana analisis ini merupakan analisis pengelompokan objek berdasarkan karakteristik tertentu, sehingga objek tersebut memiliki kesamaan satu sama lain, sedangkan objek di *cluster* yang berbeda menunjukkan perbedaan yang signifikan (Everitt, 2011). Analisis *cluster* terbagi beberapa metode yaitu metode *hierarki* dan *non-hierarki*. Metode *hierarki* melibatkan pembentukan *cluster* secara bertahap. Metode ini cocok digunakan ketika jumlah *cluster* tidak diketahui dan cocok untuk penelitian yang objeknya tidak terlalu banyak (Purnomo dan Sutadji, 2022). Di sisi lain, metode *non-hierarki* melibatkan penentuan *cluster*, sehingga metode ini objek-objek akan dikelompokkan ke dalam jumlah *cluster* yang telah ditentukan. Metode ini umumnya lebih efisien karena prosesnya lebih cepat, mudah diterapkan, dan mampu menghasilkan kelompok yang lebih kompak dibandingkan metode *hierarki* (Suyanto, 2017).

Pada penelitian ini metode yang dipakai yaitu metode *non-hierarki*. Metode *non-hierarki* yang umum diterapkan adalah metode K-Means dan K-Medoids. Keduanya adalah metode pengelompokan yang dimulai dengan menentukan terlebih dahulu sejumlah *cluster* awal yang diinginkan. Namun, kedua metode ini memiliki perbedaan pada penentuan pusat *clusternya*. K-Means pusat *clusternya* berdasarkan rata-rata dalam *cluster*, sehingga K-Means ini sensitif pada *outlier*. Sedangkan K-Medoids pusat *clusternya* menggunakan data asli yang ada di objek jadi metode ini lebih stabil jika data memiliki *outlier* (Han et al, 2012). Permasalahan data padi ditandai dengan adanya keragaman antar provinsi dan penyebaran data yang mengandung *outlier*, sehingga menjadi alasan utama digunakannya kedua metode tersebut secara bersamaan. Misalnya, perbedaan ekstrem antara provinsi dengan luas panen yang tinggi seperti Provinsi Jawa Timur dengan provinsi luas panen yang rendah seperti Provinsi Papua Pegunungan menunjukkan adanya sebaran data yang tidak merata. Tanpa metode yang memperhitungkan *outlier*, hasil pengelompokan berpotensi bias.

Kemudian setelah ditentukan metode pengelompokan yang digunakan, langkah selanjutnya adalah menentukan jumlah *cluster* yang optimal. Dalam penelitian ini digunakan metode *Elbow*, yaitu metode yang melihat perubahan nilai *Within Cluster Sum of Squares* (WSS) pada setiap penambahan jumlah *cluster*. Secara umum, nilai WSS akan semakin menurun seiring bertambahnya jumlah *cluster*, namun pada titik tertentu penurunan tersebut mulai melambat

dan membentuk pola “siku” (*elbow*). Titik siku inilah yang dianggap sebagai jumlah *cluster* optimal (Wiroromi et al., 2024).

Pada penggunaan metode K-means dan K-Medoids tersebut tentu nantinya akan menghasilkan hasil pengelompokan yang berbeda-beda. Oleh karena itu, untuk menilai hasil pengelompokan, penelitian ini menggunakan metode validasi *Davies-Bouldin Index* (DBI) yang mengukur kekompakan (kohesi) dan keterpisahan antar *cluster* (separasi), sehingga efektif membandingkan metode K-Means dan K-Medoids pada data potensi padi yang bervariasi antarwilayah (Dewi dan Pramita, 2019). Dengan demikian, DBI menjadi metode paling sesuai karena dapat membuat anggota dalam satu *cluster* saling berdekatan sehingga *cluster* yang terbentuk lebih kompak, dan memaksimalkan jarak antar *cluster* sehingga perbedaan antar kelompok provinsi tetap jelas. Dengan dasar tersebut, algoritma dengan nilai DBI terendah dipilih sebagai metode terbaik untuk membagi 38 provinsi berdasarkan potensi padi secara akurat, terstruktur, dan sesuai dengan variasi data yang ada.

Penelitian ini relevan terhadap penelitian yang sebelumnya telah dilakukan oleh beberapa peneliti lainnya. Diantaranya penelitian yang dilakukan Rizki Hesnanda (2025) yaitu dengan menerapkan K-Means *clustering* terhadap produktivitas padi di pulau sumatera sebagai strategi mendukung ketahanan Pangan. Penelitian ini menggunakan metode evaluasi yaitu *Davies-Bouldin Index* (DBI) dan *Silhouette Score*. Berdasarkan hasil validasi menunjukkan model *clustering* sangat baik, *Silhouette Score* sebesar 0,7147 dan DBI sebesar 0,3752, yang menandakan *cluster* yang terpisah dengan jelas dan valid.

