

Nama: Rafi Muhammad Luthfi  
05TPLE013

Link Repository : <https://github.com/gitluthfi/uas-kecerdasan-buatan>

## Pendahuluan

Kecerdasan Buatan (Artificial Intelligence) merupakan cabang ilmu komputer yang berfokus pada pengembangan sistem yang mampu meniru kecerdasan manusia. Salah satu pendekatan yang umum digunakan dalam machine learning adalah tree-based methods, khususnya Decision Tree, yang banyak dipakai untuk tugas klasifikasi dan regresi karena sifatnya yang mudah dipahami dan diinterpretasikan.

## Teori Singkat

### 1. Pengertian Decision Tree

Decision Tree adalah algoritma supervised learning yang digunakan untuk klasifikasi maupun regresi. Model ini bekerja dengan membagi data ke dalam beberapa cabang berdasarkan fitur tertentu sehingga membentuk struktur seperti pohon keputusan. Setiap keputusan dibuat berdasarkan aturan (if-then) yang dihasilkan dari data latih.

### 2. Konsep dalam Decision Tree

- Node  
Bagian dari pohon yang merepresentasikan sebuah fitur atau keputusan tertentu.
- Root Node  
Node paling atas pada pohon yang menjadi awal proses pengambilan keputusan.
- Leaf Node  
Node akhir yang tidak memiliki cabang lagi dan berisi hasil prediksi (kelas atau nilai).
- Splitting

Proses membagi data pada suatu node berdasarkan nilai fitur tertentu untuk meningkatkan homogenitas data.

- Pruning  
Teknik untuk memangkas cabang pohon yang tidak signifikan agar mengurangi overfitting dan meningkatkan generalisasi model.

### 3. Perbedaan Decision Tree, Random Forest, dan Gradient Boosting

Metode	Penjelasan
Decision Tree	Satu pohon keputusan, mudah dipahami namun rentan overfitting
Random Forest	Ensemble dari banyak decision tree dengan teknik bagging untuk meningkatkan stabilitas
Gradient Boosting	Ensemble berbasis boosting, membangun model secara bertahap untuk memperbaiki kesalahan sebelumnya



### 4. Kelebihan dan Kekurangan Tree-Based Methods

Kelebihan:

- Mudah dipahami dan diinterpretasikan
- Tidak memerlukan normalisasi data
- Dapat menangani data numerik dan kategorikal

Kekurangan:

- Rentan terhadap overfitting (khususnya decision tree tunggal)
- Performa kurang optimal dibanding metode ensemble
- Sensitif terhadap perubahan kecil pada data

## Metodologi

Di tugas ini saya menggunakan dataset iris, yaitu dataset klasik dalam klasifikasi yang berisi data morfologi bunga iris dengan tiga kelas berbeda.

Tujuan dari studi ini adalah membangun dan menganalisis model Decision Tree

untuk melakukan klasifikasi jenis bunga iris menggunakan python dengan library sklearn, pandas, dan matplotlib

Dataset Iris terdiri dari 150 data dengan 4 fitur numerik, diantaranya

- Sepal Length
- Sepal Width
- Petal Length
- Petal Width

Dengan target kelas ada 3

- Iris-setosa
- Iris-versicolor
- Iris-virginica

## Hasil dan Analisis

```
sierra@IO-228N-RAFI1LUTF1:/mnt/c/Users/rafi.lutfi.AW000MAIN/Documents/sierra6/kuliah/artificial_intelegence/uas$ python3.10 uas.py
```

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	target
0	5.1	3.5	1.4	0.2	0
1	4.9	3.8	1.4	0.2	0
2	4.7	3.2	1.3	0.2	0
3	4.6	3.1	1.5	0.2	0
4	5.0	3.6	1.4	0.2	0

