

1. Dataset

Dataset yang digunakan pada praktikum ini merupakan dataset titanic dari seaborn, yang dapat diunduh atau di export melalui notebook yang sudah tercantumkan pada folder yang sama. Dataset Titanic berisi data penumpang kapal RMS Titanic yang tenggelam pada tahun 1912.

2. Berikut merupakan skenario model untuk dataset titanic/

a. Melakukan Identifikasi Teknik Pemodelan

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	embark_town	alive	alone
2	0	3	male	22	1	0	7.25	S	Third	man	TRUE		Southampton	no	FALSE
3	1	1	female	38	1	0	71.2833	C	First	woman	FALSE	C	Cherbourg	yes	FALSE
4	1	3	female	26	0	0	7.925	S	Third	woman	FALSE		Southampton	yes	TRUE
5	1	1	female	35	1	0	53.1	S	First	woman	FALSE	C	Southampton	yes	FALSE
6	0	3	male	35	0	0	8.05	S	Third	man	TRUE		Southampton	no	TRUE
7	0	3	male		0	0	8.4583	Q	Third	man	TRUE		Queenstown	no	TRUE
8	0	1	male	54	0	0	51.8625	S	First	man	TRUE	E	Southampton	no	TRUE
9	0	3	male	2	3	1	21.075	S	Third	child	FALSE		Southampton	no	FALSE
10	1	3	female	27	0	2	11.1333	S	Third	woman	FALSE		Southampton	yes	FALSE

Pada dataset ini terdapat fitur-fitur sebagai berikut:

- Pclass : Kelas tiket penumpang (1, 2, 3).
- Sex : Jenis kelamin penumpang.
- Age : Usia penumpang.
- SibSp : Jumlah saudara/ pasangan di kapal.
- Parch : Jumlah orang tua/anak di kapal.
- Fare : Biaya tiket.
- Embarked : Pelabuhan embarkasi (C = Cherbourg, Q = Queenstown, S = Southampton).
- Target (Survived) : Apakah penumpang selamat (1) atau tidak (0).

Dataset relatif bersih, walau ada beberapa missing value seperti di kolom Age atau Deck. Terdapat fitur campuran numerik (Age, Fare, SibSp, Parch) dan kategorikal (Sex, Embarked). Kemudian juga terdapat fitur target berupa variabel biner (Survived: 0/1). Berdasarkan dataset diatas teknik pemodelan yang cocok yaitu decision tree atau logistic regression. Decision tree cocok digunakan karena dataset mengandung data numerik dan kategorikal, sedangkan logistic regression cocok digunakan karena cocok untuk klasifikasi biner dan interpretasi sederhana.

b. Menentukan Teknik Pemodelan yang Sesuai

Pada kasus kali ini decision tree dipilih untuk percobaan awal, karena dapat menangani fitur campuran numerik dan kategorikal, mudah divisualisasikan, dan menghasilkan interpretasi sederhana. Decision tree dapat menghasilkan

aturan klasifikasi untuk memprediksi peluang selamat berdasarkan atribut seperti kelas tiket, jenis kelamin, dan usia.

c. Menyiapkan Skenario Pengujian

- Skenario Pengujian yang digunakan yaitu 5-fold cross-validation. Dataset dibagi menjadi 5 subset, tiap subset jadi data uji secara bergantian.
- Metrik Evaluasi:
 - Akurasi → Persentase prediksi benar.
 - Precision → Seberapa tepat prediksi “selamat” itu benar.
 - Recall → Seberapa banyak korban selamat yang benar-benar terdeteksi.
 - F1-Score → Harmonis antara precision dan recall.

d. Kesimpulan

```
Akurasi tiap fold: [0.81118881 0.7972028 0.77464789 0.81690141 0.78169014]
Rata-rata Akurasi: 0.7963262090022654
```

Tabel Evaluasi Model:

	Fold	Accuracy	Precision	Recall	F1-Score
0	1.0	0.811189	0.969697	0.551724	0.703297
1	2.0	0.797203	0.837209	0.620690	0.712871
2	3.0	0.774648	0.711864	0.736842	0.724138
3	4.0	0.816901	0.918919	0.596491	0.723404
4	5.0	0.781690	0.764706	0.672414	0.715596
Rata-rata	3.0	0.796326	0.840479	0.635632	0.715861

Decision Tree berhasil digunakan untuk memprediksi keselamatan penumpang Titanic dengan performa yang cukup baik. Rata-rata akurasi model berada di kisaran 79%, dengan akurasi tiap fold berkisar antara 77% – 81% dan nilai precision dan recall yang relatif seimbang sehingga model tidak terlalu bias terhadap salah satu kelas. Dari hasil visualisasi pohon keputusan, terlihat bahwa faktor-faktor seperti kelas tiket (pclass), jenis kelamin (sex), dan usia (age) berperan penting dalam menentukan kemungkinan penumpang selamat. Model ini dapat menjadi dasar analisis lebih lanjut untuk memahami pola keselamatan penumpang pada kasus Titanic dan juga dapat dijadikan acuan dalam penerapan model klasifikasi pada dataset sejenis.