

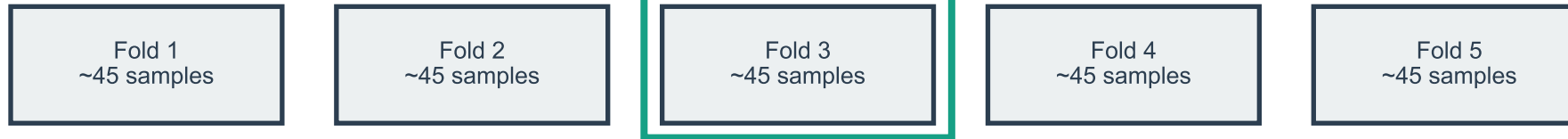
# 5-Fold Cross-Validation: Preventing Data Leakage

**227 Training Samples**

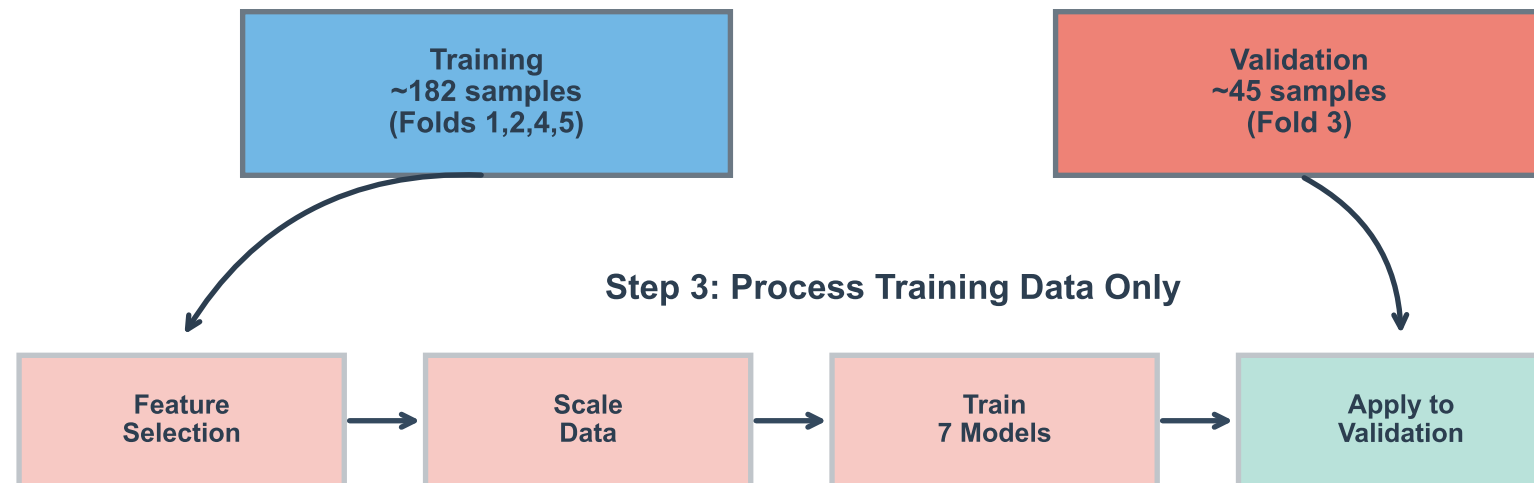
159 Responders (70%) | 68 Non-responders (30%)



## Step 1: Create 5 Stratified Folds



## Step 2: For Each Fold (Example: Fold 3)



## Step 4: Record Validation Performance

AUC = 0.XXX for Fold 3



**Repeat for all 5 folds → Mean AUC: 0.70**

**KEY INSIGHT:** Feature selection happens **INSIDE** each fold using only training data. This prevents overfitting and gives us honest performance estimates.