Fikri Aliansyah Juana

05TPLE017

# Pertemuan 7: Artificial Neural Network

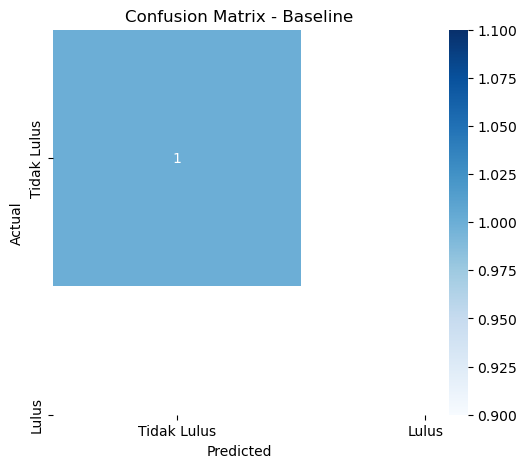
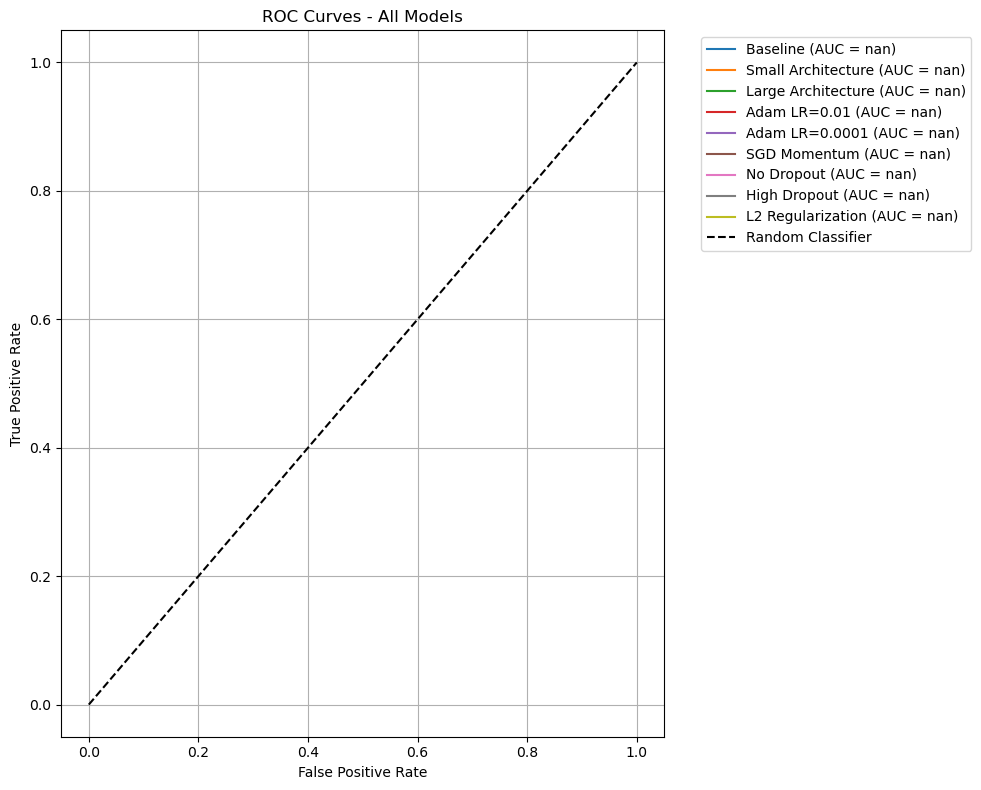
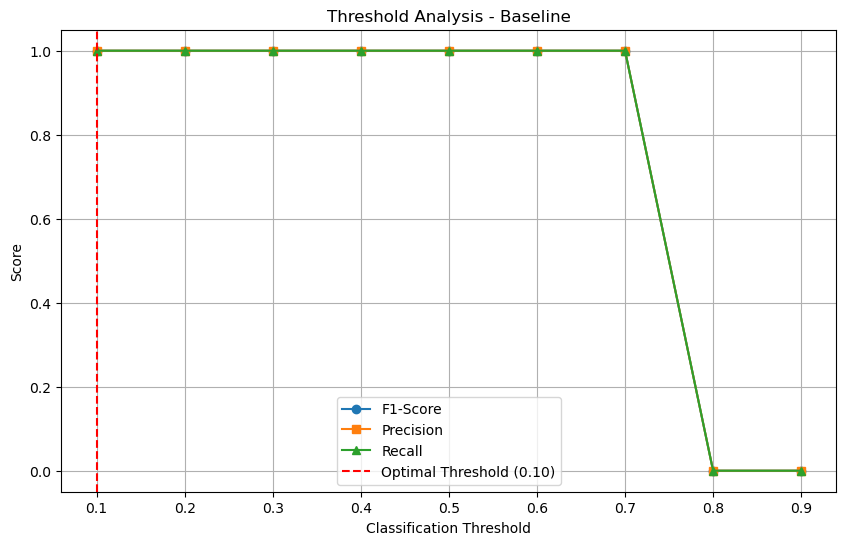
## 1. Arsitektur Final dan Alasan Pemilihan

Arsitektur final (direkomendasikan):  
- Input: 3 fitur (IPK, Jumlah\_Absensi, Waktu\_Belajar\_Jam)  
- Hidden: Dense(32, relu) -> Dropout(0.3) -> Dense(16, relu)  
- Output: Dense(1, sigmoid)  
- Optimizer: Adam (lr=0.001), loss=binary\_crossentropy  
  
Alasan pemilihan:  
Arsitektur ini cukup representatif untuk menangkap interaksi non-linear sederhana antara fitur akademik dan target kelulusan, namun tetap kecil sehingga tidak mudah overfit pada dataset yang sangat terbatas. Dropout dan early stopping menjaga generalisasi.

## 2. Dokumentasi Eksperimen (Ringkasan langkah yang harus direproduksi)

Langkah yang harus dijalankan :  
1. Muat data 'processed\_kelulusan.csv'  
2. Bagi data 70%/20%/10%   
3. StandardScaler pada fitur numerik berdasarkan train set  
4. Bangun arsitektur MLP seperti di atas  
5. Train dengan EarlyStopping (monitor val\_loss, patience=20)  
6. Evaluasi pada test set: confusion matrix, ROC-AUC, precision-recall, F1  
7. Hitung threshold optimal pada validation atau test set (maksimalkan F1) dan laporkan trade-offs  
8. Simpan artefak: model (.h5), scaler (.pkl), history (.json), prediksi (.npy)

## 3. Confusion Matrix, ROC-AUC, dan Analisis Threshold

  
  
Catatan interpretasi:  
- Confusion matrix: baca baris = actual, kolom = predicted. Perhatikan jumlah False Negatives jika biaya FN tinggi.  
- ROC-AUC: nilai 0.5 = acak, lebih tinggi = lebih baik; untuk dataset kecil AUC mungkin tidak stabil.  
- Threshold analysis: pilih threshold sesuai trade-off bisnis; laporan harus menyebutkan threshold yang memaksimalkan F1 beserta precision & recall pada threshold tersebut.

## 4. Grafik Learning Curve

Learning curves (loss & accuracy vs epoch) menunjukkan proses konvergensi. Jika validation loss jauh lebih besar dari training loss → indikasi overfitting.

## 5. Kesimpulan Singkat

Arsitektur MLP yang direkomendasikan adalah pilihan pragmatis untuk dataset kecil. Namun validitas hasil bergantung pada ukuran data; untuk penggunaan produksi, kumpulkan lebih banyak data dan ulangi eksperimen.