

BAB I

PENDAHULUAN

Bab ini membahas pendahuluan yang terdiri dari beberapa subbab, yaitu latar belakang, perumusan masalah, tujuan dan manfaat, batasan penelitian, serta sistematika skripsi. Setiap subbab disusun untuk memberikan gambaran menyeluruh mengenai dasar dilaksanakannya penelitian, ruang lingkup yang ditetapkan, serta arah dan struktur penulisan proposal.

1.1 Latar Belakang

Kabupaten Bangkalan adalah wilayah agraris dengan pola lahan dan praktik pertanian yang beragam. Banyak petani masih menentukan jenis tanam berdasarkan pengalaman turun-temurun atau rekomendasi umum, sehingga seringkali pilihan komoditas tidak cocok dengan kondisi tanah dan iklim setempat. Akibatnya, produktivitas tidak optimal dan risiko gagal panen meningkat, khususnya pada lahan yang memiliki variasi mikro-iklim dan sifat tanah dalam skala desa atau kecamatan (Volume et al., 2009).

Perbedaan kondisi tersebut sebenarnya dipengaruhi oleh faktor-faktor agroekologi utama, seperti karakteristik tanah (pH, kandungan unsur hara NPK, tekstur, dan elevasi) serta faktor iklim (curah hujan, suhu, kelembapan, dan radiasi matahari), yang secara langsung menentukan tingkat kesesuaian lahan bagi komoditas tertentu. Namun, tantangan muncul karena data terkait faktor-faktor tersebut memiliki format yang berbeda; sebagian diperoleh dari survei lapangan berbasis titik, sementara sebagian lainnya berupa peta spasial dengan cakupan luas namun resolusi yang masih terbatas. Perbedaan skala, format, dan tingkat detail data ini menyebabkan proses integrasi informasi menjadi tidak sederhana, sehingga pemetaan zona agroekologi yang representatif dan akurat untuk mendukung pengambilan keputusan pertanian di tingkat lokal masih sulit dilakukan (Poggio et al., 2021).

Sebagai pendekatan untuk mengatasi permasalahan tersebut, konsep *Agroecological Zones* (AEZ) dikembangkan untuk mengelompokkan wilayah berdasarkan kombinasi karakteristik tanah, iklim, dan topografi, sehingga setiap zona merepresentasikan kondisi lingkungan yang sesuai dengan karakteristik lahan

dan dapat digunakan sebagai dasar perencanaan geografis maupun praktik agronomis. Pemetaan zona agroekologi akan lebih efektif apabila dibangun dari data yang lengkap, terukur, dan terintegrasi, bukan hanya berdasarkan observasi lokal yang bervariasi. Berbagai studi agroekologi sebelumnya menunjukkan bahwa pendekatan klasterisasi mampu menghasilkan zonasi yang bermanfaat dalam perencanaan pertanian. Namun, sebagian besar penelitian masih bergantung pada dataset berskala makro yang belum memadai untuk analisis mendetail pada tingkat desa atau kecamatan, sehingga hasil zonasi belum sepenuhnya mendukung kebutuhan keputusan spesifik lokasi (Waongo et al., 2025).

Seiring perkembangan teknologi geospasial, peluang untuk mengatasi keterbatasan data dan metode dalam pemetaan agroekologi semakin terbuka. Platform pemrosesan geospasial berbasis *cloud* seperti Google Earth Engine (GEE) memungkinkan pengolahan variabel iklim dan tutupan lahan dalam skala besar secara berulang tanpa keterbatasan komputasi lokal. Selain itu, produk seperti SoilGrids menyediakan peta sifat tanah global, sementara produk reanalisis iklim seperti ERA5 menyediakan informasi iklim historis yang konsisten. Integrasi berbagai sumber data tersebut memungkinkan pembentukan dataset terpadu dengan cakupan spasial dan temporal yang lebih baik, sehingga berpotensi menghasilkan pemetaan geografis yang lebih akurat untuk analisis zonasi agroekologi (Zhao et al., 2021).

