

BAB I

PENDAHULUAN

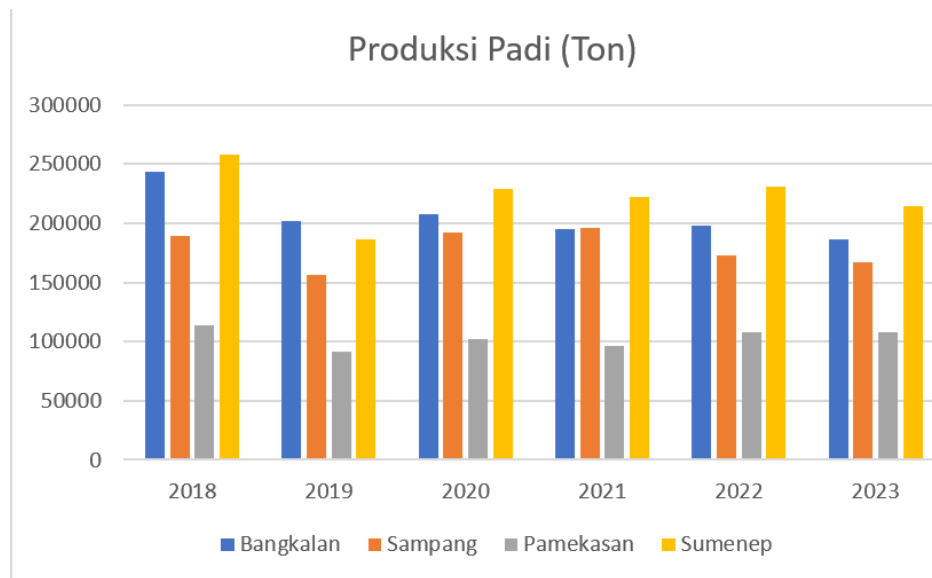
1.1 Latar Belakang

Pertanian merupakan sektor pemegang peranan penting dalam kesejahteraan masyarakat Indonesia. Sebagai negara agraris, sebagian penduduknya menggantungkan hidup mereka pada sektor pertanian atau perkebunan[1]. Pertanian tetap menjadi sektor paling utama dalam perekonomian Indonesia, karena 70 persen penduduknya bekerja di sektor ini dan aktivitas terkait sebagai sumber mata pencaharian mereka[2]. Luasnya sektor pertanian di Indonesia memberikan peluang besar terhadap pembangunan negara terutama pada produksi beras yang menjadi makanan pokok bagi masyarakat. Beras sebagai hasil dari tanaman padi merupakan komoditas penting bagi Indonesia. Keberadaan beras yang terjangkau menjadi salah satu faktor penting bagi ketahanan nasional. Beras merupakan tanaman pangan utama yang masih menjadi makanan pokok dan merupakan komoditas strategis di Indonesia.

Produksi padi selalu menjadi topik utama yang menarik perhatian oleh berbagai pihak, hal ini dikarenakan tingkat produksi padi yang tinggi ataupun rendah menjadi tolak ukur bagi ketahanan pangan di Indonesia. Ketahanan pangan ini juga merupakan isu dalam prioritas untuk pembangunan nasional. Terjadinya penurunan dalam produksi padi akan dapat mengancam kestabilan pangan, sehingga akan berdampak pada ketergantungan terhadap impor beras yang mana dapat mempengaruhi perekonomian nasional secara signifikan[3]. Dalam mengantisipasi permasalahan tersebut, langkah strategis yang dapat dilakukan adalah dengan menjaga serta meningkatkan produksi beras dalam negeri dengan memperhatikan potensi daerah-daerah penghasil beras utama di Indonesia guna memperkuat ketahanan pangan nasional secara berkelanjutan.

