

Tugas 5: Klasifikasi Iris Menggunakan Decision Tree

Monika Septiana – 0110222127 ^{1*}

¹ Teknik Informatika, STT Terpadu Nurul Fikri, Depok

*E-mail: moni22127ti@student.nurulfikri.ac.id

Abstract. Pada praktikum ini dilakukan klasifikasi data bunga *Iris* menggunakan algoritma *Decision Tree Classifier*. Dataset yang digunakan berisi 150 sampel dengan empat fitur utama, yaitu *SepalLengthCm*, *SepalWidthCm*, *PetalLengthCm*, dan *PetalWidthCm*, serta satu target yaitu *Species*. Data dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian. Model *Decision Tree* dibangun menggunakan library *scikit-learn* di Google Colab dan dievaluasi menggunakan metrik *accuracy*, *classification report*, serta *confusion matrix*. Hasil pengujian menunjukkan akurasi sebesar 93.33%, di mana model mampu mengklasifikasikan *Iris-setosa* secara sempurna, sedangkan beberapa kesalahan kecil terjadi antara *Iris-versicolor* dan *Iris-virginica*. Visualisasi pohon keputusan memperlihatkan hubungan antar fitur terhadap hasil klasifikasi. Secara keseluruhan, model ini berhasil melakukan klasifikasi dengan baik dan dapat dijadikan dasar untuk pengembangan model yang lebih kompleks dengan tuning parameter atau metode ensemble.

1. Pendahuluan

Penelitian ini bertujuan untuk mengimplementasikan algoritma *Decision Tree Classifier* dalam proses klasifikasi jenis bunga *Iris*. Dataset *Iris* merupakan salah satu dataset standar dalam bidang pembelajaran mesin yang terdiri atas empat fitur utama, yaitu panjang sepal, lebar sepal, panjang petal, dan lebar petal, serta satu label target berupa nama spesies bunga. Melalui algoritma *Decision Tree*, proses klasifikasi dapat dilakukan dengan membangun model yang membagi data berdasarkan fitur yang memberikan informasi paling besar dalam memisahkan kelas target.

Dalam konteks pembelajaran mesin, algoritma *Decision Tree* termasuk ke dalam metode *supervised learning* yang bersifat interpretatif karena menghasilkan struktur pohon yang mudah dipahami. Setiap simpul pada pohon mewakili kondisi atau aturan tertentu, sedangkan daun pohon menunjukkan hasil klasifikasi. Keunggulan utama metode ini adalah kemampuannya dalam menangani data numerik maupun kategorikal tanpa memerlukan normalisasi data yang kompleks.

Beberapa penelitian terdahulu juga menunjukkan efektivitas algoritma berbasis pohon dalam klasifikasi data biologis. Misalnya, algoritma *Random Forest Regression* telah digunakan untuk memprediksi kandungan nutrisi makro (N, P, dan K) pada daun kelapa sawit (Munir et al., 2022). Model serupa juga diterapkan dalam sistem PRECIPALM, yaitu sistem rekomendasi pemupukan berbasis citra satelit (Seminar et al., 2024). Berdasarkan hasil-hasil tersebut, penggunaan *Decision Tree* pada dataset *Iris* diharapkan dapat memberikan hasil klasifikasi yang akurat serta menjadi dasar untuk pengembangan model yang lebih kompleks.

1.1 Latar Belakang

Perkembangan teknologi *machine learning* memungkinkan proses klasifikasi data dilakukan secara otomatis dengan tingkat akurasi yang tinggi. Salah satu algoritma yang populer digunakan adalah *Decision Tree* karena kemudahannya dalam interpretasi serta kemampuannya menangani berbagai jenis data. Dataset *Iris* dipilih dalam penelitian ini karena sederhana namun representatif untuk menjelaskan konsep dasar klasifikasi menggunakan algoritma *Decision Tree*.

