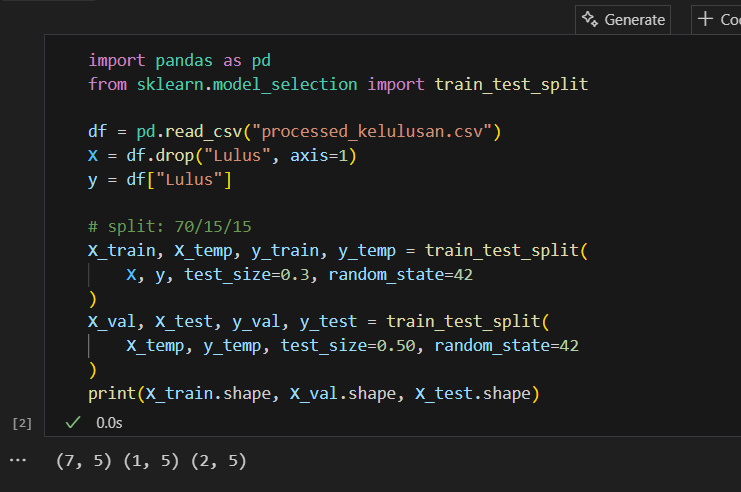
1. **Langkah 1 — Muat Data**

Pilihan A (gunakan processed\_kelulusan.csv)



1. **Langkah 2 — Pipeline & Baseline Random Forest**

Langkah ini digunakan untuk membangun baseline Random Forest yang aman dari data leakage, sekaligus jadi pembanding performa sebelum melakukan tuning model lebih lanjut.

A screen shot of a computer program

AI-generated content may be incorrect.

1. **Langkah 3 — Validasi Silang**

Terjadi eror karena salah satu kelas di y\_train memiliki kurang dari 5 data. Jadi, saya menggunakan KFold atau resampling data agar seimbang.

A screenshot of a computer program

AI-generated content may be incorrect.

A screen shot of a computer program

AI-generated content may be incorrect.

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

Tuning ringkas dengan GridSearchCV digunakan untuk memilih parameter terbaik secara sistematis agar Random Forest mencapai performa optimal dan lebih generalisasi.

A screen shot of a computer program

AI-generated content may be incorrect.

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

Langkah ini berguna untuk menguji performa model yang sudah dituning pada data yang benar-benar baru, mengevaluasi metrik penting (F1, ROC-AUC, PR Curve), dan memastikan model layak untuk deployment.

A screenshot of a computer

AI-generated content may be incorrect.

1. **Langkah 6 — Pentingnya Fitur**

Langkah ini berguna untuk memahami, menjelaskan, dan memvalidasi fitur apa yang benar-benar penting bagi model.

A screen shot of a computer program

AI-generated content may be incorrect.

1. **Langkah 7 — Simpan Model**

Bertujuan agar model yang sudah dilatih bisa disimpan, dibagikan, dan digunakan ulang untuk prediksi tanpa harus melakukan training dari awal.

A screen shot of a computer code

AI-generated content may be incorrect.

1. **Langkah 8 — Cek Inference Lokal**

Digunakan untuk menguji apakah model yang sudah disimpan benar-benar bisa melakukan prediksi (inference) pada data baru sebelum diterapkan ke aplikasi produksi.

