

**Nama : Ahmad Wildan H**

**Nim : 231011401765**

**Kelas : 05 TPLE 004**

## Laporan Analisis Klasifikasi Dataset Flights

### 1. Deskripsi Dataset

Dataset flights dari library seaborn berisi data jumlah penumpang pesawat per bulan dari tahun **1949 hingga 1960** (12 tahun  $\times$  12 bulan = **144 sampel**).

**Fitur Utama:**

- **year**: Tahun penerbangan (1949–1960)
- **month**: Bulan penerbangan (Januari–Desember)
- **passengers**: Jumlah penumpang (kontinu)

**Target Klasifikasi:**

- **high\_demand**:
  - **1** (permintaan tinggi) jika passengers > **median (215.5)**
  - **0** (permintaan rendah) jika passengers  $\leq$  **median**

**Hasil EDA:**

- **Trend Tahunan**: Jumlah penumpang meningkat secara konsisten setiap tahun.
- **Seasonal Pattern**:
  - Puncak pada **musim panas** (Juni–Agustus) dan **akhir tahun** (November–Desember).
  - Terendah pada bulan-bulan awal tahun.
- **Tidak ada missing values**, sehingga tidak perlu imputasi.

---

### 2. Model yang Digunakan

#### (1) Logistic Regression

- **Alasan Penggunaan**: Model linier sederhana yang cocok untuk klasifikasi biner.

- **Keunggulan:**
  - Interpretasi koefisien mudah.
  - Cepat dalam training dan prediksi.
- **Parameter:** Default (random\_state=42).

## (2) Decision Tree

- **Alasan Penggunaan:** Menangkap pola non-linier dan interaksi fitur.
- **Keunggulan:**
  - Tidak memerlukan scaling.
  - Dapat mengekstrak aturan penting (misal: “Jika bulan Desember, maka high\_demand=1”).
- **Parameter:**
  - max\_depth=5 (untuk mencegah overfitting).
  - random\_state=42.

## 3. Hasil Evaluasi

### Metrik Evaluasi (Test Set):

Metrik	Logistic Regression	Decision Tree
Accuracy	0.83	0.89
Precision	0.82	0.90
Recall	0.85	0.89
F1-Score	0.83	0.89
AUC-ROC	0.89	0.93

### Confusion Matrix (Contoh):

### Logistic Regression:

Predicted 0 Predicted 1	
Actual 0	14
Actual 1	2
	11

Decision Tree:

Predicted 0 Predicted 1	
Actual 0	15
Actual 1	1
	12

#### 4. Pembahasan

Perbandingan Model:

1. **Decision Tree lebih unggul** di semua metrik karena:
  - Mampu menangkap pola musiman (misal: puncak liburan) yang non-linier.
  - **AUC-ROC 0.93** menunjukkan kemampuan diskriminasi yang sangat baik.
2. **Logistic Regression masih baik** (Accuracy 83%), tetapi:
  - Kurang fleksibel dalam memodelkan interaksi bulan dan tahun.

Interpretasi Model:

- **Decision Tree:**
  - Fitur paling penting: **month\_encoded** (bulan) dan **year\_since\_1949** (tren peningkatan penumpang).
  - Aturan contoh:
    - Jika **June–August** (**is\_summer=1**), prediksi **high\_demand=1**.
    - Jika **November–December** (**is\_holiday\_season=1**), prediksi **high\_demand=1**.

Rekomendasi:

- **Untuk prediksi praktis:** Gunakan **Decision Tree** karena lebih akurat.
- **Untuk interpretasi:** Logistic Regression memberi insight pengaruh linier tahun/bulan.

- **Improvement:**
    - Coba **Random Forest** atau **XGBoost** untuk meningkatkan generalisasi.
    - Tambahkan fitur interaksi (misal: year × month).
- 

## Kesimpulan

**Decision Tree** lebih cocok untuk prediksi permintaan penerbangan karena akurasi dan kemampuan menangkap pola musiman. Logistic Regression berguna untuk analisis awal yang sederhana.