Ketersediaan data geospasial yang berjumlah ratusan ribu titik (*big data*) tersebut menuntut pemilihan metode analisis yang tangguh. Dalam analisis pengelompokan spasial, pendekatan seperti *DBSCAN* (Density-Based Spatial Clustering of Applications with Noise) atau *Hierarchical Clustering* sering digunakan. Namun, *DBSCAN* sangat sensitif terhadap variasi kepadatan data, di mana data agroekologi antar desa seringkali memiliki distribusi populasi yang heterogen, sehingga berisiko mengabaikan wilayah-wilayah perbatasan sebagai *noise*. Di sisi lain, *Hierarchical Clustering* memiliki kelemahan fatal pada skalabilitas komputasi ($O(n^3)$), sehingga sangat tidak efisien dan rentan mengalami kegagalan sistem (*out of memory*) ketika digunakan untuk memproses populasi data yang melebihi ratusan ribu titik spasial (Rizalde & Mubarak, 2024).

Oleh karena itu, algoritma K-Means clustering dipilih sebagai metode yang paling relevan dan superior untuk menjawab kebutuhan komputasi berskala besar ini. K-Means beroperasi dengan kompleksitas waktu yang linier ($O(n^3)$), sehingga sangat efisien, stabil, dan skalabel untuk memproses ratusan ribu kombinasi variabel iklim dan tanah yang telah dinormalisasi. Melalui inisialisasi iteratif, serta penentuan jumlah kluster secara sistematis menggunakan metrik evaluasi matematis yang ketat seperti *Elbow Method*, *Silhouette Score*, dan *Davies–Bouldin Index (DBI)*, K-Means mampu meminimalkan ambiguitas zonasi dan menghasilkan pengelompokan yang objektif serta terukur (Fang et al., 2025).

Meskipun peluang teknologi dan metode telah tersedia, pemetaan zonasi agroekologi berbasis data masih menghadapi beberapa *research gap* yang signifikan. Pertama, terdapat kesenjangan pada aspek data skala lokal, di mana berbagai sumber data tanah dan iklim telah tersedia secara global namun belum menyatu secara konsisten menjadi dataset zonasi yang representatif pada tingkat administrasi kecil seperti desa atau kecamatan. Kondisi ini menyebabkan hasil analisis sering kehilangan konteks spasial lokal yang penting untuk pengambilan keputusan pertanian secara presisi. Kedua, terdapat kesenjangan pada aspek metode yang sulit direplikasi dan diulang, karena banyak studi zonasi agroekologi sebelumnya masih menggunakan data yang tercatat atau alur kerja manual sehingga proses analisis sulit diulang, kurang penyesuaian terhadap pembaruan data, serta belum selalu memungkinkan evaluasi stabilitas kluster secara numerik. Ketiga, meskipun platform komputasi berbasis *cloud* seperti GEE telah berkembang pesat, pemanfaatannya masih terbatas pada tahap tertentu dan belum banyak diterapkan sebagai *pipeline* yang menggabungkan ekstraksi data, pemrosesan variabel, dan proses klasterisasi secara otomatis untuk menghasilkan peta zona agroekologi yang siap mendukung rekomendasi kebijakan pertanian spesifik lokasi. Berdasarkan hal tersebut, penelitian ini berupaya menjembatani ketiga *research gap* tersebut melalui sistem data geospasial dan penerapan metode klasterisasi yang sesuai untuk menghasilkan zonasi agroekologi yang lebih objektif (Pérez-cutillas et al., 2023).

