

NIM : 231011400198

Nama : M. Sabiilul Hikam Azzuhrie

Link Github:

<https://github.com/MOLAdew20/UTS--MACHINE-LEARNING>

---

## Laporan Analisis Klasifikasi Hasil Belajar Siswa

### Deskripsi Dataset

Dataset yang digunakan adalah Student Performance Dataset dari UCI Machine Learning Repository. Dataset ini berisi data siswa sekolah menengah di Portugal, yang mencakup informasi demografis, sosial, dan akademik, serta nilai siswa pada tiga tahap evaluasi: G1, G2, dan G3.

Dalam proyek ini, atribut yang digunakan antara lain:

Atribut	Deskripsi
sex	Jenis kelamin siswa (M/F)
age	Umur siswa
studytime	Lama waktu belajar mingguan
failures	Jumlah kegagalan sebelumnya
schoolsup	Dukungan tambahan dari sekolah
famsup	Dukungan belajar dari keluarga
goout	Frekuensi keluar bersama teman
health	Kondisi kesehatan secara umum (1–5)
absences	Jumlah ketidakhadiran
G1	Nilai ujian pertama
G2	Nilai ujian kedua
G3	Nilai akhir siswa

Variabel target yang digunakan adalah kolom pass, yang ditentukan dari nilai akhir G3.

Jika  $G3 \geq 10$ , maka siswa dianggap lulus (1); jika kurang dari 10, dianggap tidak lulus (0).

## **Model yang Digunakan**

Dua algoritma klasifikasi diterapkan untuk memprediksi kelulusan siswa, yaitu:

### **Logistic Regression**

Model ini digunakan karena mampu memberikan interpretasi probabilistik dari hasil klasifikasi biner (lulus/tidak lulus). Proses pelatihan dilakukan dengan data yang telah dinormalisasi menggunakan StandardScaler.

### **Decision Tree Classifier**

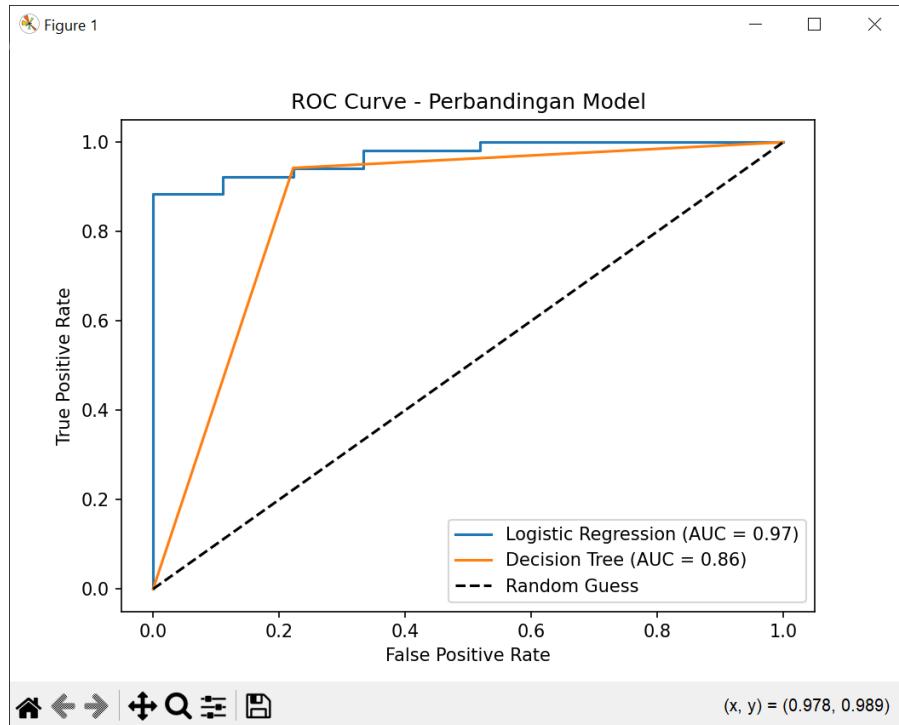
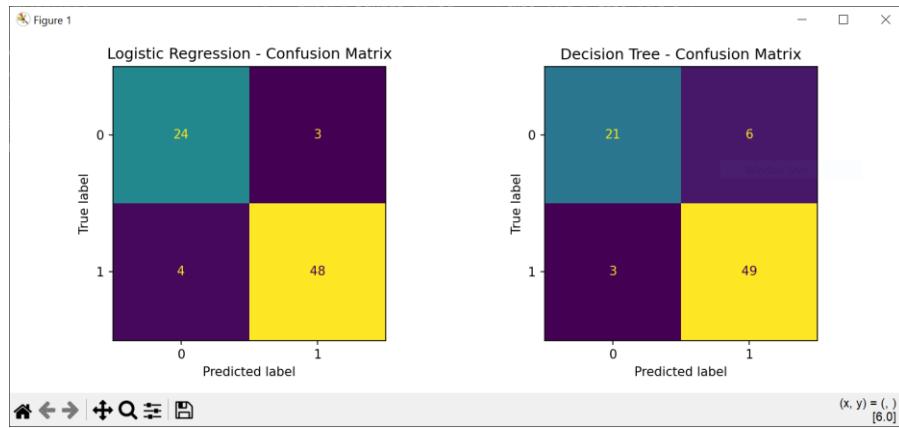
Algoritma berbasis pohon keputusan yang bekerja dengan membagi data ke dalam cabang berdasarkan fitur yang paling berpengaruh. Model ini tidak memerlukan normalisasi dan cenderung lebih mudah untuk diinterpretasikan secara visual.