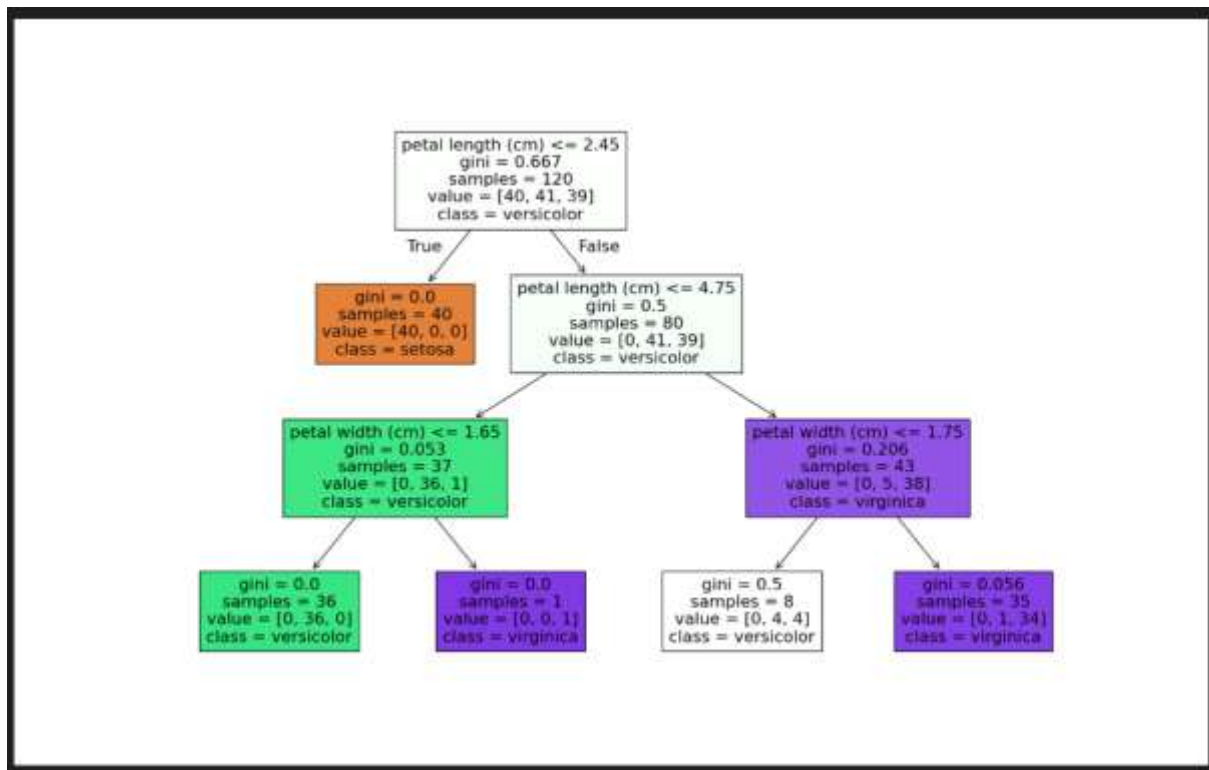
  

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	target
count	150.000000	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.057333	3.758000	1.199333	1.000000
std	0.828066	0.435866	1.765298	0.762238	0.819232
min	4.300000	2.000000	1.000000	0.100000	0.000000
25%	5.100000	2.800000	1.600000	0.300000	0.000000
50%	5.800000	3.000000	4.350000	1.300000	1.000000
75%	6.400000	3.300000	5.100000	1.800000	2.000000
max	7.900000	4.400000	6.900000	2.500000	2.000000

Accuracy: 1.0

Classification Report:

	precision	recall	f1-score	support
setosa	1.00	1.00	1.00	10
versicolor	1.00	1.00	1.00	9
virginica	1.00	1.00	1.00	11
accuracy			1.00	30
macro avg	1.00	1.00	1.00	30
weighted avg	1.00	1.00	1.00	30



## 1. Hasil Eksplorasi Data

Dataset Iris yang digunakan terdiri dari 150 data dengan 4 fitur numerik, yaitu sepal length, sepal width, petal length, dan petal width, serta satu variabel target yang merepresentasikan tiga kelas bunga: Iris-setosa, Iris-versicolor, dan Iris-virginica.

