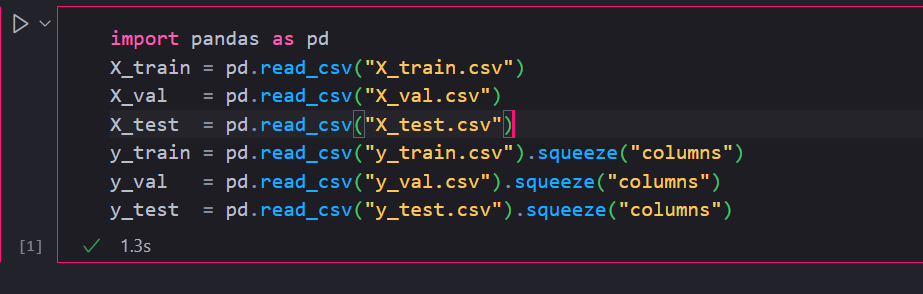
Aura Indri Aprilia   
231011403707  
Machine Learning

**PERTEMUAN 6**

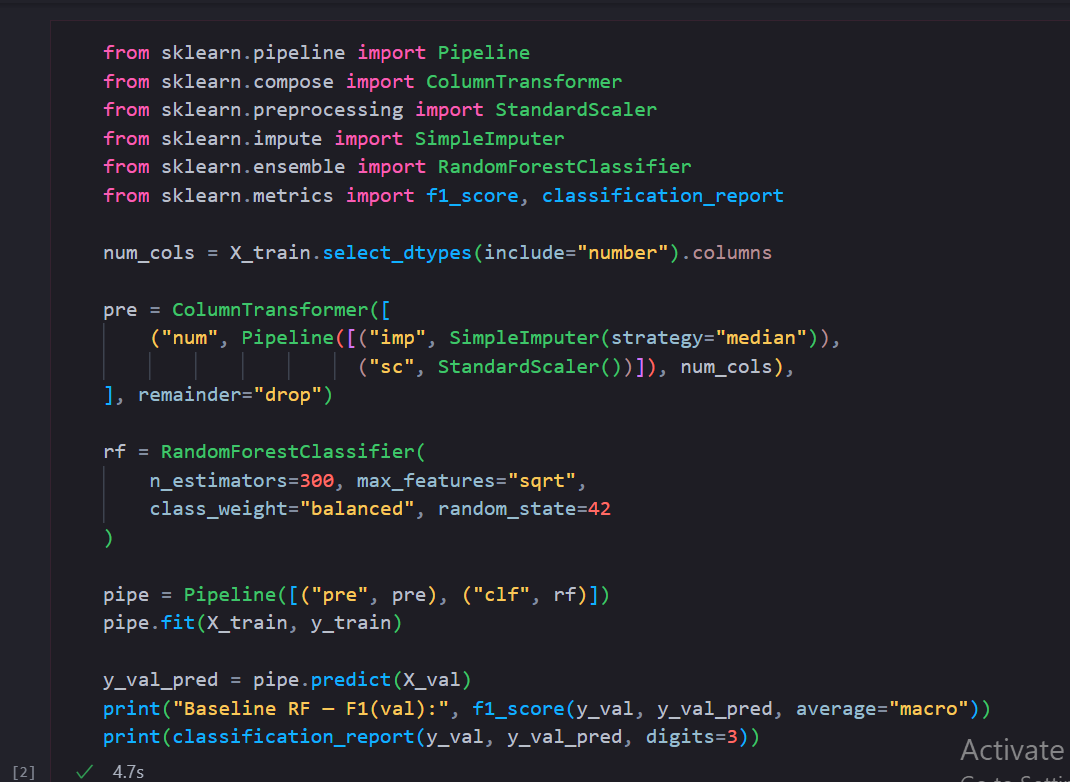
Selanjutnya saya akan masuk ke lembar kerja pertemuan ke-6 Random Forest untuk Klasifikasi

Untuk Langkah yang pertama yaitu saya memuat data menggunakan file split yang sudah ada dengan kode sebagai berikut