Penelitian lainnya yang dilakukan oleh Ismail et al., (2025), dengan melakukan penerapan teknik pengelompokan K-Means dan K-Medoids dengan menggunakan metode evaluasi yaitu *Silhouette* pada hasil padi di Nigeria. Hasil dalam penelitian ini, K-Medoid menghasilkan nilai *Silhouette* yang sedikit lebih tinggi yaitu sebesar 0,42, menunjukkan *clustering* yang lebih baik dibanding K-Means yang sebesar 0,39 untuk konteks data padi di Nigeria.

Berdasarkan permasalahan yang telah dipaparkan di atas, dapat terlihat bahwa karakteristik potensi padi di Indonesia ditandai oleh adanya perbedaan yang sangat jauh antar provinsi. Baik dari sisi luas panen, produktivitas, produksi padi, maupun produksi beras. Sebagian provinsi memiliki kapasitas luas panen, produktivitas, produksi padi dan produksi beras yang tinggi, sementara sebagian lainnya justru berada pada tingkat yang sangat rendah, sehingga membentuk pola distribusi data yang tidak merata dan mengandung

outlier. Kondisi ini menjadi dasar penting dilakukannya pengelompokan, agar provinsi-provinsi dengan potensi serupa dapat dikelompokkan dalam satu *cluster* sehingga karakteristiknya lebih jelas terlihat, sedangkan provinsi dengan potensi berbeda digolongkan pada *cluster* lain. Dengan demikian, hasil pengelompokan tidak hanya menggambarkan potensi padi di Indonesia, tetapi juga memberikan informasi karakteristik masing-masing *cluster*. Oleh karena itu, maka dari latar belakang di atas peneliti mengangkat penelitian dengan judul yaitu **“Analisis Pengelompokan Provinsi Berdasarkan Potensi Padi di Indonesia Tahun 2024 Dengan Algoritma K-Means dan K-Medoids Menggunakan *Davies-Bouldin Index*”**.

1.2 Rumusan Masalah

Adapun rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana hasil pengelompokan provinsi di Indonesia berdasarkan potensi padi tahun 2024 menggunakan algoritma K-Means dan K-Medoids?
2. Apa algoritma yang lebih sesuai digunakan untuk mengelompokkan provinsi di Indonesia berdasarkan potensi padi tahun 2024 jika dilihat dari nilai validasi *Davies-Bouldin Index*?

1.3 Tujuan Penelitian

Adapun tujuan dalam penelitian ini adalah sebagai berikut:

Mengetahui hasil dari pengelompokan provinsi yang ada di Indonesia berdasarkan potensi padi pada tahun 2024 dengan menggunakan algoritma K-Means dan K-Medoids.

Membandingkan kinerja algoritma K-Means dan K-Medoids dengan menggunakan validasi *Davies-Bouldin Index* sehingga dapat ditentukan algoritma yang paling sesuai dalam pengelompokan potensi padi di Indonesia.

1.4 Manfaat Penelitian

Adapun manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Bagi penulis adalah menambah wawasan dalam analisis data menggunakan algoritma K-Means dan K-Medoids serta memperoleh pengalaman dalam mengolah dan mengevaluasi data menggunakan *Davies-Bouldin Index* (DBI).
2. Bagi pembaca adalah memberikan pemahaman tentang penerapan metode *clustering* terutama algoritma K-Means dan K-medoids dan

menambah wawasan tentang kelebihan dan kekurangan algoritma K-Means dan K-Medoids.

3. Bagi pemerintah adalah menyediakan informasi terkait pengelompokan provinsi di Indonesia berdasarkan potensi padi pada tahun 2024 untuk memperbaiki sistem pertanian di Indonesia dan mengurangi tingkat impor bahan pangan tersebut.

1.5 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Penelitian ini menggunakan 4 variabel yaitu luas panen padi, produksi padi, produktivitas padi, dan produksi beras.
2. Penelitian ini menggunakan grafik metode *Elbow* untuk menentukan jumlah *cluster* yang optimal.
3. Penelitian ini menggunakan jarak *Euclidean* sebagai ukuran jarak dalam analisis yang dilakukan.
4. Penelitian ini menggunakan metode validasi *Davies-Bouldin Index* untuk menentukan hasil *cluster* yang optimal antara algoritma K-Means dan K-Medoids.