Data Splitting: Train / Validation / Test

Full Dataset

Typical Workflow

1. Split data BEFORE any preprocessing (prevent data leakage)

2. Train on training set, evaluate on validation set

3. Iterate: adjust hyperparameters based on validation performance

4. Final model: evaluate ONCE on test set (report this!)

WARNING: Never optimize based on test set performance (data leakage)!