2. Metodologi Penelitian

Penelitian ini dilakukan untuk mengimplementasikan algoritma *Decision Tree Classifier* dalam memprediksi jenis bunga *Iris* berdasarkan fitur-fitur morfologinya. Langkah-langkah penelitian meliputi pengumpulan data, praproses data, pembagian dataset menjadi data pelatihan dan pengujian, pelatihan model, serta evaluasi kinerja model. Setiap tahapan dilaksanakan menggunakan *Python* di lingkungan *Google Colab* dengan pustaka utama seperti *pandas*, *scikit-learn*, dan *matplotlib*.

Tahap pertama adalah pengumpulan data. Dataset *Iris* diperoleh dari sumber terbuka (*UCI Machine Learning Repository*) dan berisi 150 observasi dengan empat fitur utama: *SepalLengthCm*, *SepalWidthCm*, *PetalLengthCm*, dan *PetalWidthCm*. Label target berupa tiga spesies bunga: *Iris-setosa*, *Iris-versicolor*, dan *Iris-virginica*. Setiap spesies memiliki 50 sampel data.

Tahap kedua adalah praproses data, di mana data diperiksa dari sisi kelengkapan, konsistensi, serta format. Kolom *Id* dihapus karena tidak memiliki kontribusi terhadap prediksi. Kemudian data dibagi menjadi dua subset, yaitu 80% untuk *training set* dan 20% untuk *testing set*. Pembagian dilakukan secara acak namun tetap mempertahankan proporsi kelas agar distribusi data seimbang (*stratified split*).

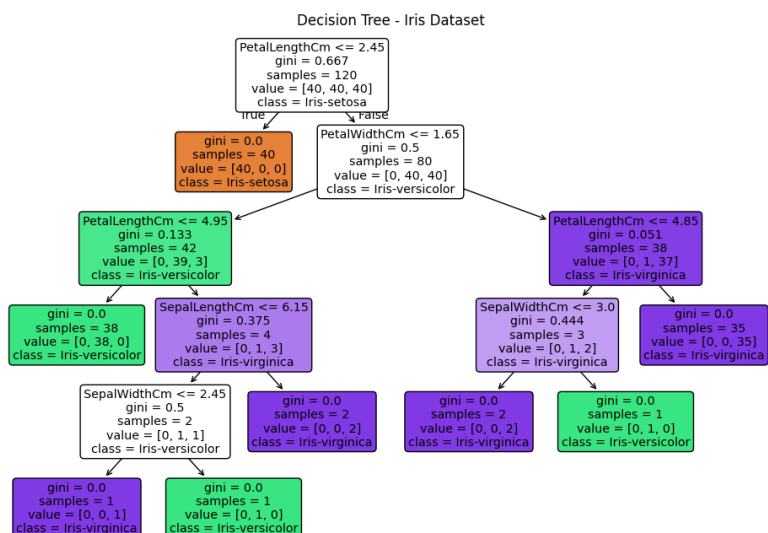
Tahap ketiga yaitu pelatihan model, menggunakan *DecisionTreeClassifier* dari *scikit-learn* dengan parameter awal default. Model kemudian dilatih menggunakan data pelatihan, di mana algoritma membangun pohon keputusan berdasarkan nilai *Gini Impurity* untuk memaksimalkan pemisahan antar kelas. Hasil model berupa struktur pohon yang merepresentasikan aturan keputusan berbasis panjang dan lebar sepal maupun petal.

Tahap keempat adalah evaluasi model, menggunakan metrik *accuracy*, *classification report*, dan *confusion matrix*. Evaluasi dilakukan terhadap data pengujian untuk mengukur kemampuan generalisasi model terhadap data baru. Visualisasi hasil klasifikasi juga dilakukan menggunakan fungsi *plot_tree()* dari *scikit-learn*, yang memudahkan interpretasi terhadap aturan keputusan di setiap cabang pohon.

