

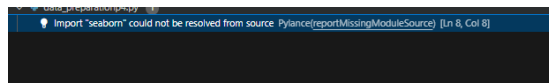
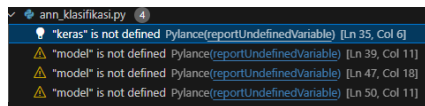
## Data Preparation

NAMA : M Yudi Mariansyah

NIM : 231011400946

KELAS : 05TPLE015

### 1. Error



**ValueError: The least populated class in y has only 1 member...**

### 2. Terjadi Pada Tahap

**Pemisahan Data (Data Splitting)**, khususnya saat membagi data  $X_{temp}$  menjadi Validation dan Test set, dan **Cross-Validation (CV)**.

### 3. Penyebab Singkat

- Data set total **sangat kecil** (sekitar 10 baris).
- Gagalnya fungsi **stratify** (`train_test_split` dan `StratifiedKFold`) karena salah satu kelas target ("Lulus" atau "Tidak Lulus") di *subset* data hanya memiliki **satu (1) sampel**. Scikit-learn tidak bisa membagi satu sampel secara merata untuk mempertahankan rasio kelas.

### 4. Solusi Utama

1. **Pemisahan Data:** Menggunakan blok **try...except** agar skrip otomatis beralih ke pemisahan *non-stratified* jika stratify gagal.
2. **Cross-Validation:** Mengubah `n_splits` dari 5 menjadi **3** pada `StratifiedKFold` agar data *training* yang kecil (7 sampel) dapat dibagi menjadi *folds* yang cukup.

Hasil setelah di perbaiki

