

BAB III

METODOLOGI PENELITIAN

A. Metode Penelitian Kuantitatif

Menurut Sugiyono (2017) Metode penelitian kuantitatif dapat diartikan sebagai metode penelitian yang berdasarkan pada filsafat positivisme, digunakan untuk meneliti pada populasi atau sampel tertentu, pengumpulan data menggunakan instrumen penelitian, analisis data bersifat kuantitatif atau statistik, dengan tujuan untuk menguji hipotesis yang telah ditetapkan.

Penelitian ini menggunakan metode kuantitatif dengan pendekatan *Machine Learning*. Metode kuantitatif dipilih karena penelitian ini berfokus pada pengumpulan dan pengolahan data numerik yang dapat diukur secara statistik. Selain itu, metode ini memungkinkan peneliti untuk melakukan analisis secara objektif terhadap pola data yang diperoleh.

Pendekatan *Machine Learning* digunakan untuk memprediksi jenis tanaman yang optimal berdasarkan data cuaca dan kondisi tanah. Dalam pendekatan ini, digunakan algoritma *Decision Tree* yang termasuk dalam cabang *Supervised Learning*. *Supervised Learning* adalah pendekatan pembelajaran mesin di mana data pelatihan sudah memiliki label atau target yang diketahui, sehingga model dapat belajar dari data tersebut untuk membuat prediksi terhadap data baru yang belum diketahui.

Dengan menggunakan algoritma *Decision Tree*, penelitian ini bertujuan membangun model klasifikasi yang dapat merekomendasikan jenis tanaman yang sesuai berdasarkan input variabel seperti suhu, kelembaban, dan curah hujan. Langkah-langkah dalam penelitian ini adalah sebagai berikut:

a. Identifikasi Masalah

Pada tahap ini, peneliti mengamati kondisi pertanian di Indonesia, khususnya terkait minimnya penggunaan teknologi berbasis kecerdasan buatan. Permasalahan utama yang diangkat adalah bagaimana cara memberikan rekomendasi jenis tanaman yang optimal berdasarkan kondisi cuaca dan tanah, sehingga hasil pertanian dapat ditingkatkan

b. Pengumpulan Data

Peneliti mengumpulkan data yang relevan, seperti data cuaca (suhu, kelembaban udara, curah hujan) dan data tanah (kelembaban tanah, jenis tanah) dari sumber yang kredibel, seperti Observasi langsung, atau dataset publik.

c. Pra-pemrosesan Data

Data yang dikumpulkan kemungkinan besar memiliki nilai kosong, duplikat, atau format yang tidak konsisten. Oleh karena itu, dilakukan pembersihan data (data cleaning), normalisasi, dan transformasi data agar layak digunakan dalam pelatihan model. Tahapan ini penting untuk memastikan bahwa model dapat belajar dari data secara optimal dan akurat.

d. Penerapan Algoritma *Decision Tree*

Setelah data siap, algoritma *Decision Tree* diterapkan untuk membangun model klasifikasi. Model ini akan belajar dari data berlabel untuk mengenali pola dan hubungan antar variabel (fitur) sehingga dapat memprediksi jenis tanaman yang sesuai berdasarkan input data cuaca dan tanah.

e. Evaluasi Model

Model yang telah dibangun akan diuji akurasinya menggunakan data uji. Evaluasi dilakukan dengan menggunakan metrik seperti *accuracy*, *precision*, *recall*, dan *confusion matrix* untuk mengetahui sejauh mana model dapat melakukan prediksi dengan tepat. Jika performa model masih kurang, penyesuaian parameter atau metode *cross-validation* dapat dilakukan.

f. Interpretasi Hasil

Hasil evaluasi dianalisis untuk mengetahui bagaimana model bekerja dan seberapa akurat rekomendasi yang diberikan. Peneliti juga membahas fitur-fitur apa saja yang paling berpengaruh dalam proses klasifikasi. Hasil ini kemudian digunakan untuk memberikan rekomendasi tanaman yang optimal sesuai kondisi input.