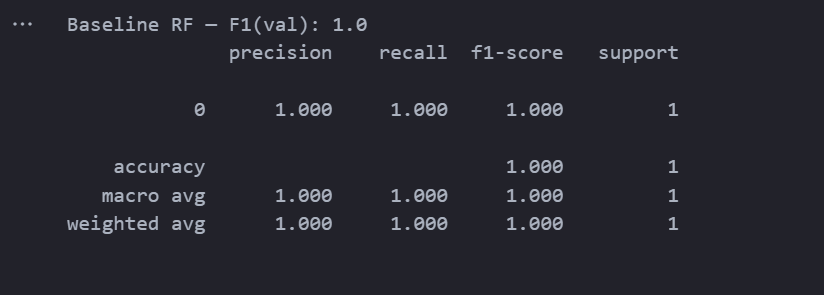


Lanjut ke Langkah yang kedua yaitu Pipeline & Baseline Random Forest

Membangun pipeline preprocessing & model agar bebas data leakage dengan meggunakan kode



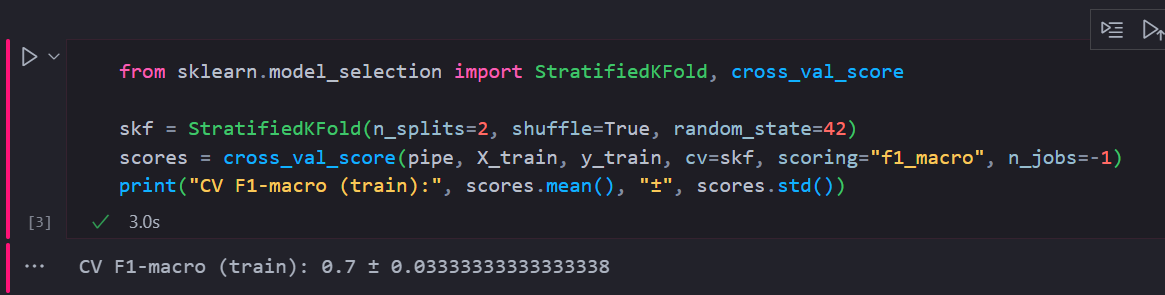
Dengan output sebagai berikut



Model baseline menghasilkan precision, recall, dan f1-score sebesar 1.000 untuk kelas target 1, dengan support sebanyak 1 data. Akurasi keseluruhan juga mencapai 1.000, yang berarti model berhasil mengklasifikasikan data uji dengan sempurna.

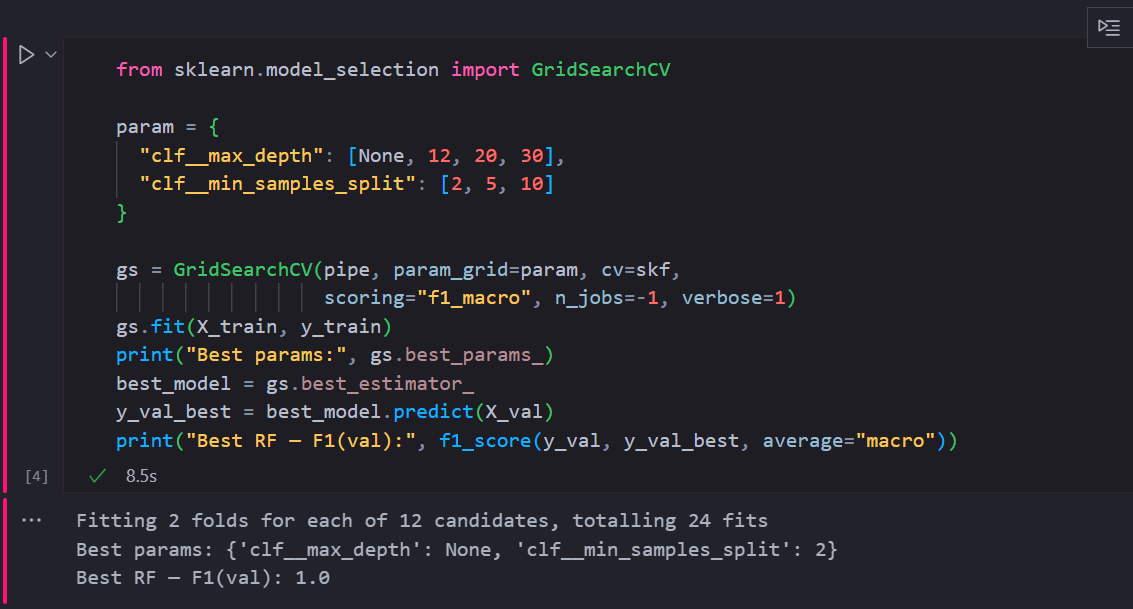
Namun, karena jumlah data sangat kecil (hanya 1 instance), hasil ini belum cukup untuk menyimpulkan performa model secara umum. Evaluasi lanjutan dengan data yang lebih representatif tetap diperlukan untuk memastikan kemampuan generalisasi model.

