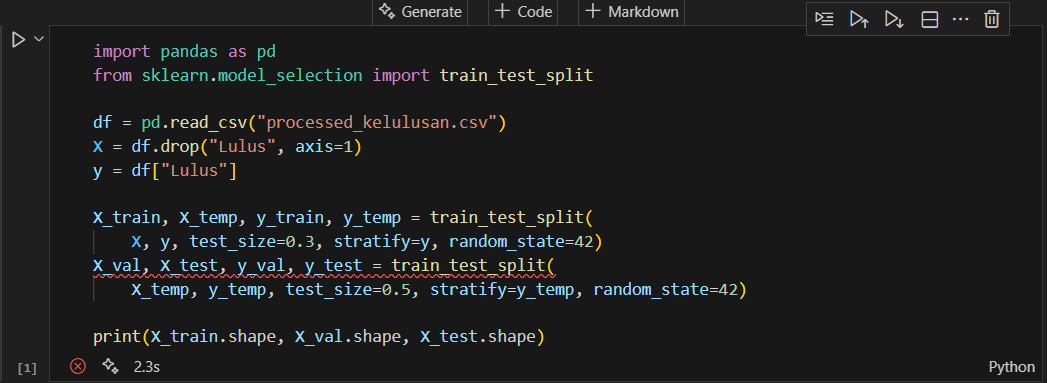
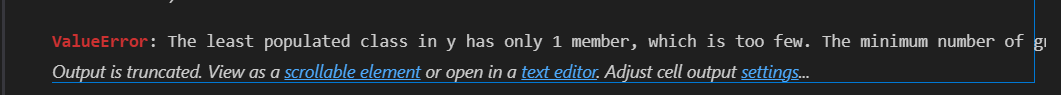
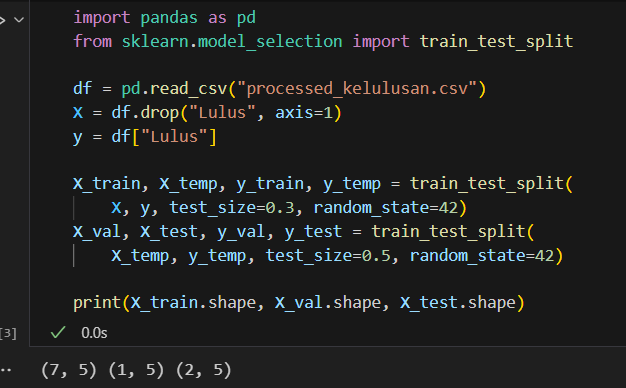
1. **Langkah 1 — Muat Data**

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.



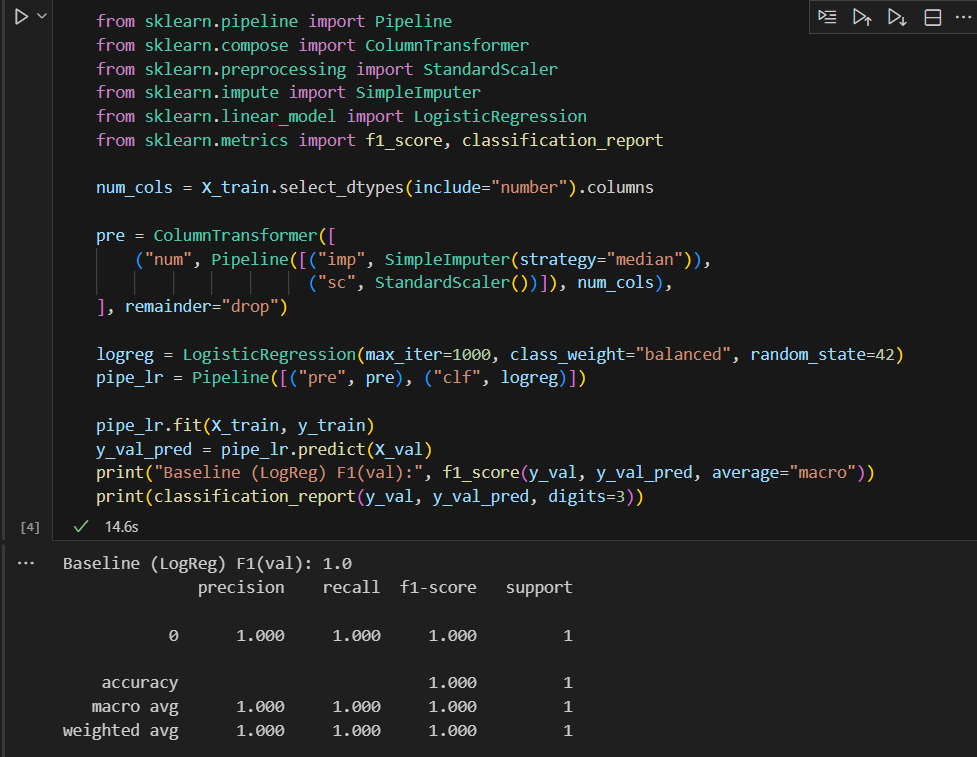


Untuk mengatasi permasalahan tersebut, dilakukan menghapus stratify pada split kedua.