B. Model yang Diusulkan

a. Deskripsi Model yang Dibangun

Model yang diusulkan dalam penelitian ini adalah sebuah sistem pendukung keputusan yang dirancang untuk memberikan rekomendasi jenis tanaman umbi berdasarkan kondisi lingkungan di Desa Cinangka. Sistem ini menggunakan

algoritma Decision Tree C4.5, yang mampu membentuk struktur pohon keputusan dari data lingkungan seperti suhu, curah hujan, kelembaban, dan intensitas cahaya. Algoritma ini efektif dalam mengklasifikasikan data ke dalam kelompok tertentu, sehingga dapat memberikan hasil prediksi berupa jenis tanaman yang paling sesuai untuk ditanam.

Sistem bekerja dengan cara membaca data lingkungan seperti suhu udara, curah hujan, kelembaban, intensitas cahaya matahari, serta parameter lainnya yang relevan terhadap pertumbuhan tanaman umbi. Berdasarkan kombinasi dari data tersebut, sistem akan menghasilkan rekomendasi jenis tanaman umbi yang optimal untuk ditanam, seperti talas, singkong, kumeli, atau bengkoang. Model klasifikasi ini akan dilatih menggunakan data observasi dan data simulasi yang telah dikategorikan berdasarkan karakteristik masing-masing tanaman. Setelah model dilatih, sistem mampu memproses data cuaca harian dan secara otomatis mengelompokkan kondisi tersebut ke dalam kelas tanaman tertentu menggunakan struktur pohon keputusan yang dihasilkan oleh algoritma C4.5.

Dengan pendekatan ini, diharapkan petani dapat memperoleh informasi yang lebih tepat dalam mengambil keputusan tanam berdasarkan data lingkungan yang aktual, serta mengurangi risiko kesalahan akibat pemilihan tanaman yang tidak sesuai dengan kondisi alam.

b. Input dan Output Sistem

Model klasifikasi yang dibangun dalam penelitian ini menggunakan delapan atribut lingkungan sebagai input, yaitu dapat dilihat pada Tabel 3.1 Tabel karakteristik

Tabel 3. 1 Tabel Karakteristik

No	Karakteristik Tanah dan Lingkungan
1	Suhu (°C)
2	pH Tanah
3	Curah Hujan (mm/tahun)
4	Ketinggian (m)
5	Jenis Tanah
6	Cahaya (jam/hari)
7	Kelembaban (%)

Semua atribut tersebut merupakan variabel yang diperoleh melalui observasi langsung serta sumber literatur, dan digunakan untuk menentukan jenis tanaman umbi yang paling cocok ditanam pada kondisi tersebut.

Adapun output dari sistem adalah rekomendasi jenis tanaman umbi yang sesuai dengan kombinasi nilai dari setiap atribut input. Dalam penelitian ini, terdapat empat kelas output yang digunakan, yaitu dapat dilihat pada Tabel 3.2 Tabel Umbi.

Tabel 3. 2 Tabel Umbi

No	Jenis Umbi-umbian
1	Talas
2	Singkong
3	Bengkuang
4	Kumeli

Masing-masing baris data mencerminkan satu kondisi lingkungan lengkap dan label tanaman umbi yang paling cocok ditanam pada kondisi tersebut. Data ini digunakan sebagai dasar untuk pelatihan dan pengujian model Decision Tree C4.5 yang diimplementasikan dalam penelitian.

c. Alur Kerja Model

Alur kerja dari model klasifikasi yang dibangun dalam penelitian ini terdiri dari beberapa tahap yang saling berkaitan. Setiap tahap memiliki fungsi penting dalam memproses data dari awal hingga menghasilkan output berupa rekomendasi jenis tanaman umbi.

1. Pengumpulan Data

Data dikumpulkan melalui metode observasi langsung dan studi pustaka. Data ini mencakup informasi lingkungan seperti suhu udara, curah hujan, kelembaban, pH tanah, jenis tanah, ketinggian lahan, dan intensitas cahaya matahari. Data yang diperoleh kemudian disusun dalam bentuk dataset untuk proses selanjutnya.