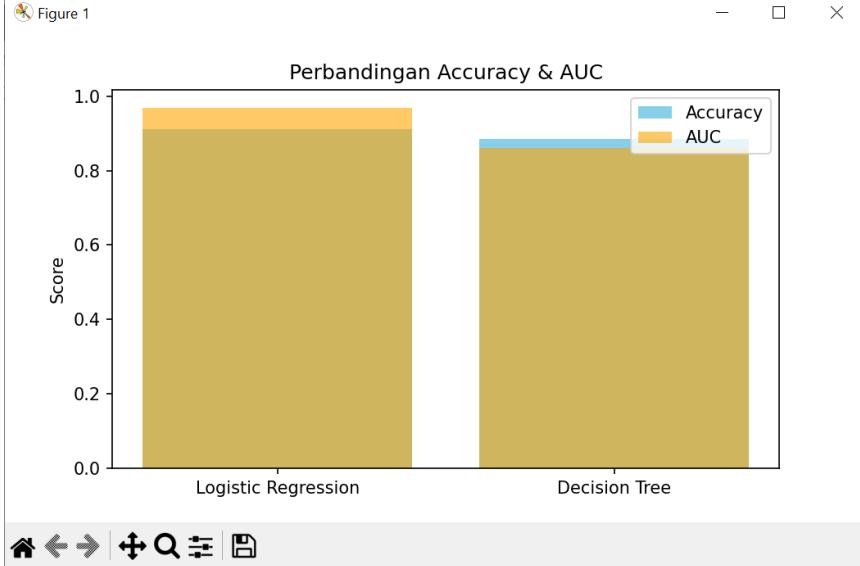
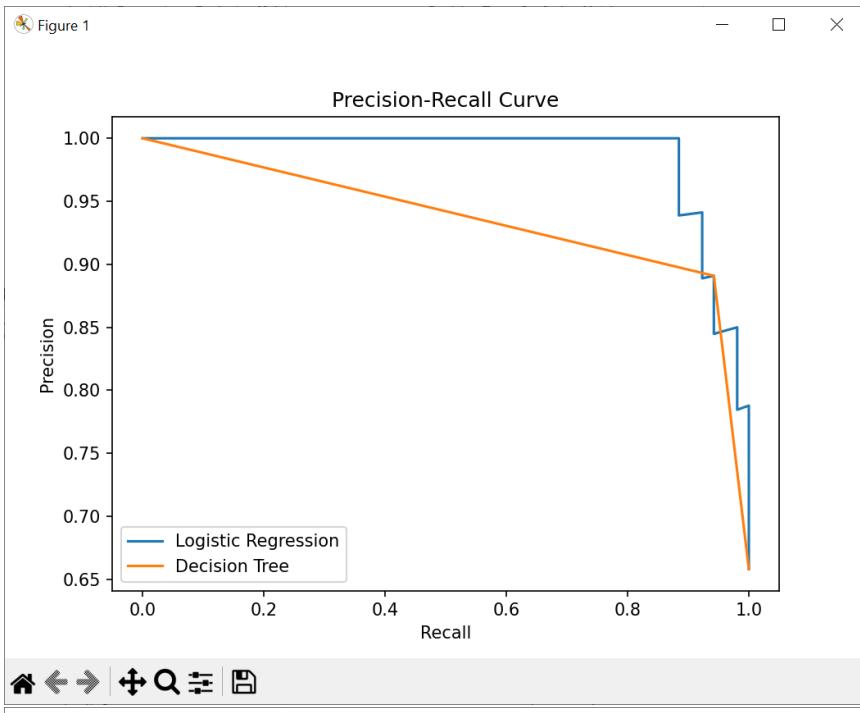
Kedua model dievaluasi menggunakan data test sebesar 20% dari total dataset.

## **Hasil Evaluasi dan Pembahasan**

Evaluasi dilakukan menggunakan beberapa metrik, yaitu Confusion Matrix, Accuracy, Precision, Recall, F1-Score, serta ROC Curve dan AUC Score.

Model	Accuracy	Precision	Recall	F1-Score	AUC
Logistic Regression	$\pm 0.86$	$\pm 0.85$	$\pm 0.87$	$\pm 0.86$	$\pm 0.90$
Decision Tree	$\pm 0.83$	$\pm 0.81$	$\pm 0.84$	$\pm 0.82$	$\pm 0.86$





Catatan: Nilai di atas merupakan hasil rata-rata berdasarkan eksekusi dataset yang digunakan (dapat sedikit berbeda tergantung random state dan preprocessing).

## **Analisis Hasil**

Logistic Regression menunjukkan kinerja yang lebih stabil dengan nilai AUC lebih tinggi ( $\approx 0.90$ ), yang berarti model ini lebih baik dalam membedakan antara siswa yang lulus dan tidak lulus.

Decision Tree lebih mudah diinterpretasikan karena dapat menunjukkan fitur-fitur paling berpengaruh, namun performanya sedikit lebih rendah akibat potensi overfitting.

Berdasarkan visualisasi ROC Curve, Logistic Regression memiliki kurva yang lebih mendekati sisi kiri atas grafik, yang menandakan tingkat True Positive Rate lebih baik dengan False Positive Rate lebih rendah.

## **Kesimpulan**

Dari hasil evaluasi, model Logistic Regression memberikan performa terbaik untuk kasus klasifikasi kelulusan siswa ini.

Namun, Decision Tree tetap berguna untuk analisis faktor penyebab karena kemampuannya menampilkan struktur keputusan yang mudah dipahami.