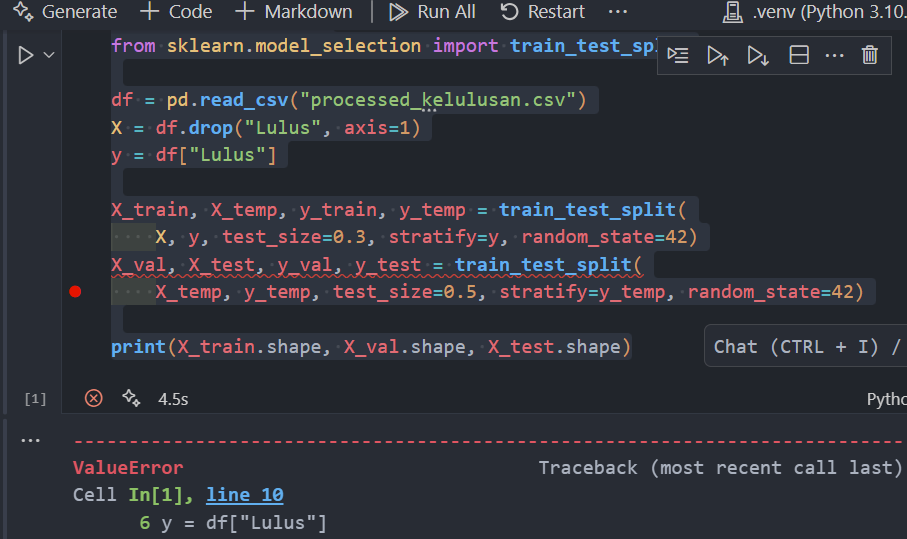
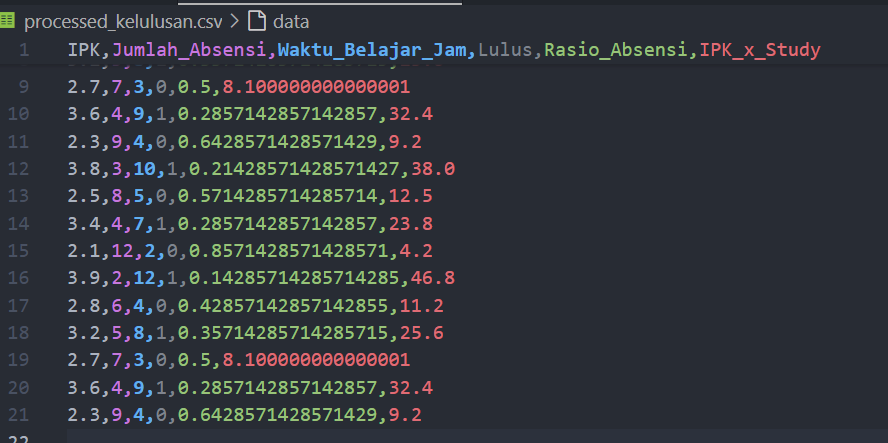
1. **Langkah 1 — Muat Data**

Pilihan B (pakai processed\_kelulusan.csv) lalu split ulang. Saat mau split lagi (train\_test\_split kedua), stratify=y\_temp butuh minimal **2 data di setiap kelas** → tapi karena ada yang hanya 1 → error.

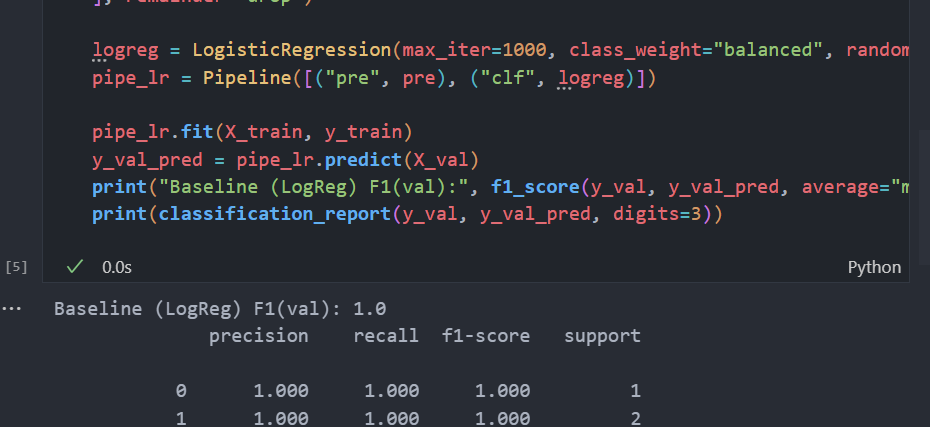


Tambahkan jumlah baris data pada “procesed\_kelulusan.csv” supaya setiap kelas punya cukup sampel.