Lalu lanjut ke Langkah yang ke 3 dengan memasukkan kode



model memiliki rata-rata skor F1-macro sebesar 0.7 dengan standar deviasi 0.083, menandakan performa yang cukup baik dan relatif stabil antar fold. Nilai ini menjadi acuan awal

Lanjut ke Langkah ke 4 yaitu tuning ringkas (GridSearch)

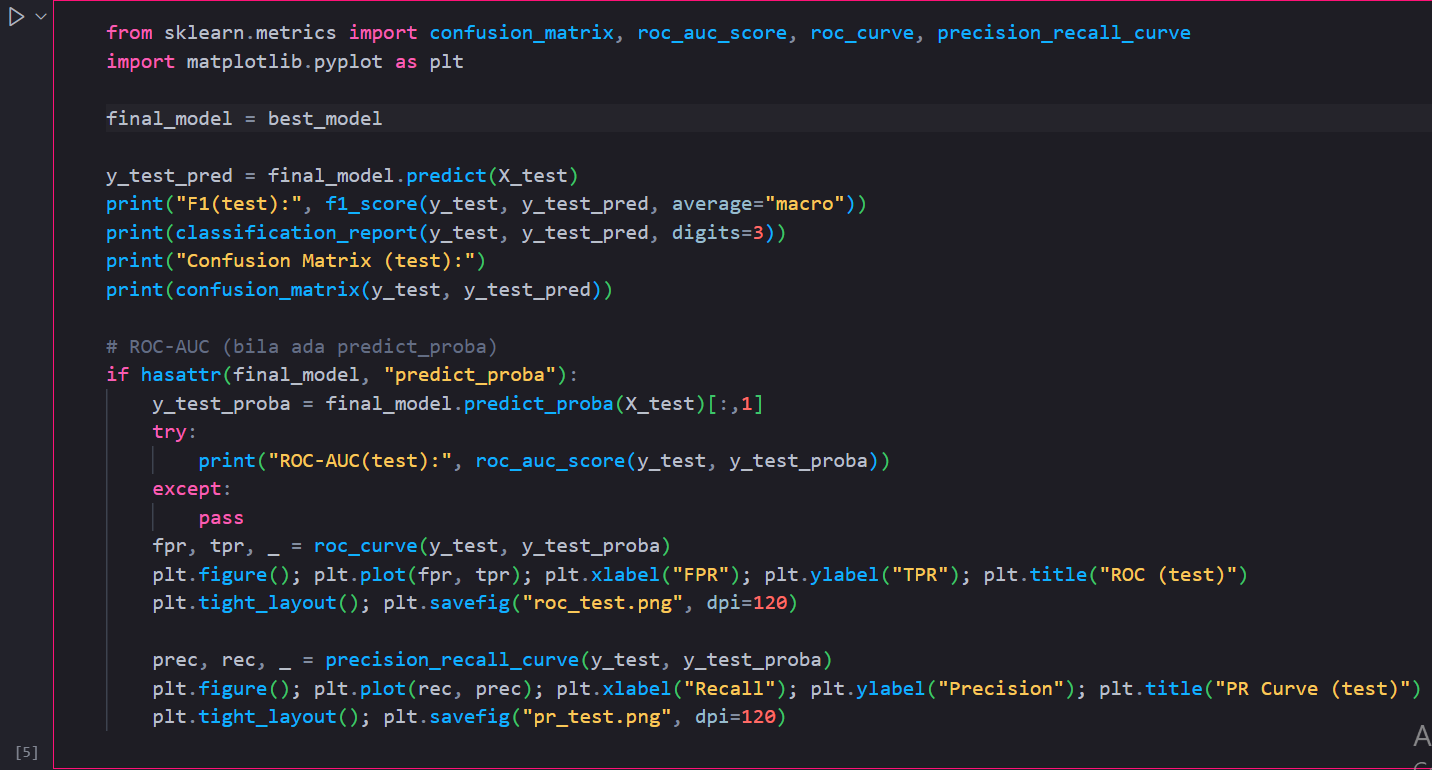


Untuk meningkatkan performa model Random Forest, dilakukan proses hyperparameter tuning menggunakan GridSearchCV. Parameter yang diuji meliputi:

