1. Persiapan & Import Library
Library: pandas, sklearn (model_selection, pipeline, preprocessing, ensemble, metrics), matplotlib, joblib
Error: ImportError: cannot import name 'SimpleImputer' dari 'sklearn.preprocessing'
Solusi: impor SimpleImputer dari sklearn.impute

2. Membaca Dataset
Dataset: processed_kelulusan.csv
Error: FileNotFoundError: No such file or directory
Solusi: gunakan path lengkap sesuai lokasi file di Pydroid3
Dataset berhasil dibaca: 10 baris, 6 kolom

3. Membagi Dataset
Split pertama: train 70%, temp 30% (stratify)
Split kedua: val 50%, test 50% dari temp (tanpa stratify karena dataset kecil)
Error sebelumnya: ValueError: The least populated class in y has only 1 member,
ValueError: n_splits=5 cannot be greater than number of members in each class
Solusi: hapus stratify di split kecil, gunakan n_splits=2

4. Pipeline Preprocessing
Kolom numerik: IPK, Jumlah_Absensi, Waktu_Belajar_Jam, Rasio_Absensi, IPK_x_Study
Proses: median imputation + standard scaling

5. Model RandomForest
Parameter baseline: n_estimators=300, max_features="sqrt", class_weight="balanced"
Pipeline: preprocessing + classifier
Model berhasil fit di training set

6. Evaluasi Validation Set
Prediksi: y_val_pred
F1-macro: 1.0
Classification report dicetak

7. Cross-Validation
StratifiedKFold n_splits=2
CV F1-macro: mean ± std dicetak
Error sebelumnya: n_splits=5 terlalu besar, solusi: disesuaikan n_splits=2

8. Hyperparameter Tuning (GridSearchCV)
Parameter grid: max_depth=[None,12,20,30], min_samples_split=[2,5,10]
CV menggunakan StratifiedKFold n_splits=2
Best estimator ditemukan
F1-macro validation terbaik dicetak

9. Evaluasi Test Set
Prediksi: y_test_pred
F1-macro, classification report, confusion matrix dicetak

10. ROC & Precision-Recall Curve
Probabilitas prediksi digunakan
ROC Curve disimpan: roc_test.png
Precision-Recall Curve disimpan: pr_test.png
Error sebelumnya: hanya satu gambar tersimpan, solusi: buat figure eksplisit (fig1, fig2)
Kedua gambar ditampilkan

11. Feature Importance
Dihitung dari RandomForest
Top 10 fitur dicetak
Jika error: "Feature importance tidak tersedia"

12. Menyimpan Model
Model disimpan: rf_model.pkl menggunakan joblib

13. Contoh Prediksi
Sample input: IPK 3.4, Absensi 4, Waktu Belajar 7, Rasio_Absensi 0.2857, IPK_x_Study 23.8
Prediksi sample dicetak: 0/1 (lulus/tidak)

Ringkasan Error
1. ImportError SimpleImputer → versi scikit-learn lama
2. FileNotFoundError CSV → path file salah
3. ValueError stratify dataset kecil → jumlah per kelas <2
4. ValueError StratifiedKFold → n_splits terlalu besar
5. Output gambar tidak tersimpan → figure tidak eksplisit