Nama : Fawzy Imam Hamdani Gunamukti

NIM : 231011401030

KLS : 05TPLE017

Penjelasan Machine Learning Pertemuan 7

1. Persiapan Data

* Membaca dataset dari temp/processed\_kelulusan.csv menggunakan pandas.
* Memisahkan fitur (X) dan label (y) dengan y = df["Lulus"].
* Melakukan standardisasi fitur menggunakan StandardScaler agar semua fitur berada pada skala yang sama.
* Membagi dataset menjadi tiga bagian:
  + X\_train, y\_train: data latih (70%)
  + X\_val, y\_val: data validasi (15%)
  + X\_test, y\_test: data uji (15%)
* Pembagian dilakukan dengan stratifikasi agar proporsi kelas tetap seimbang.

1. Pembangunan Model

* Menggunakan Keras Sequential API.
* Arsitektur model:
  + Input layer dengan jumlah neuron sesuai jumlah fitur.
  + Hidden layer 1: 32 neuron, aktivasi ReLU.
  + Dropout 0.3 untuk mencegah overfitting.
  + Hidden layer 2: 16 neuron, aktivasi ReLU.
  + Output layer: 1 neuron dengan aktivasi sigmoid (karena klasifikasi biner).
* Optimizer: Adam (learning rate 0.001).
* Loss function: Binary Crossentropy.
* Metrics: Accuracy dan AUC.

1. Proses Pelatihan

* Model dilatih selama maksimal 100 epoch dengan batch size 32.
* Menggunakan callback EarlyStopping dengan parameter:
  + monitor = "val\_loss"
  + patience = 10
  + restore\_best\_weights = True
* Early stopping berfungsi menghentikan training bila performa validasi tidak meningkat selama 10 epoch.
* Data validasi digunakan untuk memantau performa model selama pelatihan.

1. Evaluasi Model

* Model dievaluasi pada data uji (X\_test, y\_test).
* Metrik utama yang dihitung: Accuracy dan AUC.
* Menghasilkan prediksi probabilitas (y\_proba) dan label akhir (y\_pred) dengan ambang batas 0.5.
* Menampilkan:
  + Confusion Matrix untuk melihat jumlah prediksi benar/salah.
  + Classification Report yang memuat precision, recall, dan f1-score.

1. Visualisasi

* Membuat grafik Learning Curve dari loss dan val\_loss untuk melihat performa model selama pelatihan.
* Menyimpan grafik hasil pelatihan ke file temp/learning\_curve.png.