2. Pra-pemrosesan Data (Preprocessing)

Tahap ini bertujuan untuk memastikan bahwa data bersih dan siap digunakan. Pada tahap ini dilakukan normalisasi nilai, pengecekan data duplikat, serta pengelompokan jenis tanaman sebagai label atau kelas. Data juga diuji untuk memastikan format dan rentang nilainya sesuai dengan kebutuhan algoritma.

3. Pelatihan Model (Training)

Dataset yang telah diproses dibagi menjadi data latih dan data uji menggunakan metode train-test split. Data latih digunakan untuk membentuk struktur pohon keputusan menggunakan algoritma Decision Tree C4.5. Pada tahap ini sistem belajar mengenali pola dari kombinasi atribut lingkungan yang mengarah pada tanaman tertentu.

4. Pengujian Model (Testing)

Setelah model terbentuk, data uji digunakan untuk menguji seberapa baik model dapat memprediksi tanaman yang sesuai berdasarkan data baru yang belum pernah dilihat sebelumnya. Hasil pengujian kemudian dievaluasi menggunakan confusion matrix dan metrik seperti akurasi, presisi, recall, dan F1-score.

5. Rekomendasi Output

Berdasarkan struktur pohon yang terbentuk, sistem akan memberikan output berupa nama tanaman umbi yang paling cocok untuk ditanam pada kondisi lingkungan tertentu. Output ini diharapkan dapat digunakan sebagai referensi oleh petani dalam mengambil keputusan tanam yang lebih tepat dan efisien.

d. Penjelasan Algoritma Decision Tree C4.5

1. Penjelasan Umum Cara Kerja C4.5

Algoritma Decision Tree C4.5 merupakan salah satu metode klasifikasi yang banyak digunakan dalam pengambilan keputusan berbasis data. C4.5 bekerja dengan membentuk struktur pohon keputusan, di mana setiap cabang mewakili suatu kondisi atau tes terhadap atribut, dan setiap daun menunjukkan hasil keputusan atau klasifikasi akhir.

Secara umum, proses kerja algoritma C4.5 dimulai dengan menganalisis seluruh atribut dalam dataset untuk menentukan atribut mana yang paling berpengaruh terhadap hasil klasifikasi. Atribut terbaik ini dipilih sebagai akar pohon dengan menggunakan nilai gain ratio, yaitu hasil dari perbandingan antara gain informasi dan tingkat keragaman data (split information).

Setelah atribut terbaik ditentukan, data dibagi menjadi subset sesuai dengan nilai atribut tersebut. Proses pemilihan atribut terbaik ini diulang untuk setiap subset data hingga tidak ada lagi atribut yang bisa dipilih, atau hingga semua data dalam subset termasuk ke dalam satu kelas yang sama. Hasil akhirnya berupa struktur pohon yang berisi logika keputusan yang dapat digunakan untuk mengklasifikasikan data baru.

C4.5 juga memiliki kemampuan untuk menangani atribut numerik dengan cara membuat batas nilai tertentu (threshold), serta dapat menangani data yang memiliki nilai kosong (missing values), yang membuatnya lebih fleksibel dibandingkan metode decision tree sebelumnya.

2. Entropy dan Gain pada Decision Tree C4.5

Dalam algoritma Decision Tree C4.5, pemilihan atribut terbaik untuk dijadikan akar pohon atau percabangan dilakukan berdasarkan perhitungan entropy, information gain, dan gain ratio. Ketiga komponen ini merupakan dasar dari cara kerja C4.5 dalam mengukur seberapa baik sebuah atribut dalam memisahkan data ke dalam kelas-kelas tertentu.

Entropy digunakan untuk mengukur tingkat ketidakpastian atau impuritas dalam sebuah dataset. Semakin tinggi nilai entropy, semakin tidak murni (heterogen) data tersebut, dan semakin besar usaha yang dibutuhkan untuk memisahkannya.

$$\text{Entropy}(S) = - \sum_{i=1}^n p_i \log_2(p_i)$$

Gambar 3. 1 Rumus Entropy

Keterangan:

S = himpunan data

Pi = proporsi data pada kelas ke-i

Setelah entropy diketahui, maka dilakukan perhitungan information gain, yaitu pengurangan entropy awal dengan rata-rata entropy setelah data dipisah berdasarkan atribut tertentu.

