

LAPORAN UAS MACHINE LEARNING



Nama : Azacky Habibilah Syahlan
Nim : 231011402531
Kelas : 05TPLE004

Program Studi Teknik Informatika
Fakultas Ilmu Komputer
Universitas Pamulang

Jl. Raya Puspiptek No. 11, Serpong, Kota Tangerang Selatan, Banten 15316

Pendahuluan

Machine Learning merupakan salah satu cabang kecerdasan buatan yang memungkinkan sistem komputer untuk belajar dari data dan membuat prediksi atau keputusan tanpa diprogram secara eksplisit. Salah satu algoritma yang sering digunakan dalam machine learning adalah **Decision Tree**, yang bekerja dengan membentuk struktur pohon keputusan berdasarkan atribut data.

Pada laporan ini dilakukan implementasi algoritma Decision Tree untuk studi kasus **klasifikasi dataset Iris**. Dataset Iris dipilih karena memiliki struktur data yang sederhana, bersih, serta sering digunakan sebagai dataset standar dalam pembelajaran machine learning. Tujuan dari penelitian ini adalah untuk membangun model klasifikasi menggunakan Decision Tree serta mengevaluasi performa model tersebut.

Teori Singkat

Decision Tree adalah algoritma supervised learning yang digunakan untuk tugas klasifikasi dan regresi. Algoritma ini bekerja dengan membagi data berdasarkan atribut tertentu sehingga membentuk struktur pohon keputusan. Setiap node merepresentasikan pengujian atribut, sedangkan leaf node merepresentasikan hasil prediksi.

Beberapa konsep penting dalam Decision Tree meliputi:

- **Node:** Titik pada pohon yang merepresentasikan keputusan atau kondisi.
- **Root Node:** Node paling atas yang menjadi awal proses pengambilan keputusan.
- **Leaf Node:** Node terakhir yang berisi hasil klasifikasi.
- **Splitting:** Proses pembagian data berdasarkan nilai atribut tertentu.
- **Pruning:** Proses pemangkasan cabang pohon untuk mengurangi overfitting.

Decision Tree memiliki kelebihan berupa kemudahan interpretasi dan tidak memerlukan normalisasi data, namun memiliki kelemahan yaitu rentan terhadap overfitting jika tidak dibatasi dengan baik.

Metodologi

Metodologi yang digunakan dalam penelitian ini terdiri dari beberapa tahapan sebagai berikut:

1. Load dan Eksplorasi Dataset

Dataset Iris dimuat menggunakan bahasa pemrograman Python dengan bantuan library *pandas*. Eksplorasi data dilakukan untuk melihat struktur data, tipe data, serta memastikan tidak terdapat missing value.

2. Preprocessing Data

Tahap preprocessing dilakukan dengan menghapus kolom **Id** karena tidak relevan terhadap proses klasifikasi. Selanjutnya data dipisahkan menjadi fitur (X) dan label (y). Label kelas *Species* kemudian diubah dari bentuk teks menjadi numerik menggunakan *LabelEncoder*.

3. Pembagian Data

Dataset dibagi menjadi data latih dan data uji dengan perbandingan 80% data latih dan 20% data uji menggunakan fungsi *train_test_split*.

4. Pembangunan Model

Model dibangun menggunakan algoritma **Decision Tree Classifier** dengan parameter *criterion = gini* dan *max_depth = 3*.

5. Evaluasi Model

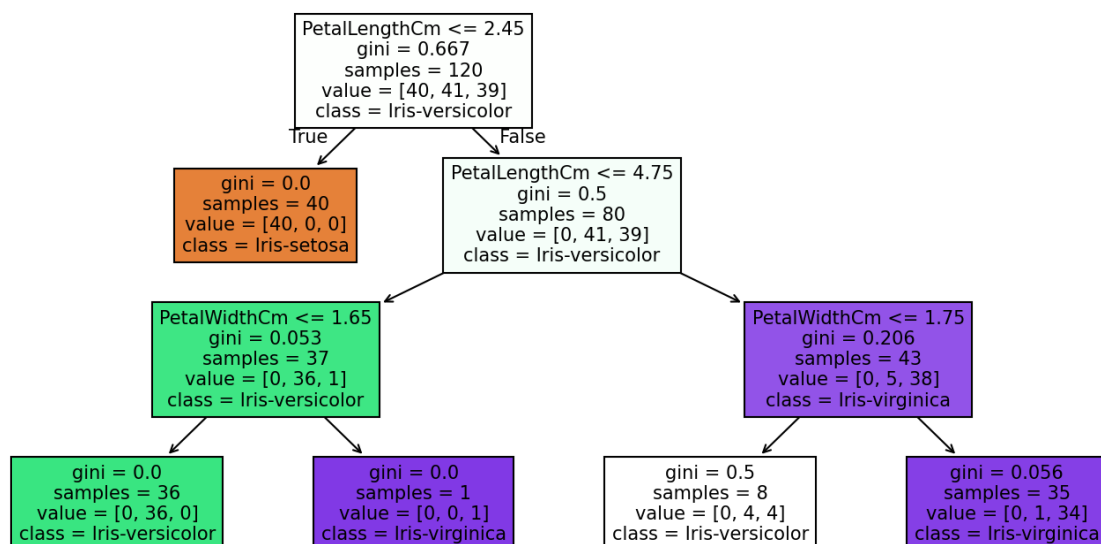
Evaluasi dilakukan menggunakan metrik **Accuracy, Precision, Recall, dan F1-score** untuk mengetahui performa model pada data uji.

6. Visualisasi Pohon Keputusan

Struktur pohon keputusan divisualisasikan untuk memudahkan interpretasi hasil klasifikasi.

Hasil dan Analisis

Berdasarkan hasil eksperimen yang telah dilakukan, model Decision Tree menghasilkan nilai **accuracy sebesar 1.0 atau 100%** pada data uji. Hal ini menunjukkan bahwa seluruh data uji berhasil diklasifikasikan dengan benar.



Hasil *classification report* menunjukkan bahwa nilai **precision, recall, dan F1-score** untuk setiap kelas (**Iris-setosa**, **Iris-versicolor**, dan **Iris-virginica**) masing-masing bernilai 1.00. Hal ini menandakan bahwa model mampu melakukan klasifikasi dengan sangat baik tanpa kesalahan prediksi.

Visualisasi pohon keputusan menunjukkan bahwa atribut **PetalLengthCm** dan **PetalWidthCm** menjadi faktor utama dalam pemisahan kelas. Beberapa leaf node memiliki nilai *Gini* = 0, yang menandakan bahwa data pada node tersebut bersifat homogen. Pembatasan kedalaman pohon dengan *max_depth* = 3 membuat model tetap sederhana dan mudah diinterpretasikan.

Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma **Decision Tree** sangat efektif digunakan untuk klasifikasi dataset Iris. Model yang dibangun mampu menghasilkan performa yang sangat baik dengan akurasi sempurna serta struktur pohon keputusan yang sederhana dan mudah dipahami. Dengan karakteristik dataset

yang bersih dan pemisahan kelas yang jelas, Decision Tree menjadi pilihan yang tepat untuk studi kasus klasifikasi data serupa.