Berpijak pada ketiga *research gap* tersebut, penelitian ini bertujuan untuk mengusulkan sebuah kerangka kerja komprehensif berlandaskan metodologi CRISP-DM. Solusi yang ditawarkan adalah membangun sistem *pipeline* yang

mengotomatisasi ekstraksi 12 variabel iklim (ERA5) dan sifat tanah (SoilGrids) secara berulang dari *Google Earth Engine*, memecahkan masalah kekosongan data melalui imputasi spasial berlandaskan geografi, dan mengelompokkannya secara matematis menggunakan algoritma K-Means $K=4$. Output dari komputasi ini selanjutnya tidak dibiarkan pasif, melainkan langsung didistribusikan (*deployment*) ke dalam bentuk *Dashboard Web* spasial interaktif terpusat. Dengan demikian, penelitian ini diharapkan dapat memberikan peta zonasi agroekologi yang terukur, lengkap dengan interpretasi agronomis di setiap wilayah, guna mendukung penetapan kebijakan dan pendistribusian bibit oleh Dinas Pertanian Kabupaten Bangkalan secara lebih cerdas dan presisi (Rahman et al., 2024).

1.2 Rumusan Masalah

Upaya optimalisasi potensi lahan pertanian di Kabupaten Bangkalan masih terhambat oleh minimnya pemetaan karakteristik lingkungan secara presisi. Kondisi iklim (mikro-klimat) dan kesuburan tanah yang sangat bervariasi antar wilayah belum dipetakan ke dalam zona-zona agroekologi yang spesifik dan terintegrasi secara digital. Ketiadaan pemetaan yang komprehensif ini mengakibatkan pengelolaan lahan termasuk penentuan komoditas tanam masih didominasi oleh asumsi subjektif yang berisiko memicu inefisiensi produksi. Penggunaan algoritma *K-Means Clustering* untuk memproses populasi data geospasial berskala besar dari *Google Earth Engine* (GEE) menawarkan solusi komputasi untuk memetakan Zona Agroekologi (ZAE) secara objektif. Namun, agar pemetaan ini berdampak nyata, efektivitas pengelompokan tersebut tidak hanya perlu divalidasi secara saintifik, tetapi hasil akhirnya juga harus didistribusikan ke dalam platform visualisasi spasial interaktif agar mudah diakses dan digunakan sebagai dasar pengambilan keputusan pertanian.

1.2.1 Permasalahan (*Problem Statement*)

Kabupaten Bangkalan belum memiliki pemetaan zonasi agroekologi terkomputerisasi yang mampu menganalisis kombinasi data iklim (suhu, curah hujan, kelembapan) dan karakteristik fisik tanah (pH, unsur hara, topografi) dalam skala mikro secara akurat. Kondisi ini mengakibatkan ketergantungan pada metode pengelolaan lahan tradisional yang subjektif. Oleh karena itu, penelitian ini

mengusulkan solusi berbasis *data mining* melalui implementasi algoritma *K-Means Clustering* di dalam kerangka kerja CRISP-DM. Masalah utama yang diselesaikan dalam penelitian ini adalah: bagaimana mengelompokkan wilayah berdasarkan homogenitas karakteristik lingkungannya, memvalidasi kualitas klaster secara matematis, serta merancang arsitektur pendistribusian hasil zonasi tersebut ke dalam sistem Peta Spasial Interaktif (*Web GIS / Dashboard*). Melalui sistem ini, profil agronomis dan rekomendasi komoditas pada masing-masing klaster dapat disajikan secara dinamis untuk menghasilkan keputusan pertanian yang lebih akurat dan terukur.

1.2.2 Metode Usulan

Untuk mengatasi permasalahan pemilihan tanaman berbasis kondisi agroekologi, penelitian ini mengembangkan solusi *data mining* dengan mengintegrasikan algoritma *K-Means* klasterisasi dalam kerangka *CRISP-DM*. Metode *unsupervised learning* menganalisis variabel lingkungan (suhu, curah hujan, kelembapan, pH tanah, dan kandungan hara) untuk mengelompokkan wilayah Kabupaten Bangkalan ke dalam zona agroekologi homogen. Pendekatan *CRISP-DM* dipilih karena menyediakan alur sistematis mulai dari pemahaman kebutuhan bisnis, preparasi data, pemodelan, evaluasi klaster, hingga implementasi solusi, sehingga menghasilkan rekomendasi tanaman berbasis data yang akurat dan siap diterapkan di lapangan.