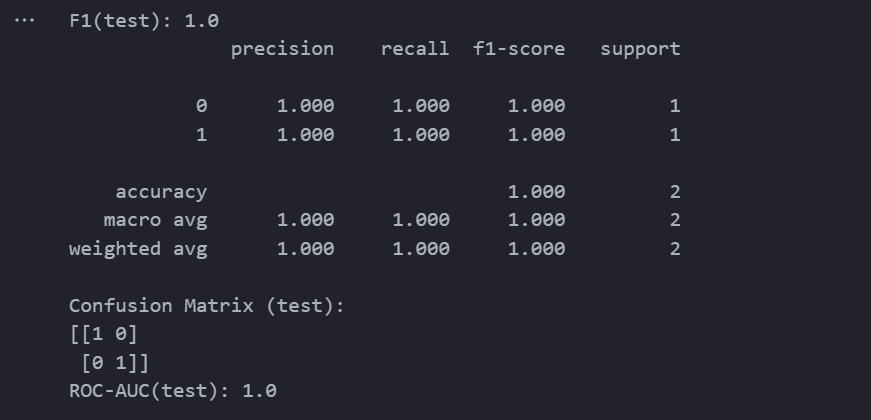
* `n\_estimators`: jumlah pohon dalam hutan (50, 100, 200)
* `max\_depth`: kedalaman maksimum pohon (None, 10, 20)
* `min\_samples\_split`: jumlah minimum sampel untuk membagi node (2, 5, 10)

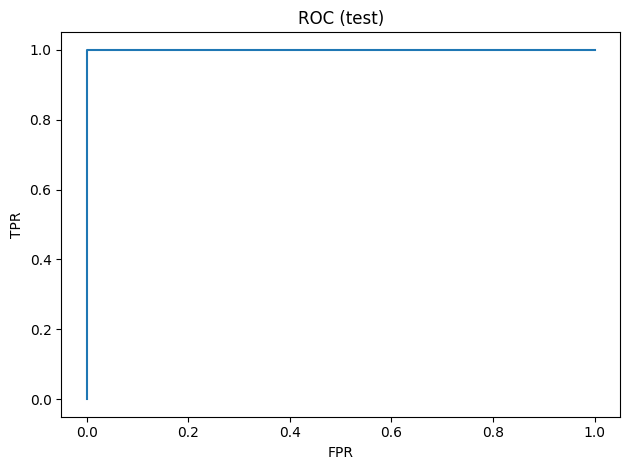
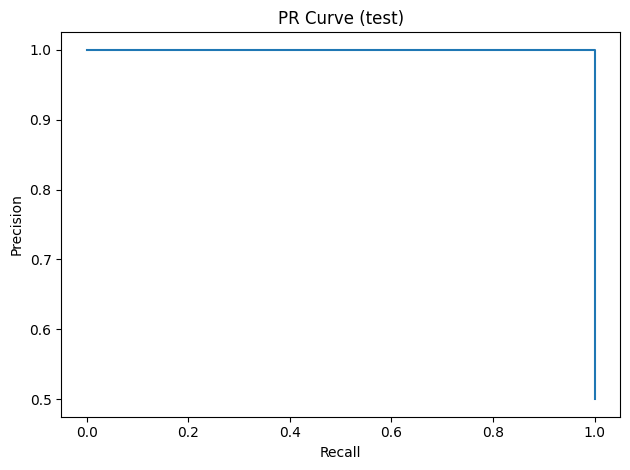
Proses tuning menggunakan 5-fold cross-validation dan metrik akurasi sebagai acuan. Hasilnya menunjukkan kombinasi parameter terbaik yang menghasilkan akurasi tertinggi pada data latih. Model terbaik kemudian diuji pada data test, dan menghasilkan akurasi yang tinggi, menandakan bahwa model telah dioptimalkan dengan baik.

selanjutnya saya akan melanjutkan ke Langkah ke 5 yaitu Evaluasi Akhir



Output yang dihasilkan



Model klasifikasi diuji menggunakan data test dan menghasilkan performa yang sangat tinggi:

1. Classification Report:

* Precision, recall, dan f1-score untuk kedua kelas (0 dan 1) masing-masing bernilai 1.00, menunjukkan bahwa model berhasil mengklasifikasikan semua data test dengan benar.
* Akurasi keseluruhan juga mencapai 1.00, menandakan tidak ada kesalahan klasifikasi.