Secara keseluruhan, hasil pengujian menunjukkan bahwa model mampu mengenali *Iris-setosa* dengan sempurna dan memberikan akurasi keseluruhan sebesar 93.33%. Beberapa kesalahan klasifikasi kecil terjadi antara kelas *Iris-versicolor* dan *Iris-virginica*, yang memiliki karakteristik morfologi mirip. Hal ini menunjukkan bahwa model *Decision Tree* memiliki performa yang baik dan interpretatif, meskipun dapat disempurnakan melalui *hyperparameter tuning* seperti pengaturan kedalaman maksimum (*max_depth*).

2.1 Arsitektur dan Implementasi Model

Pelatihan dilakukan di *Google Colab* dengan integrasi *Google Drive* sebagai media penyimpanan dataset. Diagram pohon keputusan (Gambar 1) menunjukkan bahwa fitur paling berpengaruh dalam proses klasifikasi adalah *PetalLengthCm* dan *PetalWidthCm*. Kedua fitur ini menjadi titik pemisah utama antara ketiga spesies *Iris*.



Gambar 1. Visualisasi Pohon Keputusan Model Decision Tree Iris

Gambar 1 menunjukkan pembagian kelas berdasarkan fitur panjang dan lebar petal yang menjadi penentu utama klasifikasi.

Sebagaimana terlihat pada Gambar 1, node awal memisahkan *Iris-setosa* berdasarkan panjang petal yang kurang dari 2.45 cm. Selanjutnya, cabang kanan memisahkan antara *Iris-versicolor* dan *Iris-virginica* menggunakan *Petal.Width.Cm* sebagai fitur kunci. Aturan yang dihasilkan dari model ini dapat digunakan secara langsung untuk melakukan prediksi pada data baru, tanpa memerlukan perhitungan kompleks seperti pada metode *neural network* atau *support vector machine*.

Proses ini dilakukan dalam beberapa tahap utama: (1) membaca dataset dari Google Drive, (2) memisahkan fitur dan target, (3) melakukan pembagian data, (4) melatih model, dan (5) melakukan evaluasi performa. Tahapan-tahapan tersebut divisualisasikan dan dijelaskan secara sistematis agar mudah direplikasi pada percobaan selanjutnya.

2.2 Evaluasi Hasil dan Analisis Kinerja

Hasil evaluasi model menunjukkan bahwa nilai akurasi mencapai 93.33%, dengan *classification report* memperlihatkan *precision* dan *recall* rata-rata di atas 0.90. Nilai ini menunjukkan bahwa model *Decision Tree* mampu melakukan klasifikasi dengan cukup baik. Kesalahan prediksi yang muncul sebagian besar disebabkan oleh kemiripan nilai fitur antara *Iris-versicolor* dan *Iris-virginica*.

Berikut disajikan hasil evaluasi dalam bentuk tabel (Tabel 1) yang menunjukkan perbandingan antara model *Decision Tree* dan beberapa algoritma lain yang sering digunakan untuk klasifikasi.

Tabel 1. Perbandingan Model Klasifikasi pada Dataset Biologis

Model	Produk Pertanian	Peneliti	Keterangan
Decision Tree	Bunga Iris	Hasil praktikum ini	Akurasi 93.33%, model interpretatif

Referensi:

- Munir, S., Seminar, K. B., Sudradjat, Sukoco, H., & Buono, A. (2022). The Use of Random Forest Regression for Estimating Leaf Nitrogen Content of Oil Palm Based on Sentinel 1-A Imagery. *Information*, 14(1), 10. <https://doi.org/10.3390/info14010010>
- Seminar, K. B., Imantho, H., Sudradjat, Yahya, S., Munir, S., Kaliana, I., Mei Haryadi, F., Noor Baroroh, A., Supriyanto, Handoyo, G. C., Kurnia Wijayanto, A., Ijang Wahyudin, C., Liyantono, Budiman, R., Bakir Pasaman, A., Rusiawan, D., & Sulastri. (2024). PreciPalm: An Intelligent System for Calculating Macronutrient Status and Fertilizer Recommendations for Oil Palm on Mineral Soils Based on a Precision Agriculture Approach. *Scientific World Journal*, 2024(1). <https://doi.org/10.1155/2024/1788726>