1.2.3 Pertanyaan Penelitian

1. Bagaimana performa algoritma *K-Means* klasterisasi dalam membentuk zona agroekologi di Kabupaten Bangkalan jika diukur berdasarkan metrik evaluasi internal *Silhouette Coefficient* dan *Davies-Bouldin Index (DBI)*?
2. Bagaimana karakteristik geografis (cuaca dan tanah) pada masing-masing klaster yang terbentuk dan bagaimana profil tersebut berkorelasi secara agronomis?

1.3 Tujuan & Manfaat

Tujuan dan manfaat penelitian ini dijelaskan pada poin 1.3.1 dan 1.3.2 sebagai berikut.

1.3.1 Tujuan Penelitian

Penelitian ini memiliki tujuan yang disusun berdasarkan pertanyaan penelitian serta mengacu pada tahapan metodologi CRISP-DM, khususnya mulai dari tahap *modeling*, *evaluation*, hingga *deployment*. Tujuan penelitian ini dijabarkan sebagai berikut:

1. Menerapkan algoritma *K-Means Clustering* untuk memetakan zona agroekologi di Kabupaten Bangkalan berdasarkan kombinasi variabel cuaca dan tanah, serta mengevaluasi kualitas hasil klusterisasinya menggunakan metrik evaluasi matematis (*Silhouette Score* dan *Davies–Bouldin Index*) guna memperoleh jumlah zona (K) yang paling optimal dan stabil.
2. Mengidentifikasi karakteristik geografis dan memprofilkan kondisi biofisik pada setiap klaster yang terbentuk, serta menginterpretasikannya secara agronomis untuk menyusun peta zonasi agroekologi komprehensif yang dapat digunakan sebagai dasar rekomendasi komoditas pertanian.
3. Mengembangkan dan mengimplementasikan hasil klusterisasi ke dalam sistem visualisasi spasial berbasis Peta Interaktif (*Web GIS / Dashboard*) pada tahap *deployment*, agar informasi pemetaan zona agroekologi dapat diakses secara dinamis dan mempermudah pengambilan keputusan oleh pemangku kepentingan.

1.3.2 Manfaat Penelitian

Penelitian ini bertujuan pengolahan data cuaca dan jenis tanah secara terpadu, agar bisa digunakan sebagai acuan oleh petani, penyuluh, maupun pihak terkait dalam memilih tanaman yang cocok di setiap wilayah. Hasil akhir dari penelitian ini berupa peta zona agroekologi dan aplikasi berbasis web yang mudah diakses, yang diharapkan mampu menunjang kegiatan bercocok tanam dengan lebih efisien, mengurangi kemungkinan gagal panen, dan mendorong pemanfaatan teknologi dalam praktik pertanian yang berkelanjutan.

1.4 Batasan Penelitian

Untuk menjaga fokus dan kedalaman analisis dalam penelitian ini, maka batasan-batasan yang ditetapkan adalah sebagai berikut:

1. Wilayah studi terbatas pada Kabupaten Bangkalan, dengan *unit* analisis berbasis koordinat titik yang berada dalam batas administrasi kecamatan dan desa.

2. Data cuaca historis yang digunakan meliputi suhu rata-rata, curah hujan, kelembapan, kecepatan angin, dan durasi penyinaran matahari, dengan rentang tahun 2021 hingga 2025. Data diperoleh dari sumber terbuka di *Google Earth Engine*
3. Data tanah mencakup variabel pH, kandungan hara (N, P, K), kapasitas tukar kation (CEC), elevasi, kemiringan lereng, jenis tanah, dan kualitas tanah. Data bersumber dari dataset SoilGrids.
4. Sistem yang dikembangkan adalah prototipe berbasis web, yang memberikan hasil klasterisasi lahan berdasarkan zona agroekologi.
5. Evaluasi penelitian terbatas pada aspek kualitas klaster (misalnya dengan *Silhouette Score* atau *Davies-Bouldin Index*) serta interpretasi tiap klaster.