Pulau Madura merupakan salah satu wilayah penghasil padi di Indonesia, yang terdiri dari empat kabupaten, yaitu Bangkalan, Sampang, Pamekasan, dan Sumenep. Berdasarkan data dari Badan Pusat Statistik (BPS) Provinsi Jawa Timur, produksi padi di keempat kabupaten tersebut mengalami fluktuasi selama periode 2018 hingga 2023. Pada tahun 2019, seluruh kabupaten mengalami penurunan produksi dibandingkan tahun sebelumnya. Selanjutnya, pada tahun 2020 terjadi

peningkatan produksi di semua kabupaten. Namun, produksi kembali menurun pada tahun 2021, kemudian meningkat kembali di tahun 2022, kecuali Kabupaten Sampang yang justru mengalami penurunan. Pada tahun 2023, produksi padi kembali menurun di seluruh kabupaten. Data produksi padi tersebut dapat dilihat secara visual pada Gambar 1.1 berikut:



Gambar 1. 1 Data Produksi Tanaman Padi Madura Tahun 2018-2023

Berdasarkan Gambar 1.1 hasil produksi padi dari tahun ke tahun menunjukkan ketidakstabilan. Produksi padi yang tidak konsisten ini mempengaruhi masyarakat karena padi merupakan komoditas utama dan sumber pangan pokok bagi masyarakat. Untuk mengatasi masalah ini perlu dilakukan langkah-langkah perbaikan, salah satunya melakukan pengelompokan potensi produksi padi guna mendukung pemerintah daerah dalam menentukan daerah yang menjadi kawasan sentra padi. Pengelompokan produksi padi ini juga berguna untuk memudahkan analisis produktivitas padi tiap daerah yang mana hal ini berfungsi sebagai dasar dalam melakukan identifikasi mengenai faktor-faktor apa saja yang mempengaruhi tingkat produksi padi. Sehingga, dengan dilakukannya pengelompokan ini akan memiliki manfaat untuk memetakan daerah mana saja yang perlu mendapatkan perhatian khusus agar produksi padi dapat ditingkatkan dan ketahanan pangan tetap terjaga[3].

Clustering merupakan bagian dari model pada teknik data mining atau juga dikenal dengan panggilan *knowledge discovery in database (KDD)* yang mana

bertujuan untuk mengelompokkan data berdasarkan kemiripan terhadap objek data serta meminimalisir kemiripan terhadap kluster lain[4]. Terdapat beberapa algoritma yang dapat digunakan untuk melakukan clustering, di antaranya adalah *K-Means*, *K-Medoids*, dan *Fuzzy C-Means*. *K-Means* adalah metode yang termasuk ke dalam *unsupervised learning* dimana melakukan pengelompokan data ke dalam *cluster* berdasarkan variabel dengan tujuan untuk meminimalisir jarak antar titik data dengan *cluster* terkait[5]. Metode ini memiliki beberapa kelebihan utama, yaitu sederhana, efisien, mudah dipahami, serta mudah diterapkan. Akan tetapi terdapat kekurangan dari metode ini, yaitu pemilihan titik pusat (*centroid*) dilakukan secara acak yang mana dapat mempengaruhi hasil akhir *clustering*. Sebagai solusi dari kekurangan tersebut, digunakan metode *K-Medoids* yaitu dengan melakukan pemilihan pusat *cluster* dari titik data aktual yang paling representatif dalam *cluster* tersebut[6]. Hal ini menjadikan metode ini lebih tahan terhadap data *outlier* serta dapat menghasilkan cluster yang lebih stabil karena pemilihan pusat *cluster* tidak bergantung pada rata-rata seperti pada metode *K-Means*, melainkan pada data aktual yang memiliki total jarak minimum terhadap anggota cluster lainnya[7]. Selain itu, Fuzzy C-Means juga menjadi alternatif yang efektif untuk mengatasi kekurangan pada K-Means. Metode ini menggunakan prinsip keanggotaan fuzzy yang mana tiap data dapat menjadi anggota dari beberapa cluster sekaligus dengan derajat keanggotaan tertentu dalam rentang 0 hingga 1. Pendekatan ini lebih fleksibel terutama dalam kasus data yang berada diantara beberapa cluster karena memungkinkan interpretasi yang lebih realistis terhadap distribusi data[8].

