

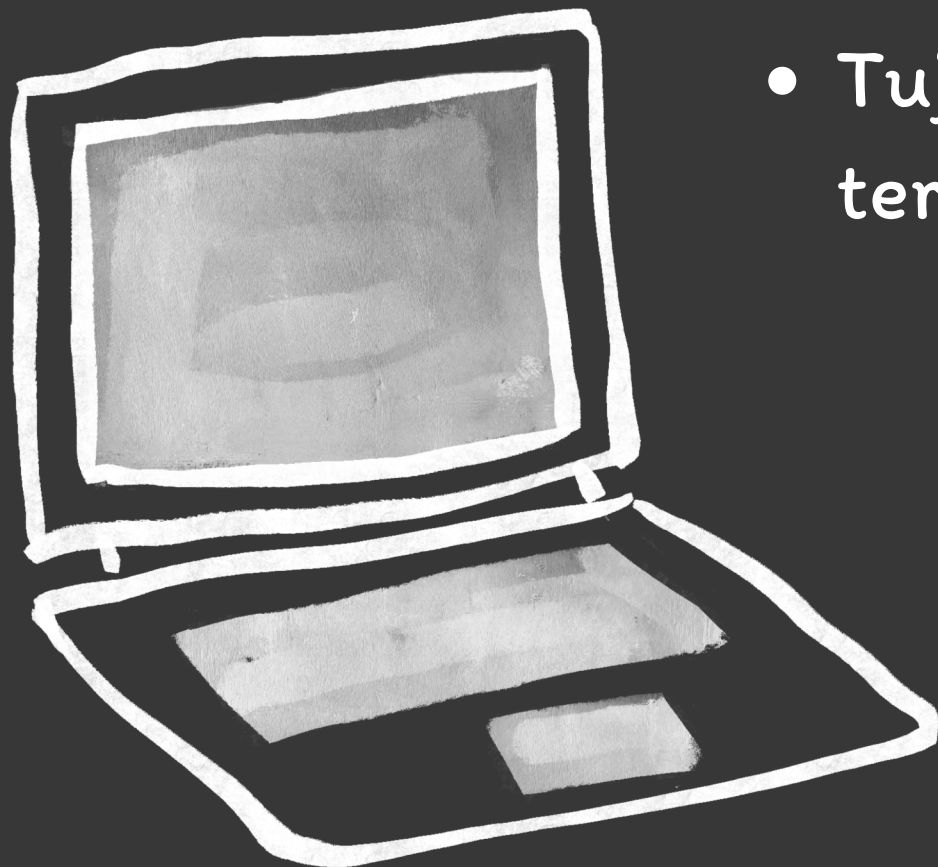
# PROJECT MINI PORTFOLIO: MACHINE

BY : ALIF FAUZAN



# PENDAHULUAN

- Dalam proyek ini, saya membangun dan mengevaluasi beberapa model machine learning untuk klasifikasi dataset Iris.
- Model yang diuji: K-Nearest Neighbors (KNN), Decision Tree, Random Forest, dan XGBoost.
- Tujuan utama: Mencari model terbaik dengan akurasi tertinggi dan generalisasi yang baik.





## **DATASET: IRIS DATASET (SCIKIT-LEARN)**

Fitur:

- Sepal Length (cm)
- Sepal Width (cm)
- Petal Length (cm)
- Petal Width (cm)

- Target:
  - Setosa (0)
  - Versicolor (1)
  - Virginica (2)
- Pembagian Data:
  - 80% untuk training
  - 20% untuk testing



# MODEL YANG DIUJI & HASIL EVALUASI

## 1 K-Nearest Neighbors (KNN) 📍

- Akurasi: 100% ✓
- Confusion Matrix: Tidak ada kesalahan klasifikasi
- Kelebihan: Sederhana dan akurat untuk dataset kecil
- Kekurangan: Kurang efisien jika dataset besar

## 2 Decision Tree 🌳

- Akurasi awal: 33.3% ✗ (underfitting)
- Setelah tuning: 93.3% ✓
- Kelebihan: Mudah diinterpretasikan
- Kekurangan: Cenderung overfitting tanpa tuning

## 3 Random Forest 🌲🌲

- Akurasi awal: 90%
- Setelah tuning: 93% ✓
- Kelebihan: Stabil & lebih tahan terhadap overfitting
- Kekurangan: Butuh lebih banyak sumber daya