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \cdot \text{Entropy}(S_v)$$

Gambar 3. 2 Information gain

Keterangan:

A = atribut

Sv = subset data pada nilai v dari atribut A

Namun, C4.5 tidak hanya menggunakan gain saja, melainkan gain ratio untuk menghindari pemilihan atribut yang memiliki terlalu banyak nilai unik (overfitting). Gain ratio adalah rasio antara information gain dan split information.

$$\text{Gain Ratio}(S, A) = \frac{\text{Gain}(S, A)}{\text{SplitInfo}(A)}$$

Gambar 3. 3 Rumus gain ratio

Atribut dengan nilai gain ratio tertinggi akan dipilih sebagai node pada pohon keputusan. Proses ini diulang secara rekursif untuk membentuk struktur pohon yang dapat mengklasifikasikan data dengan akurasi tinggi.

3. Keunggulan Decision Tree C4.5 untuk Penerapan di Bidang Pertanian

Pemilihan algoritma Decision Tree C4.5 dalam penelitian ini didasarkan pada beberapa keunggulan yang menjadikannya sangat relevan untuk kasus klasifikasi kondisi lingkungan terhadap jenis tanaman umbi yang optimal. Salah satu keunggulan utama C4.5 adalah kemampuannya dalam menangani atribut numerik dan kategorik sekaligus. Hal ini sangat penting dalam konteks pertanian, di mana data seperti suhu, kelembaban, curah hujan, dan pH tanah umumnya berbentuk angka (numerik), sedangkan atribut seperti jenis tanah bersifat kategorikal.

Selain itu, C4.5 juga mampu membentuk batasan nilai numerik otomatis (thresholding) saat membangun pohon keputusan, misalnya dalam bentuk aturan seperti “suhu > 28°C”. Fitur ini membuat model lebih fleksibel dan mudah diinterpretasikan oleh pengguna akhir, seperti petani atau penyuluh pertanian. Struktur pohon yang dihasilkan bersifat eksplisit dan transparan, sehingga keputusan yang diambil oleh sistem dapat dijelaskan secara logis.

Keunggulan lainnya adalah kemampuannya dalam menangani data yang tidak lengkap atau memiliki missing values, serta kestabilannya terhadap data yang memiliki jumlah atribut yang besar. Dalam penelitian ini, C4.5 dipandang sebagai algoritma yang paling sesuai karena mampu mengolah data lingkungan pertanian secara efektif dan menghasilkan rekomendasi tanaman umbi yang praktis dan mudah dipahami oleh pengguna.

C. Prodeur Penelitian

Penelitian ini dilakukan melalui beberapa tahapan sistematis yang dirancang untuk menghasilkan model klasifikasi berbasis algoritma Decision Tree C4.5. Adapun prosedur pelaksanaan penelitian adalah sebagai berikut:

1. Identifikasi Masalah dan Tujuan Penelitian

Peneliti melakukan studi awal untuk memahami permasalahan yang dihadapi petani di Desa Cinangka, khususnya dalam hal pemilihan jenis tanaman umbi yang sesuai dengan kondisi lingkungan. Berdasarkan studi tersebut, ditentukan tujuan penelitian, yaitu membangun sistem rekomendasi tanaman berbasis kondisi cuaca menggunakan algoritma C4.5.

2. Pengumpulan Data

- a) Observasi langsung, dilakukan di lapangan untuk mencatat kondisi lingkungan seperti suhu udara, kelembaban, curah hujan, intensitas cahaya, pH tanah, dan jenis tanah.
- b) Studi pustaka, digunakan untuk mengumpulkan informasi tentang karakteristik pertumbuhan masing-masing tanaman umbi (Talas, Singkong, Bengkoang, Kumeli), serta menentukan label tanaman berdasarkan kondisi idealnya.

3. Penyusunan Dataset Simulasi

Berdasarkan hasil observasi dan studi pustaka, disusun dataset berisi kombinasi kondisi lingkungan dan tanaman umbi yang sesuai. Dataset ini disusun dalam format tabel dengan kolom atribut sebagai input dan label tanaman sebagai output.