Penelitian terdahulu mengenai *Clustering* atau pengelompokan data sudah banyak dilakukan. Penelitian yang dilakukan oleh Ratih HafSarah Maharrani, Prih Diantono Abda'u, dan Hety Dwi Hastuti (2021) melakukan pengelompokan kesesuaian lahan kayu putih di Kabupaten Cilacap menggunakan algoritma *K-Means*. Hasil penelitian yang diperoleh, lahan terbagi menjadi dua *cluster* yaitu lahan yang sangat berpotensi dan tidak berpotensi terhadap serangan hama rayap. Sebanyak 13,56% lahan (16 lahan) teridentifikasi sebagai sangat berpotensi, sementara 86,44% lahan (112 lahan) tidak berpotensi terhadap serangan rayap. Peneliti juga melakukan evaluasi hasil *cluster* menggunakan *Davies Boulding Index* (DBI), diperoleh hasil optimal sebesar 0,055 yang mana ini menunjukkan

pemisahan *cluster* yang baik[9]. Kemudian penelitian yang dilakukan oleh Luth Fimawahib, Imam Rangga Bakti, dan Asep Supriyanto (2022) melakukan pengelompokan produksi padi dan beras di Provinsi Riau dari tahun 2019 hingga 2021 menggunakan metode *K-Medoids*. Dalam penelitian tersebut menghasilkan pengelompokan wilayah produksi padi dan beras terbagi menjadi dua *cluster*, yaitu cluster dengan produksi tinggi berjumlah 4 kabupaten dan produksi rendah berjumlah 8 kabupaten. Hasil evaluasi menggunakan *Davies-Bouldin Index* (DBI) diperoleh sebesar 0,626, yang menunjukkan bahwa hasil *clustering* tergolong cukup baik[10]. Selain itu, penelitian yang dilakukan oleh Ira Halimatuz Zahro, Ulya Anisatur Rosyidah, dan Luluk Handayani (2024) mengimplementasikan metode Fuzzy C-Means untuk mengelompokkan 34 provinsi di Indonesia berdasarkan kualitas perguruan tinggi menggunakan data Angka Partisipasi Kasar (APK) dengan rentang tahun 2019 hingga 2022, sebanyak 136 data. Hasil penelitian menunjukkan hasil terbaik pada 2 cluster dengan nilai PCI sebesar 0,83491[11].

Penelitian terdahulu terkait perbandingan algoritma *clustering* sudah banyak dilakukan. Penelitian yang dilakukan oleh Siti Nadilla, Wirda Syahrifa, dan rekan-rekannya (2024) melakukan perbandingan algoritma K-Means dan *K-Medoids* dalam klasterisasi produksi padi di Pulau Sumatra dengan rentang tahun 2010 hingga 2020. Hasil menunjukkan bahwa *K-Medoids* menghasilkan *cluster* yang lebih stabil dan akurat dibandingkan dengan *K-Means*, dimana untuk nilai *Davies-Bouldin Index* (DBI) *K-Medoids* sebesar 0,43, lebih rendah dibandingkan dengan *K-Means* yaitu sebesar 0,440[6]. Kemudian penelitian yang dilakukan oleh Lily Wulandari dan Bima Olga Yogantara (2022) membandingkan algoritma *K-Means* dan *Fuzzy C-Means* untuk mengelompokkan 167 negara berdasarkan faktor ekonomi dan kesehatan. Dalam penelitian ini untuk evaluasi digunakan empat parameter yaitu *Davies Bouldin Index* (DBI), *Calinski Harabasz*, dan *Execution Time*. Hasil yang diperoleh algoritma *K-Means* menghasilkan 32 negara maju dan 135 negara berkembang, sedangkan *Fuzzy C-Means* menghasilkan 33 negara maju dan 134 negara berkembang. Hasil evaluasi menunjukkan bahwa *K-Means* unggul pada nilai DBI dan waktu eksekusi yaitu 0,6606 dan 0,001996 detik, sementara *Fuzzy C-Means* lebih baik pada *Silhouette Coefficient* yaitu 0,896[12]. Selanjutnya, penelitian yang dilakukan oleh Syafa Marwa Sabrin dan Tabah Heri Setiawan

