**LAPORAN LEMBAR KERJA PERTEMUAN 6**

Nama : Muhammad Hafith Ghiffari Almughni

NIM : 231011402136

Kelas : 05TPLE015

Mata Kuliah : Machine Learning

1. Tujuan

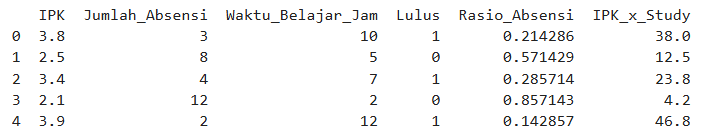
Mengetahui bagaimana Random Forest digunakan untuk klasifikasi, melakukan baseline dengan Decision Tree, membandingkan performa, melakukan tuning hyperparameter, dan mengevaluasi model akhir menggunakan F1, Precision, Recall, ROC-AUC, confusion matrix, serta feature importance.

2. Dataset

Dataset: processed\_kelulusan.csv

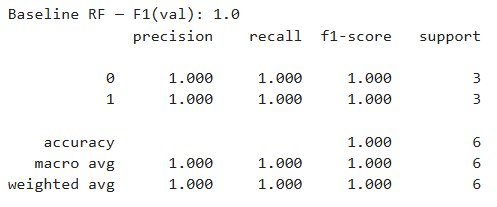
Target: Lulus (0 = tidak lulus, 1 = lulus)

Fitur: IPK, Jumlah Absensi, Waktu Belajar Jam, Rasio Absensi, IPK\_x\_Study.

Data dibagi dengan stratified split: 70% train, 15% validation, 15% test.

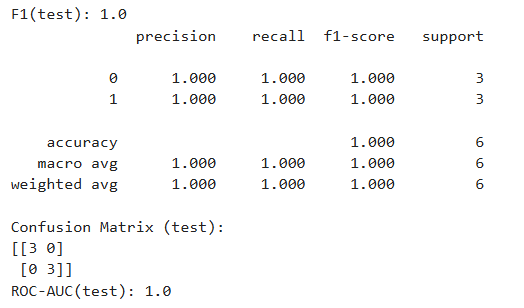
= Hasil dari Train, Validation, dan Test

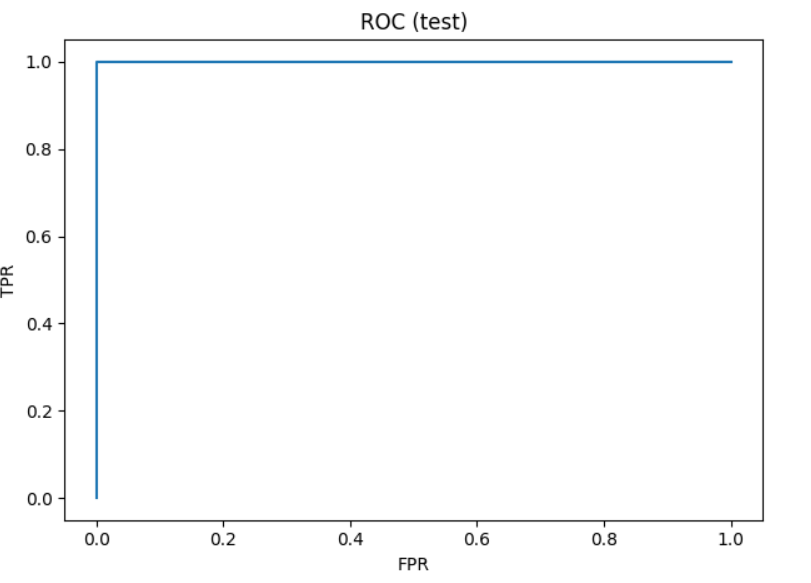
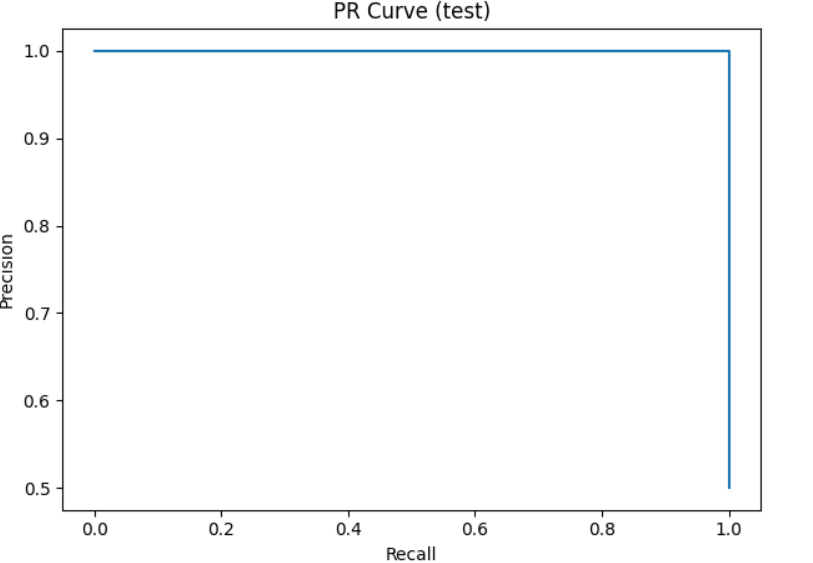
3. Baseline Model