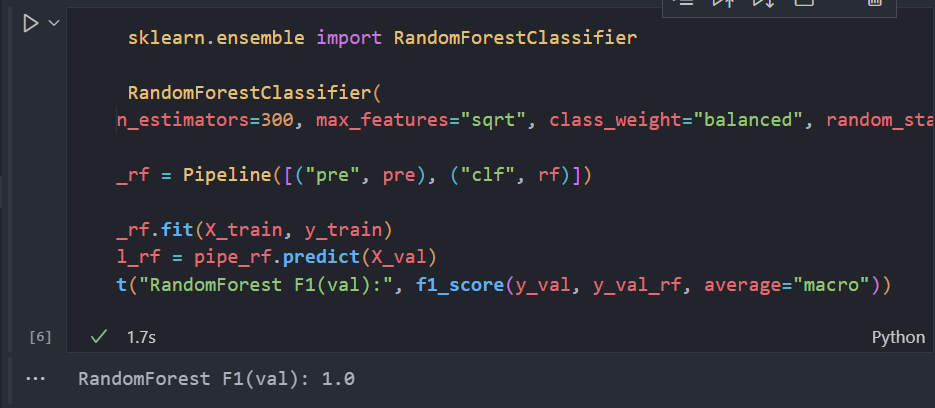
1. **Langkah 2 — Baseline Model & Pipeline**

Baseline berfungsi memberikan gambaran performa minimum yang dapat dicapai menggunakan model sederhana, sehingga setiap peningkatan performa pada model berikutnya dapat diukur secara objektif melalui metrik evaluasi. Sementara itu, penggunaan Pipeline memastikan tahapan preprocessing (seperti imputasi nilai hilang dan normalisasi fitur) berjalan konsisten, terstruktur, dan mudah direplikasi pada data latih maupun data uji.

****

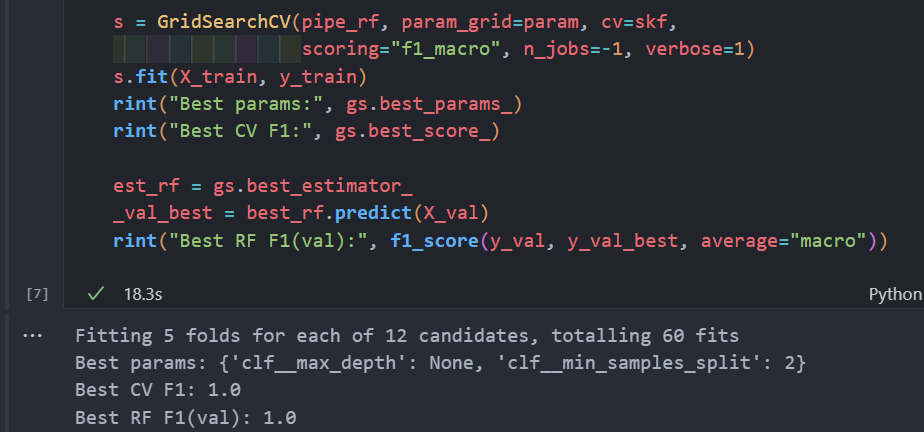
1. **Langkah 3 — Model Alternatif (Random Forest)**

Langkah ini dipakai untuk membangun model alternatif yang lebih kuat dibanding baseline, lalu dibandingkan kinerjanya.



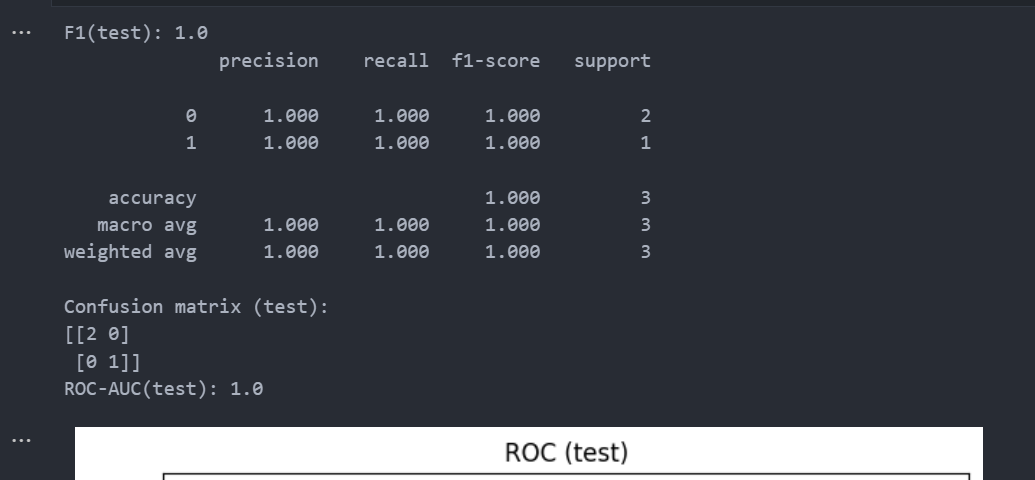
1. **Langkah 4 — Validasi Silang & Tuning Ringkas**

Langkah ini digunakan untuk meningkatkan kinerja Random Forest melalui validasi silang dan pencarian parameter terbaik, sehingga model akhir lebih handal, generalisasi lebih baik, dan tidak bergantung pada kebetulan split data.

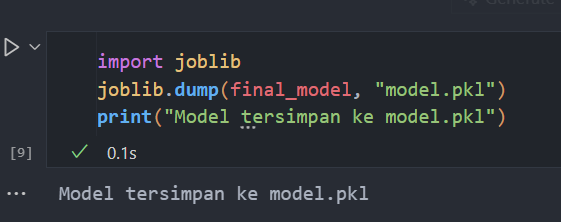


1. **Langkah 5 — Evaluasi Akhir (Test Set)**

tujuan utama evaluasi akhir adalah memvalidasi generalisasi model sebelum dipakai di dunia nyata.



1. **Langkah 6 (Opsional) — Simpan Model**

Penyimpanan model memastikan hasil kerja tidak hilang dan siap dipakai kapan saja tanpa retraining.