(2025) melakukan perbandingan tiga metode *clustering*, yaitu *K-Means*, *K-Medoids*, dan *Fuzzy C-Means* untuk mengelompokkan provinsi di Indonesia dengan berdasarkan Indikator Indeks Pembangunan Manusia (IPM) di tahun 2023. Evaluasi yang digunakan pada penelitian ini adalah *Silhouette Coefficient* dan rasio standar deviasi dalam dan antar *cluster* (Sw/Sb). Hasil yang diperoleh menunjukkan bahwa *K-Medoids* memberikan performa terbaik yaitu dengan nilai *Silhouette Coefficient* sebesar 0,5398 pada $k = 3$, dimana hasil ini menandakan bahwa struktur *cluster* tergolong baik. Sementara itu, *K-Means* dan *Fuzzy C-Means* menghasilkan nilai *Silhouette Coefficient* yang lebih rendah yaitu sebesar 0,3164 pada $k = 3$ dan 0,2877 pada $k = 3$, yang menandakan bahwa struktur *cluster* tergolong lemah[13].

Pada penelitian ini menganalisa 4 (empat) parameter tanaman padi yaitu luas tanam, luas panen, produktivitas, dan produksi. Mengetahui pengelompokan produksi padi di Pulau Madura dengan membandingkan metode *K-Means*, *K-Medoids*, dan *Fuzzy C-Means*.

Oleh karena itu, penelitian ini mengusulkan pengelompokan data tanaman padi dengan menggunakan metode *K-Means*, *K-Medoids*, dan *Fuzzy C-Means* serta dilakukan uji evaluasi menggunakan *Davies-Bouldin Index* (DBI). Adanya pengelompokan ini diharapkan dapat membantu dan mempermudah Dinas Pertanian dan Ketahanan Pangan dalam pemantauan, perkembangan produksi padi serta pengetahuan dalam mengambil kebijakan mengenai strategi dalam meningkatkan produktivitas padi yang ada di Madura dengan lebih efektif.

1.2 Rumusan Masalah

1.2.1 Permasalahan

Permasalahan utama dalam penelitian ini adalah belum meratanya tingkat produktivitas padi di berbagai wilayah Pulau Madura tepatnya Kabupaten Bangkalan, Sampang, Pamekasan, dan Sumenep. Variasi produktivitas ini menyulitkan pemerintah dalam menyusun strategi peningkatan hasil panen secara konsisten dan berkelanjutan. Belum adanya sistem analisis yang mampu memetakan karakteristik wilayah secara objektif berbasis data. Oleh karena itu, diperlukan metode pengelompokan berbasis data mining untuk memetakan wilayah berdasarkan produktivitas padinya. Hasil pemetaan ini dapat menjadi dasar bagi

pemerintah daerah dalam merancang kebijakan yang lebih tepat sasaran, seperti penyediaan sarana dan prasarana pertanian serta penyusunan program strategis guna meningkatkan ketahanan pangan dan mengoptimalkan potensi produksi padi di Pulau Madura.

1.2.2 Metode Usulan

Pada penelitian ini mengusulkan penggunaan Algoritma *K-Means*, *K-Medoids*, dan *Fuzzy C-Means* untuk mengelompokkan wilayah di Pulau Madura berdasarkan tingkat produktivitas padi. Evaluasi kualitas *cluster* dilakukan dengan *Davies-Bouldin Index* (DBI). Selain itu, evaluasi efisiensi juga dilakukan berdasarkan waktu eksekusi dan jumlah iterasi yang dibutuhkan pada masing-masing metode. Data yang digunakan mencakup empat kabupaten di Madura dengan parameter data yaitu luas tanam, luas panen, produktivitas, dan produksi.

1.2.3 Pertanyaan penelitian

Bagaimana perbandingan hasil pengelompokan produktivitas padi di Pulau Madura menggunakan metode K-Means, K-Medoids, dan Fuzzy C-Means berdasarkan nilai evaluasi Davies-Bouldin Index, waktu eksekusi, dan jumlah iterasi yang dihasilkan dari masing-masing metode?