1.5 Sistematika Skripsi

Sistematika penulisan menggambarkan urutan bab yang menjadi dasar pembahasan dalam laporan skripsi, disusun secara ringkas dan terstruktur seperti berikut.

BAB I PENDAHULUAN

Bab ini membahas latar belakang penelitian yang menjelaskan dasar pemikiran dilakukannya studi tentang klasterisasi jenis tanaman berdasarkan cuaca dan tanah di Kabupaten Bangkalan. Selanjutnya, dirumuskan permasalahan utama yang ingin dipecahkan beserta metode yang digunakan untuk menjawabnya. Pada bagian ini juga dijelaskan tujuan penelitian yang ingin dicapai, manfaat yang diharapkan baik secara akademis maupun praktis, serta batasan penelitian yang menjelaskan ruang lingkup agar pembahasan lebih terarah. Di akhir bab, dijabarkan pula sistematika penulisan sebagai panduan struktur keseluruhan laporan skripsi.

BAB II TINJAUAN PUSTAKA

Bab ini memuat tinjauan teori yang menjadi landasan ilmiah bagi penelitian. Pembahasan mencakup konsep-konsep dasar seperti data mining, metodologi CRISP-DM (Cross Industry Standard Process for Data Mining), dan algoritma K-Means klasterisasi yang digunakan sebagai metode utama. Selain itu, dikaji pula teori mengenai rekomendasi tanaman, konsep agroekologi, serta zona agroekologi

yang relevan dengan konteks Bangkalan. Bab ini juga menyajikan hasil studi pustaka terhadap penelitian-penelitian terdahulu yang memiliki kemiripan tema, sebagai bahan perbandingan dan penguat posisi penelitian yang dilakukan.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan secara rinci tahapan penelitian yang digunakan untuk mencapai tujuan penelitian. Penjelasan mencakup kerangka kerja *CRISP-DM* yang terdiri dari tahap *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*. Selain itu, diuraikan pula mengenai sumber data, variabel yang digunakan, metode pengumpulan dan preprocessing data, serta implementasi algoritma K-Means dalam proses klasterisasi. Bab ini juga mencakup penentuan jumlah klaster optimal, teknik evaluasi model menggunakan metrik seperti *Silhouette Score* dan *Davies-Bouldin Index*, hingga tahap implementasi hasil ke dalam sistem berbasis web. Pada bagian akhir skenario uji coba untuk memastikan model dapat diterapkan sesuai tujuan penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab IV membahas hasil penerapan algoritma K-Means pada data agroekologi di Kabupaten Bangkalan menggunakan tahapan CRISP-DM. Pembahasan dimulai dari pemahaman tujuan penelitian, analisis dataset, hingga tahap persiapan data seperti seleksi variabel, transformasi, dan normalisasi. Selanjutnya dilakukan proses klasterisasi dengan penentuan jumlah klaster optimal menggunakan metode Elbow, *Silhouette Score*, dan *Davies-Bouldin Index*. Hasil klaster kemudian dievaluasi, divisualisasikan dalam bentuk grafik dan peta spasial, serta diinterpretasikan berdasarkan karakteristik cuaca dan tanah untuk membentuk zona agroekologi. Pada tahap akhir, hasil penelitian diimplementasikan dalam bentuk peta zonasi dan visualisasi berbasis web sebagai pendukung pengambilan keputusan pertanian yang lebih tepat.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari hasil penelitian yang telah dilakukan berdasarkan analisis dan pembahasan pada bab sebelumnya. Kesimpulan disusun untuk menjawab rumusan masalah serta tujuan penelitian secara ringkas dan jelas. Selain itu, pada bab ini juga disampaikan saran yang ditujukan untuk pengembangan penelitian selanjutnya, baik dari sisi metode, data, maupun

implementasi sistem, sehingga dapat memberikan perbaikan dan penyempurnaan terhadap hasil yang telah diperoleh.

DAFTAR PUSTAKA