

Tugas UTS Machine Learning

Nama : Egidius Edi Putrawan Halawa

Nim : 231011403453

Kelas : 05TPLE005

1. Deskripsi Dataset

Titanic dataset

dataset Titanic dalam format **CSV (Comma Separated Values)**, yang biasa digunakan dalam analisis data dan pembelajaran mesin.

Berikut penjelasan singkat tiap kolom:

- **PassengerId** → Nomor identitas penumpang unik.
- **Survived** → Status selamat (1 = selamat, 0 = tidak selamat).
- **Pclass** → Kelas tiket (1 = kelas satu, 2 = kelas dua, 3 = kelas tiga).
- **Lname, Name** → Nama penumpang.
- **Sex** → Jenis kelamin (male/female).
- **Age** → Usia penumpang.
- **SibSp** → Jumlah saudara atau pasangan yang ikut di kapal.
- **Parch** → Jumlah orang tua atau anak yang ikut.
- **Ticket** → Nomor tiket.
- **Fare** → Harga tiket.
- **Cabin** → Nomor kabin tempat tinggal di kapal.
- **Embarked** → Pelabuhan keberangkatan (C = Cherbourg, Q = Queenstown, S = Southampton).

Dataset ini sering digunakan untuk latihan **analisis data dan model prediksi** (misalnya memprediksi siapa yang selamat dalam tragedi Titanic).

2. Deskripsi Model Algoritma klasifikasi,: a. Logistic Regression b. Decision Tree

a. Logistic Regression

Logistic Regression adalah metode **klasifikasi** yang digunakan untuk **memprediksi dua kemungkinan** (misalnya: selamat atau tidak).

Model ini menggunakan **persamaan matematika** untuk menghitung peluang suatu data termasuk ke dalam suatu kelas, kemudian hasilnya dinyatakan dalam bentuk **probabilitas (0–1)**.

Contoh penggunaan: memprediksi apakah penumpang Titanic selamat (**1**) atau **tidak selamat (0)** berdasarkan usia, jenis kelamin, dan kelas.

b. Decision Tree

Decision Tree adalah model yang berbentuk **pohon keputusan**.

Data dianalisis dengan membuat pertanyaan berurutan, seperti:

- Apakah usia < 15?
- Jika ya, lanjut ke keputusan berikut.
- Jika tidak, ke cabang yang lain.

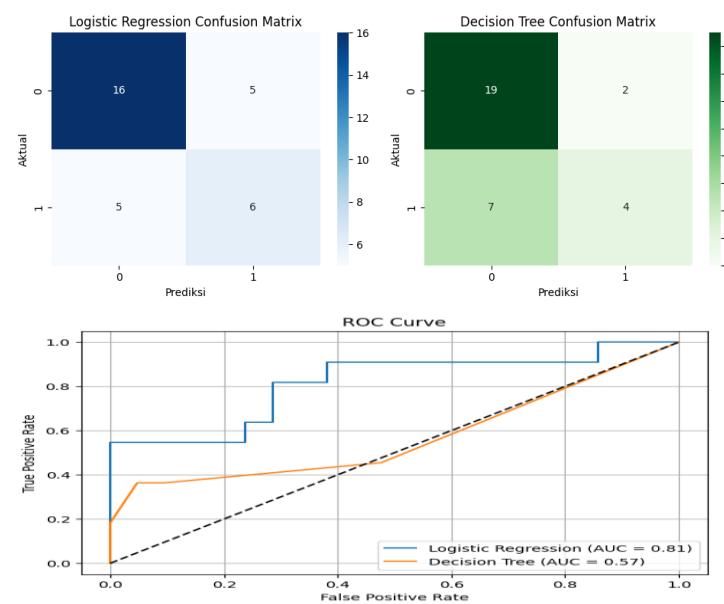
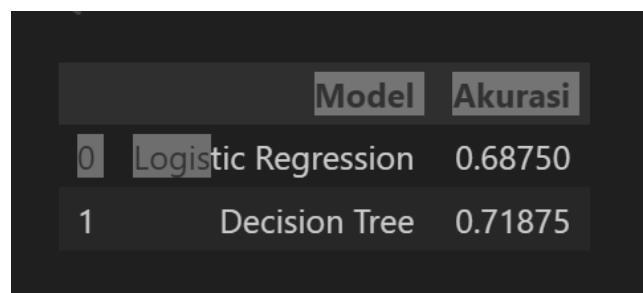
Model ini mudah dipahami karena bekerja mirip seperti **logika manusia** dalam mengambil keputusan.

Contoh penggunaan: menentukan apakah penumpang selamat berdasarkan kondisi langkah demi langkah (umur, jenis kelamin, kelas, dll).

Singkatnya:

- **Logistic Regression** → Menghitung **peluang** dan menghasilkan **kelas (0/1)**.
- **Decision Tree** → Mengambil keputusan melalui **percabangan aturan** seperti pohon.

3. Hasil evaluasi dan pembahasan



Metrik	Logistic Regression	Decision Tree
Accuracy	0.6875	0.7188
Precision	0.5455	0.6667
Recall	0.5455	0.3636
F1-Score	0.5455	0.4706
AUC	0.81	0.57

Interpretasi Logistic Regression :

- Logistic Regression memiliki performa cukup seimbang antara prediksi benar untuk kelas 0 (tidak selamat) dan kelas 1 (selamat).
- Nilai AUC 0.81 menunjukkan model ini cukup baik membedakan antara penumpang yang selamat dan tidak.
- Namun, nilai accuracy 68.75% masih menunjukkan beberapa kesalahan prediksi (terutama false negatives & false positives yang cukup banyak di confusion matrix).

Interpretasi Decision Tree:

- Model ini memiliki akurasi sedikit lebih tinggi (71.88%) dibanding Logistic Regression.
- Namun, recall rendah (0.36) artinya banyak kasus “selamat” yang tidak berhasil dikenali (False Negative tinggi).
- AUC 0.57 menunjukkan kemampuan pemisahan kelasnya lemah, mendekati prediksi acak.

Kesimpulan Akhir

- Logistic Regression lebih direkomendasikan untuk kasus Titanic ini, karena meskipun akurasinya sedikit lebih rendah, model ini memiliki kemampuan generalisasi yang lebih baik (AUC tinggi dan metrik seimbang).
- Decision Tree cocok jika ingin interpretasi yang lebih mudah, tapi hasilnya kurang stabil dan kurang baik dalam mendeteksi kelas minoritas (selamat).