1.3 Tujuan Dan Manfaat

1.3.1 Tujuan Penelitian

Untuk mengetahui perbandingan hasil pengelompokan produktivitas padi di Pulau Madura menggunakan metode K-Means, K-Medoids, dan Fuzzy C-Means berdasarkan nilai evaluasi Davies-Bouldin Index, waktu eksekusi, dan jumlah iterasi yang dihasilkan dari masing-masing metode.

1.3.2 Manfaat Penelitian

Adapun manfaat yang diperoleh dari penelitian ini adalah sebagai berikut :

A. Bagi Penulis

Dapat mengimplementasikan ilmu yang diperoleh selama perkuliahan mengenai data mining dan algoritma *clustering* dalam studi kasus nyata, serta

memperluas wawasan terkait pengelompokan wilayah berdasarkan tingkat produksi padi.

B. Bagi Pembaca

Dapat meningkatkan pengetahuan dan wawasan pembaca mengenai metode *K-Means*, *K-Medoids*, dan *Fuzzy C-Means* serta sebagai referensi bagi pembaca untuk melakukan penelitian mengenai *clustering*.

1.4 Batasan Masalah

Berdasarkan pemaparan yang telah disampaikan sebelumnya, maka dilakukan penentuan masalah guna menghindari penyimpangan dari sasaran yang diharapkan. Oleh karena itu, penulis melakukan pembatasan ruang lingkup permasalahan. Adapun batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan merupakan data produksi padi dari Dinas Ketahanan Pangan dan Pertanian di empat kabupaten yang ada di pulau Madura yaitu Bangkalan, Sampang, Pamekasan, dan Sumenep dari tahun 2018 hingga 2023, dengan total keseluruhan data sebanyak 426 data.
2. Terdapat 7 variabel berdasarkan data yang diperoleh yaitu kabupaten, kecamatan, tahun, luas tanam (Ha), luas panen (Ha), produktivitas (Ton/Ha), dan produksi (Ton). Dalam penelitian ini akan digunakan 4 variabel yaitu luas tanam (Ha), luas panen (Ha), produktivitas (Ton/Ha), dan produksi (Ton).
3. Data diolah dengan membandingkan metode *K-Means*, *K-Medoids*, dan *Fuzzy C-Means*.
4. Evaluasi model *clustering* menggunakan *Davies Bouldin Index* (DBI).
5. Dilakukan perbandingan waktu eksekusi dan jumlah iterasi dari masing-masing metode.

1.5 Sistematika Penulisan

Sistematika Penulisan dalam penyusunan laporan skripsi ini terdiri dari beberapa bab, yakni sebagai berikut:

BAB I PENDAHULUAN

Bab pertama merupakan bagian pendahuluan yang mana pada bab ini berisi latar belakang penelitian, perumusan masalah, tujuan dan manfaat dari penelitian yang dilakukan serta batasan-batasan masalah dalam penelitian.

BAB II TINJAUAN PUSTAKA

Bab kedua merupakan bagian tinjauan pustaka yang mana pada bab ini melakukan pembahasan mengenai teori dan penjelasan secara mendetail tentang istilah berdasarkan penelitian yang dilakukan.

BAB III METODE PENELITIAN

Bab ketiga merupakan bagian metode penelitian, dimana pada bab ini memberikan penjelasan mengenai tahapan tahapan penelitian yang akan dilakukan seperti pengumpulan data, analisi permasalahan, metode pengembangan sistem, gambaran umum sistem dan skenario uji coba.

BAB IV HASIL DAN PEMBAHASAN

Bab keempat merupakan hasil implementasi metode yang digunakan dalam penelitian disertai pembahasan, dimana pada bab ini juga akan ditampilkan potongan kode program, hasil analisis, serta implementasi antarmuka dalam bentuk website yang dibangun menggunakan streamlit.

BAB V PENUTUP

Bab kelima merupakan bab terakhir yang memuat kesimpulan yang diperoleh dari hasil penelitian yang telah dilakukan serta saran untuk penelitian selanjutnya yang diberikan oleh penulis.