Berdasarkan statistik deskriptif, seluruh fitur memiliki rentang nilai yang jelas dan tidak ditemukan missing value. Hal ini menunjukkan bahwa dataset berada dalam kondisi bersih dan siap digunakan tanpa proses imputasi data. Distribusi kelas target juga relatif seimbang, sehingga tidak diperlukan teknik penanganan data tidak seimbang (imbalanced data).

## 2. Hasil Pelatihan dan Evaluasi Model

Model Decision Tree dilatih menggunakan 80% data training dan diuji menggunakan 20% data testing. Parameter utama yang digunakan adalah:

- criterion: gini
- max\_depth: 3

Hasil evaluasi model menunjukkan performa yang sangat baik, dengan metrik sebagai berikut:

- Accuracy: 1.00 (100%)
- Precision: 1.00 pada seluruh kelas
- Recall: 1.00 pada seluruh kelas
- F1-score: 1.00 pada seluruh kelas

Classification report memperlihatkan bahwa semua data uji berhasil diklasifikasikan dengan benar tanpa kesalahan prediksi. Tidak terdapat false positive maupun false negative pada setiap kelas.

### 3. Analisis Performa Model

Performa sempurna yang dihasilkan oleh model Decision Tree dipengaruhi oleh beberapa faktor utama:

1. Dataset Iris memiliki separabilitas kelas yang tinggi, terutama pada fitur petal length dan petal width, sehingga proses pemisahan kelas menjadi sangat efektif.
2. Jumlah fitur yang sedikit dan seluruhnya numerik memudahkan algoritma dalam menentukan titik pemisahan (splitting) yang optimal.
3. Pembatasan kedalaman pohon (max\_depth = 3) membantu model menghindari overfitting meskipun tetap mampu mencapai akurasi maksimum.
4. Tidak adanya noise dan missing value membuat model mampu belajar pola data dengan stabil dan konsisten.

Berdasarkan visualisasi pohon keputusan, fitur petal length dan petal width menjadi faktor dominan dalam pengambilan keputusan, yang sesuai dengan karakteristik biologis bunga Iris.

Meskipun model menghasilkan akurasi 100%, perlu dicatat bahwa:

- Dataset Iris merupakan dataset sederhana dan terstruktur dengan baik
- Hasil serupa belum tentu dapat dicapai pada dataset dunia nyata yang lebih kompleks
- Decision Tree tunggal tetap berpotensi mengalami overfitting jika digunakan tanpa pembatasan parameter

Namun demikian, untuk studi kasus akademik dan pembelajaran konsep machine learning, hasil ini menunjukkan keunggulan Decision Tree dalam hal interpretabilitas dan efektivitas.

## Kesimpulan

Berdasarkan hasil eksperimen dan analisis yang telah dilakukan, dapat disimpulkan bahwa:

1. Algoritma Decision Tree mampu melakukan klasifikasi dataset Iris dengan performa yang sangat tinggi, ditunjukkan oleh nilai akurasi, precision, recall, dan F1-score sebesar 1.00.
2. Parameter max\_depth berperan penting dalam menjaga keseimbangan antara kompleksitas model dan kemampuan generalisasi.
3. Tree-based methods sangat cocok digunakan pada dataset yang memiliki pola pemisahan kelas yang jelas dan jumlah fitur yang terbatas.
4. Keunggulan utama Decision Tree terletak pada kemudahan interpretasi dan visualisasi aturan keputusan, sehingga sangat sesuai untuk analisis berbasis pemahaman.
5. Untuk dataset yang lebih besar dan kompleks, metode ensemble seperti Random Forest atau Gradient Boosting dapat menjadi alternatif untuk meningkatkan performa dan stabilitas model.

Dengan demikian, Decision Tree terbukti efektif sebagai model klasifikasi pada dataset Iris dan menjadi pendekatan yang relevan dalam pembelajaran kecerdasan buatan.