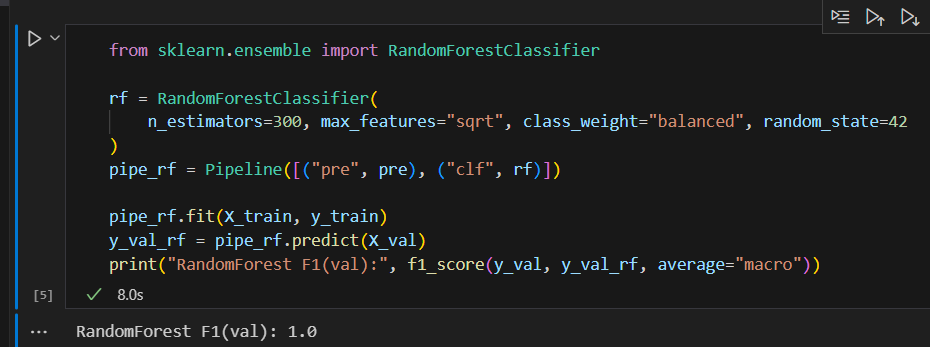
1. **Langkah 2 — Baseline Model & Pipeline**

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.

****

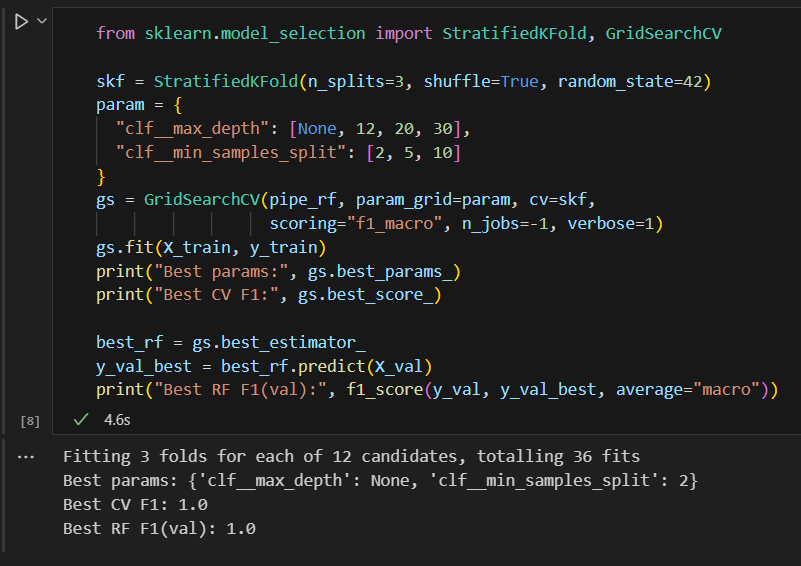
1. **Langkah 3 — Model Alternatif (Random Forest)**

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.

A screen shot of a computer program

Description automatically generated

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

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

A screen shot of a computer program

Description automatically generated

1. **Langkah 7 (Opsional) — Endpoint Inference (Flask)**

**Error karena  server ini tidak dirancang untuk digunakan dalam lingkungan produksi karena tidak efisien, tidak stabil, dan kurang aman untuk melayani trafik sebenarnya secara besar.**

A screen shot of a computer program

Description automatically generated

Alternatif yang saya gunakan yaitu menambahkan **🧠 Kesimpulan singkat**

Kode **setelahnya** menambahkan:

1. **Route /** → agar bisa diakses di browser tanpa 404,
2. **Threading** → supaya bisa dijalankan aman di **VS Code .ipynb**,
3. **Pesan status & pengaturan Flask** → lebih stabil dan tidak hang.

Jadi versi **setelah** adalah versi yang **lebih aman, interaktif, dan kompatibel** untuk dijalankan di Jupyter Notebook / VS Code.

A screenshot of a computer

AI-generated content may be incorrect.A screen shot of a computer program

AI-generated content may be incorrect.