

TUGAS UTS MACHINE LEARNING

NAMA : NURDIANSYAH

NIM : 231011400211

KELAS : 05TPLE005

1. Deskripsi dataset

Titanic dari Kaggle.

Deskripsi Dataset Titanic (Kaggle):

Dataset ini berisi data penumpang kapal **RMS Titanic** yang tenggelam tahun 1912. Tujuannya adalah **memprediksi apakah penumpang selamat atau tidak** berdasarkan fitur seperti **jenis kelamin, usia, dan kelas tiket**. Dataset memiliki **891 baris** dan **12 kolom**, dengan variabel target **Survived (0 = tidak selamat, 1 = selamat)**.

2. Model yang digunakan:

Model yang digunakan Adalah Logistic Regression dan Decision Tree

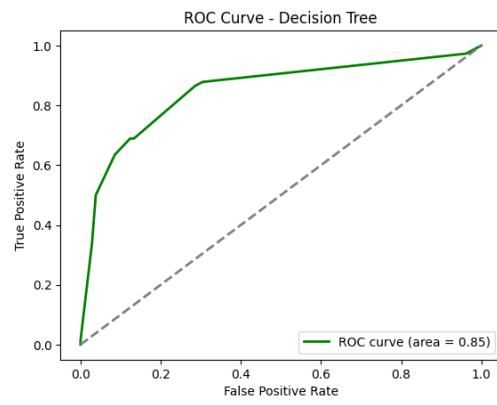
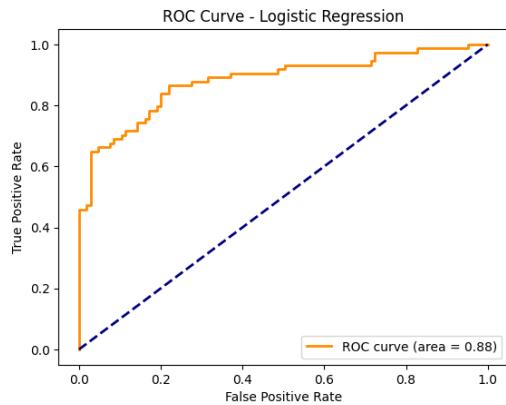
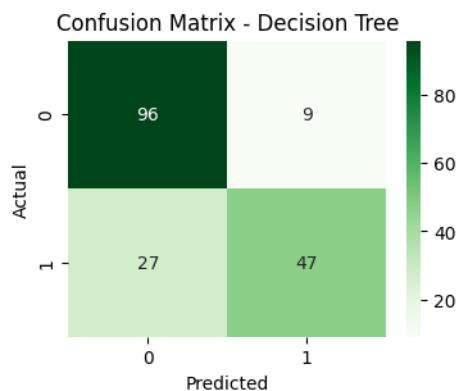
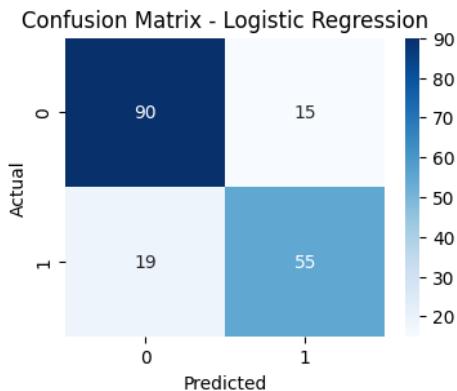
② Logistic Regression

Model ini digunakan untuk **klasifikasi biner**, memprediksi kemungkinan seseorang **selamat (1)** atau **tidak selamat (0)**. Logistic Regression bekerja dengan menghitung **probabilitas** berdasarkan hubungan linier antara fitur input dan output.

③ Decision Tree

Model ini membuat keputusan dengan membagi data ke dalam **cabang-cabang berdasarkan kondisi tertentu** (misalnya jenis kelamin, usia, kelas tiket). Decision Tree mampu menangkap **pola non-linear** dan mudah dipahami melalui bentuk visual pohon keputusan.

3. Hasil evaluasi dan pembahasan



Perbandingan Hasil Model

Model	Akurasi	ROC AUC	Ciri Utama
Logistic Regression	0.81	0.88	Linear, lebih stabil terhadap noise dan generalisasi baik
Decision Tree	0.80	0.85	Non-linear, mudah menjelaskan pola, tapi sedikit lebih rawan overfitting

Analisis Kinerja Model

🧠 Logistic Regression

- Accuracy: 81%
- Precision (kelas 1 / selamat): 0.79

- Recall (kelas 1 / selamat): 0.74
- F1-score: 0.76
- Model ini cukup seimbang dalam mengenali penumpang yang selamat dan tidak selamat.
- Nilai ROC AUC = 0.88 menunjukkan kemampuan diskriminatif yang sangat baik (semakin mendekati 1, semakin bagus).

Decision Tree

- Accuracy: 79.9%
- Precision (kelas 1 / selamat): 0.84
- Recall (kelas 1 / selamat): 0.64
- F1-score: 0.72
- Model ini lebih tepat dalam memprediksi yang selamat (precision tinggi), tapi lebih banyak meleset dalam mendekripsi semua yang selamat (recall rendah).
- Nilai ROC AUC = 0.85 juga tergolong baik, meski sedikit di bawah Logistic Regression.

Kesimpulan

- Kedua model memiliki performa yang mirip (selisih $\pm 1\%$), tetapi Logistic Regression unggul sedikit pada hampir semua metrik.
- Logistic Regression lebih stabil, generalisasi lebih baik, dan cocok untuk dataset dengan variabel numerik & kategorikal sederhana seperti Titanic.
- Decision Tree tetap bermanfaat karena memberikan interpretasi yang lebih mudah (misalnya, kita bisa melihat aturan keputusan seperti “penumpang perempuan kelas 1 memiliki peluang selamat lebih tinggi”).
- Untuk implementasi nyata, Logistic Regression lebih direkomendasikan jika tujuan utama adalah akurasi dan generalisasi, sedangkan Decision Tree cocok jika fokusnya pada interpretabilitas.
- **Rangkuman Akhir**
- Model terbaik untuk kasus Titanic ini adalah Logistic Regression karena memberikan akurasi (81%) dan ROC AUC (0.88) tertinggi, serta performa yang seimbang antara precision dan recall.