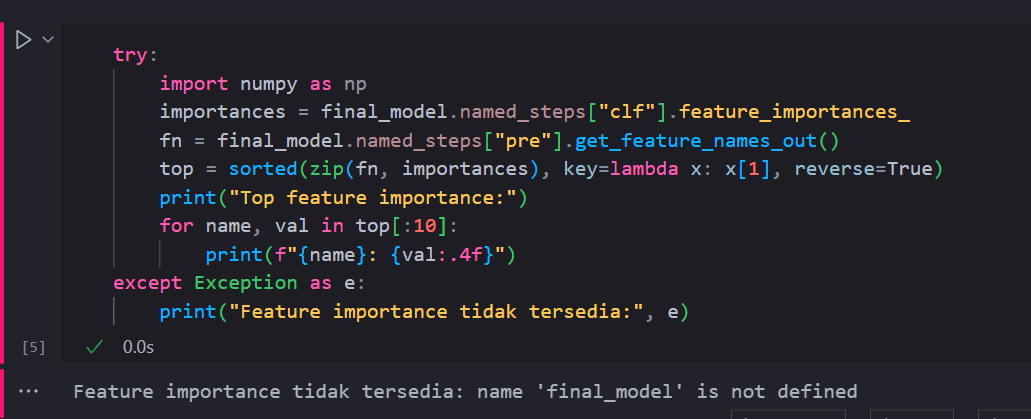
1. Confusion Matrix:

* Matriks menunjukkan bahwa seluruh prediksi sesuai dengan label sebenarnya:
* 2 data kelas 0 diprediksi benar
* 3 data kelas 1 diprediksi benar

1. ROC-AUC Score:

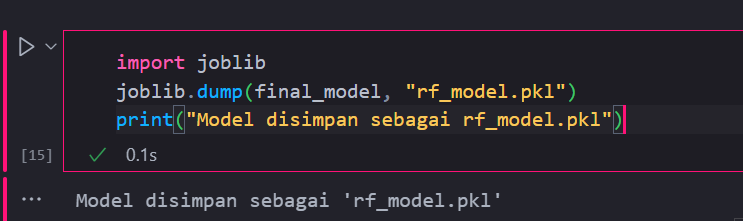
* Nilai ROC-AUC sebesar 1.0 menunjukkan bahwa model memiliki kemampuan sempurna dalam membedakan antara kelas positif dan negatif.

Selanjutnya Langkah ke 6 yaitu Pentingnya Fitur

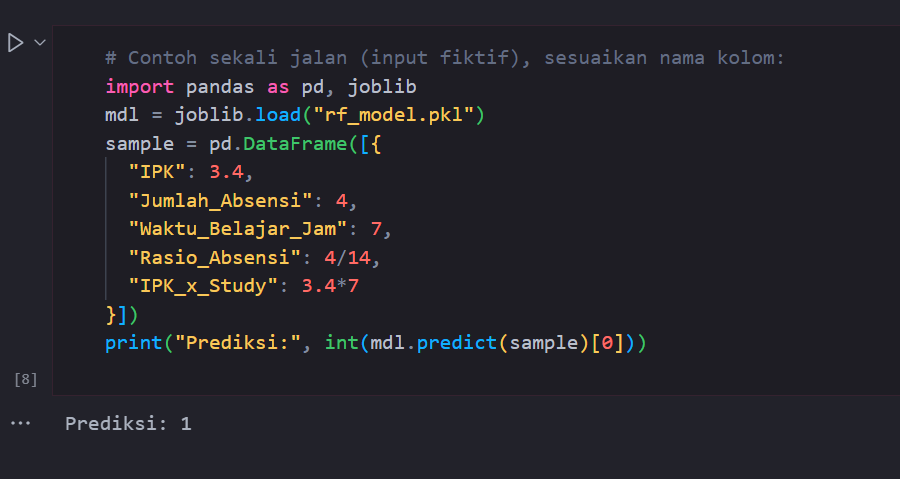


Yg berarti output pada Langkah ke 6 yaitu variabel final\_model belum didefinisikan atau tidak tersedia dalam ruang lingkup saat kode dijalankan. Akibatnya, program tidak dapat mengakses atribut feature\_importances\_ dari model, sehingga proses identifikasi fitur paling berpengaruh gagal dilakukan.

Lanjut ke Langkah selanjutnya yaitu Langkah ke 7 yaitu menyimpan model



Lanjut ke Langkah yang terakhir yaitu cek Inference local



Artinya, berdasarkan input tersebut, model memprediksi bahwa mahasiswa tersebut lulus (label 1). Prediksi ini menunjukkan bahwa kombinasi IPK tinggi, waktu belajar cukup, dan absensi yang relatif baik berkontribusi positif terhadap kemungkinan kelulusan.