## 4 XGBoost ⚡

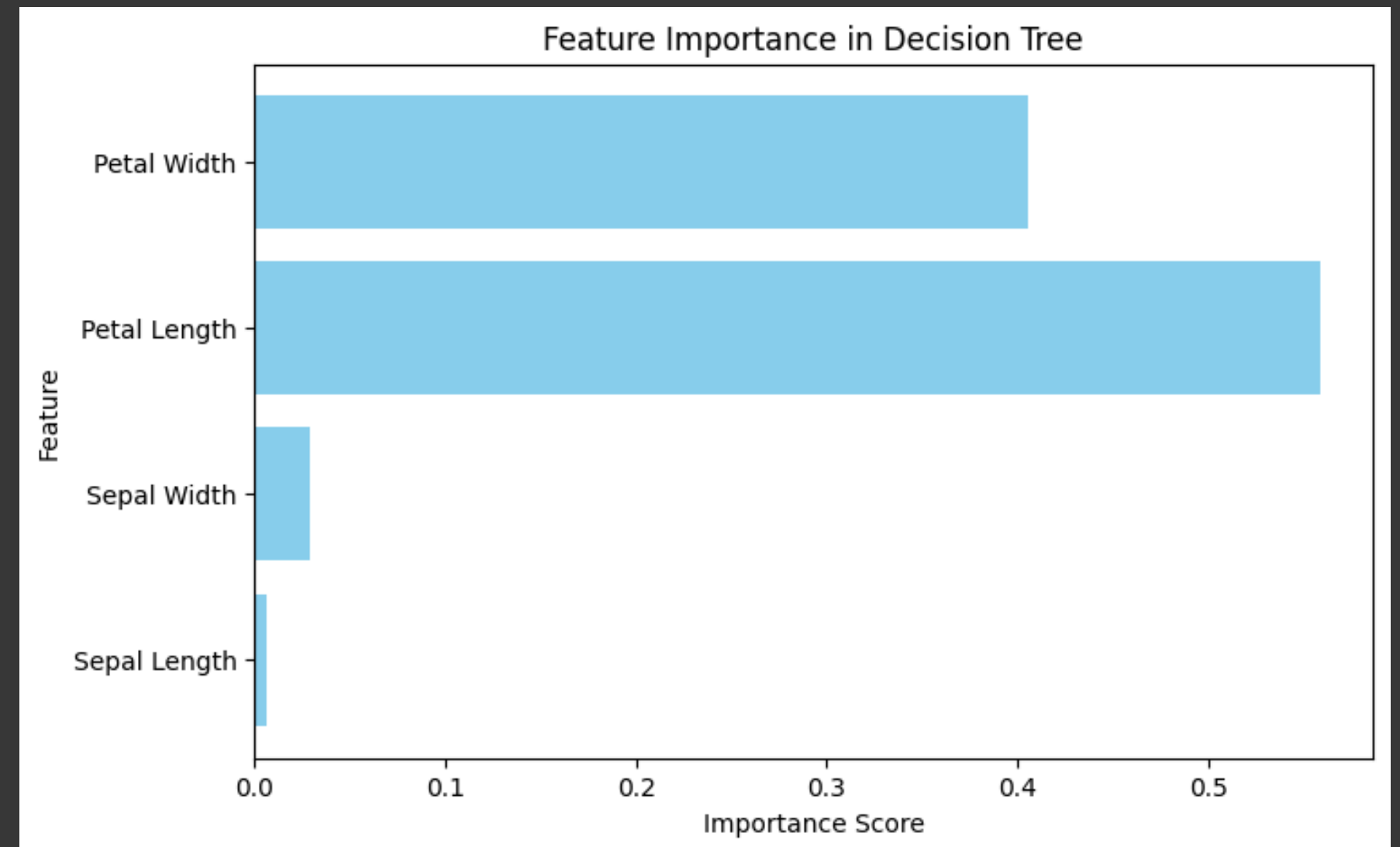
- Akurasi awal: 100% ✓
- Setelah tuning: 100% ✓
- Kelebihan: Cepat, efisien, dan kuat dalam menangani dataset besar
- Kekurangan: Lebih kompleks dibanding model lain



# ANALISIS FEATURE IMPORTANCE

Dari model Decision Tree, Random Forest, dan XGBoost:

- Fitur paling berpengaruh:  Petal Length & Petal Width
- Fitur kurang signifikan:  Sepal Width & Sepal Length





# Kesimpulan & Model Terbaik



- XGBoost & KNN mencapai akurasi 100% → Keduanya adalah model terbaik.
- XGBoost lebih cocok untuk dataset besar, sedangkan KNN lebih baik untuk dataset kecil.
- Random Forest cukup stabil dan bisa jadi alternatif jika XGBoost terlalu kompleks.



## NEXT STEPS

- **1** Uji model dengan dataset berbeda untuk melihat apakah tetap akurat. **2** Deploy model ke API/web agar bisa digunakan lebih luas. **3** Gunakan SHAP untuk interpretasi lebih dalam tentang keputusan model.

• Terima kasih! 





**THANK  
YOU**