Baseline dipilih Decision Tree (default) karena model ini sederhana dan interpretatif. Model ini digunakan sebagai titik awal pembanding sebelum dilakukan tuning atau penggunaan model lain

4. Random Forest

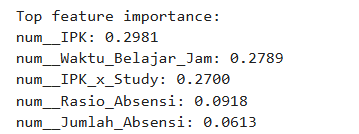
Sebagai model alternatif, digunakan Random Forest Classifier.

Alasan: lebih robust, dapat mengurangi overfitting, dan mendukung evaluasi feature importance.



* Baseline, Random Forest sudah sangat baik dengan F1 1.0, dan hasil tuning tidak memberikan peningkatan yang signifikan karena metrik sudah maksimal

5. Analisis Feature Importance

- Tiga fitur teratas (importance) dan implikasinya:

1. IPK (num\_\_IPK) → 29.8%

-IPK menjadi faktor paling berpengaruh dalam memprediksi kelulusan.

-Logis, karena semakin tinggi IPK maka kemungkinan lulus tepat waktu juga semakin tinggi.

1. Waktu Belajar per Jam (num\_\_Waktu\_Belajar\_Jam) → 27.9%

-Durasi belajar mandiri mahasiswa juga sangat menentukan.

-Semakin konsisten waktu belajar, semakin besar peluang mahasiswa menyelesaikan studi tepat waktu.

1. Interaksi IPK × Study Hours (num\_\_IPK\_x\_Study) → 27.0%

-Kombinasi antara prestasi akademik (IPK) dan intensitas belajar memberi sinyal yang kuat.

Artinya: bukan hanya IPK tinggi saja, tetapi bila IPK tinggi dan waktu belajar juga konsisten, peluang kelulusan lebih terjamin.

Implikasi:

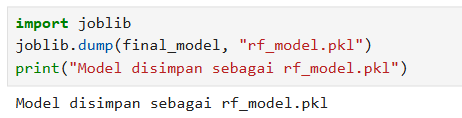
-Untuk mahasiswa: fokus pada menjaga IPK tetap baik sambil menjaga rutinitas belajar yang cukup.

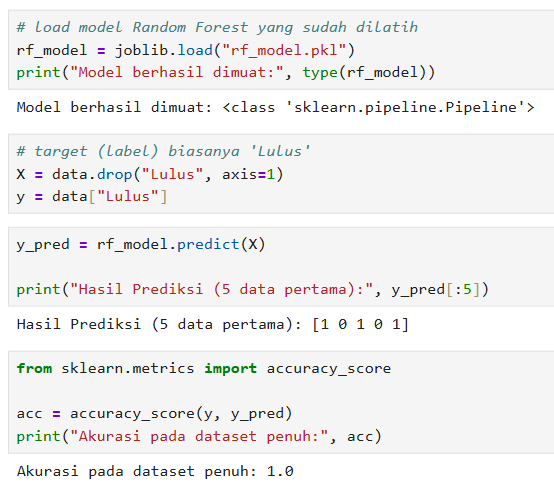
-Untuk dosen atau pengelola: bisa memberi intervensi pada mahasiswa dengan IPK rendah dan waktu belajar minim, karena dua hal ini adalah indikator utama keterlambatan.

-Secara model: ketiga fitur ini mendominasi hampir 85% pengaruh prediksi (29.8 + 27.9 + 27.0).

model dapat membantu universitas mengidentifikasi mahasiswa berisiko tidak lulus tepat waktu, sehingga dapat diberikan intervensi.

6. Penyimpanan Model

Model terbaik disimpan ke file rf\_model.pkl agar bisa digunakan tanpa perlu retraining.

Hasil prediksi pada 5 data pertama adalah [1 0 1 0 1], dan ketika dibandingkan dengan label sebenarnya, model mencapai akurasi sempurna (1.0) pada seluruh dataset, artinya model mampu mengklasifikasikan semua data dengan benar tanpa kesalahan.