

Nama : Ilham Novriadi

NIM : 231011403539

Kelas : 05TPLE004

## 1. Informasi Dataset

- **Nama Dataset:** Bank Marketing Dataset
  - **Sumber:** [Kaggle – Bank Marketing Dataset](#)
  - **Jumlah Data:** 11.162 baris dan 17 kolom
  - **Tujuan:** Memprediksi apakah seorang nasabah akan membuka deposito berjangka (kolom target: `deposit`)
  - **Tipe Data:**
    - Numerik: `age`, `balance`, `duration`, `campaign`, `pdays`, `previous`
    - Kategorikal: `job`, `marital`, `education`, `default`, `housing`,  
`loan`, `contact`, `month`, `poutcome`
  - **Distribusi Target:**
    - `no` : 5.873 data
    - `yes` : 5.289 data → Dataset relatif seimbang.
  - **Kualitas Data:** Tidak ada nilai hilang (`missing values = 0` untuk seluruh kolom).
- 

## 2. Model yang Digunakan

Dua model utama digunakan untuk klasifikasi biner (`deposit` : yes/no):

### 1. Logistic Regression

- Pipeline terdiri dari:
  - *Preprocessing:*
    - Standardisasi (StandardScaler) untuk fitur numerik
    - One-Hot Encoding untuk fitur kategorikal
  - *Model:* Logistic Regression (`max_iter=1000`,  
`class_weight='balanced'`)
- Digunakan karena sifatnya yang interpretable dan cocok untuk klasifikasi biner.

### 2. Decision Tree Classifier

- Pipeline terdiri dari:

- *Preprocessing*: Sama seperti Logistic Regression
  - *Model*: DecisionTreeClassifier ( `max_depth=5` , `random_state=42` , `class_weight='balanced'` )
  - Cocok untuk data campuran (numerik dan kategorikal) serta memberikan interpretasi berbasis aturan.
- 

### 3. Hasil Evaluasi dan Pembahasan

#### a. Logistic Regression

Metrik	Nilai
Accuracy	0.8289
Precision (kelas 1)	0.8207
Recall (kelas 1)	0.8176
F1-Score (kelas 1)	0.8191
ROC-AUC	<b>0.9071</b>

**Analisis:** Model Logistic Regression memiliki performa terbaik dengan ROC-AUC tinggi ( $>0.9$ ), menunjukkan kemampuan diskriminatif yang sangat baik antara nasabah yang berpotensi membuka deposito dan yang tidak.

---

#### b. Decision Tree

Metrik	Nilai
Accuracy	0.8039
Precision (kelas 1)	0.7798
Recall (kelas 1)	0.8166
F1-Score (kelas 1)	0.7978
ROC-AUC	<b>0.8745</b>

**Analisis:** Model Decision Tree memberikan hasil yang cukup baik namun masih di bawah Logistic Regression. Kemungkinan karena kedalaman pohon yang dibatasi (`max_depth=5`) untuk menghindari overfitting.

---

#### c. Perbandingan Umum

Model	ROC-AUC	Akurasi	Kesimpulan
Logistic Regression	<b>0.9071</b>	0.8289	Performa terbaik, seimbang antara recall & precision
Decision Tree	0.8745	0.8039	Cukup baik, tapi kurang optimal

**Visualisasi:** ROC Curve menunjukkan Logistic Regression memiliki area di bawah kurva yang lebih luas dibanding Decision Tree, mengindikasikan performa klasifikasi yang lebih unggul.

---

## Kesimpulan Akhir

- Logistic Regression merupakan model terbaik untuk dataset ini dengan **ROC-AUC = 0.9071** dan **akurasi = 82.9%**.
  - Decision Tree tetap berguna untuk interpretasi aturan, namun performanya lebih rendah.
  - Dataset yang bersih dan relatif seimbang membantu kedua model mencapai performa yang stabil.
  - Langkah selanjutnya dapat mencakup tuning hyperparameter lebih lanjut (misalnya `max_depth` atau `min_samples_split` untuk Decision Tree) dan uji model lain seperti Random Forest atau XGBoost.
- 

Apakah Anda ingin saya ubah ringkasan ini menjadi **laporan formal (format docx atau pdf)** dengan struktur akademik dan visualisasi hasilnya?