4. Pra-pemrosesan Data (Preprocessing)

Data yang telah dikumpulkan dilakukan proses pembersihan (data cleaning), konversi ke format yang sesuai (misalnya kategorisasi), dan normalisasi nilai jika diperlukan. Langkah ini bertujuan untuk memastikan bahwa data dapat diproses secara optimal oleh algoritma.

5. Pelatihan dan Pengujian Model

Dataset yang telah siap dibagi menjadi dua bagian, yaitu data latih dan data uji. Data latih digunakan untuk membentuk pohon keputusan menggunakan algoritma C4.5. Selanjutnya, model diuji dengan data uji untuk mengukur performanya menggunakan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score.

6. Analisis dan Penarikan Kesimpulan

Hasil pengujian dan performa model dianalisis untuk mengetahui tingkat keberhasilan sistem dalam memberikan rekomendasi tanaman yang sesuai.

Berdasarkan hasil tersebut, disusun kesimpulan dan saran untuk pengembangan sistem di masa depan..

D. Pengujian Model

Pengujian model dalam penelitian ini dilakukan untuk mengevaluasi performa algoritma Decision Tree C4.5 dalam memberikan rekomendasi jenis tanaman umbi berdasarkan kondisi lingkungan. Pengujian dilakukan dengan menggunakan data uji yang sebelumnya dipisahkan dari dataset utama melalui teknik train-test split, yaitu membagi dataset menjadi dua bagian: data latih untuk pelatihan model dan data uji untuk menguji akurasi prediksi.

Setelah model dilatih, dilakukan pengujian terhadap data uji dengan melihat hasil klasifikasi yang dihasilkan oleh model, lalu dibandingkan dengan label sebenarnya. Evaluasi performa model dilakukan menggunakan metode confusion matrix, yaitu matriks yang menggambarkan hasil klasifikasi dalam empat kategori: True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Berdasarkan confusion matrix tersebut, dihitung beberapa metrik evaluasi, yaitu :

- 1) Akurasi (*Accuracy*) mengukur persentase prediksi benar secara keseluruhan terhadap total data uji. Metrik ini dipilih untuk menilai performa umum model dalam mengklasifikasikan data cuaca dan rekomendasi tanaman.

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN}$$

Gambar 3. 4 Metrik Evaluasi (*Accuracy*)

TP (True Positive): Prediksi benar kelas positif.

TNTN (True Negative): Prediksi benar kelas negatif.

FPFP (False Positive): Prediksi salah kelas positif.

FNFN (False Negative): Prediksi salah kelas negatif.

Akurasi digunakan untuk menilai performa model secara keseluruhan. Nilai akurasi tinggi menunjukkan bahwa model dapat memprediksi dengan benar sebagian besar data uji.

- 2) Selanjutnya Presisi (*Precision*) menilai ketepatan model dalam merekomendasikan tanaman tertentu. Nilai presisi tinggi menunjukkan

minimnya rekomendasi tanaman yang tidak sesuai. Presisi (*Precision*) dipilih untuk meminimalkan kesalahan rekomendasi tanaman yang tidak sesuai, sementara recall memastikan tidak ada kondisi cuaca kritis yang terlewat.

$$\text{Presisi} = \frac{TP}{TP + FP}$$

Gambar 3. 5 Metrik Evaluasi (*Precision*)

- 3) *Recall (Sensitivity)* mengukur kemampuan model mendeteksi seluruh instance aktual dari suatu kelas. Recall tinggi penting untuk memastikan tidak ada kondisi cuaca kritis yang terlewat.

$$\text{Recall} = \frac{TP}{TP + FN}$$

Gambar 3. 6 Metrik Evaluasi (*Sensitivity*)

- 4) *F1-Score* merupakan rata-rata harmonik dari presisi dan recall. Metrik ini berguna ketika kelas data tidak seimbang.

$$\text{F1-Score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}}$$

Gambar 3. 7 Metrik Evaluasi (*F1-Score*)

Penggunaan keempat metrik ini memberikan gambaran yang lebih menyeluruh terhadap kinerja model, terutama dalam hal ketepatan dan kelengkapan hasil klasifikasi yang dihasilkan oleh algoritma Decision Tree C4.5