Model Evaluation and Selection

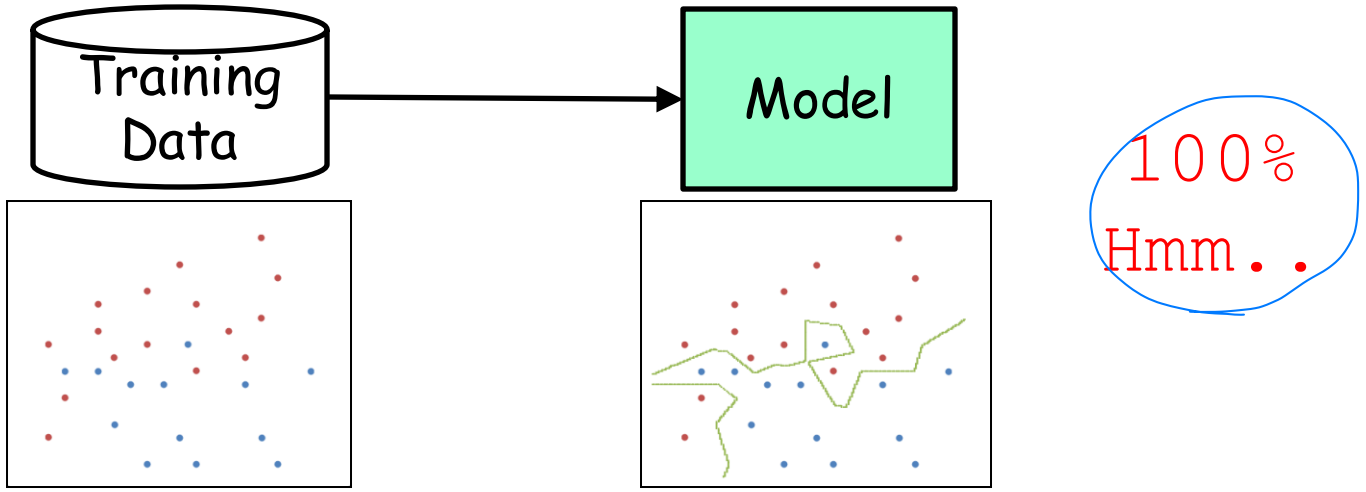
Introduction

- **Model Evaluation**
  - Hold-out
  - k-fold Cross Validation (CV)

- **Model Selection and Evaluation**
  - Using Hold-out
  - Using k-fold Cross Validation

Model Evaluation

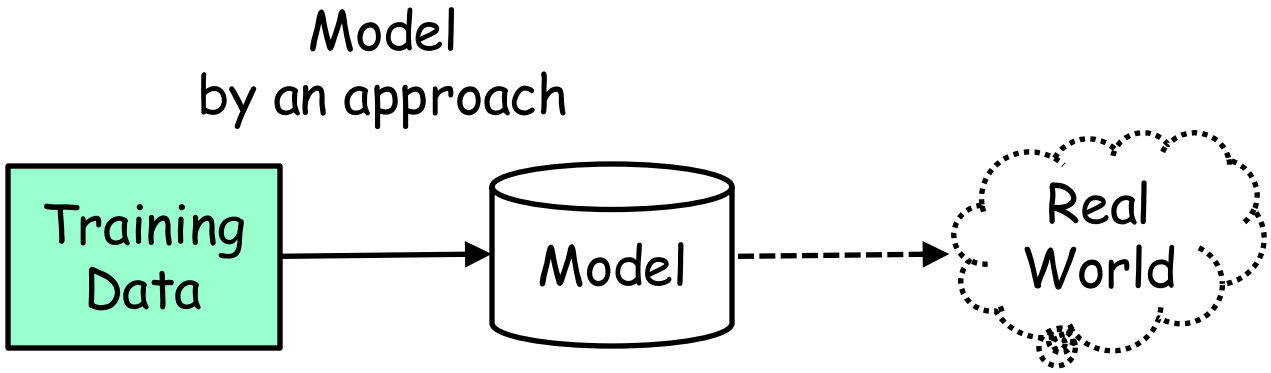
- **I want to estimate the performance of my model**
  - I have built a model
  - My model is 100% correct for the Training Set
  - Can I trust it?



  - We need to estimate its performance using "unknown new data" ← Test

Evaluation Method: Hold-out

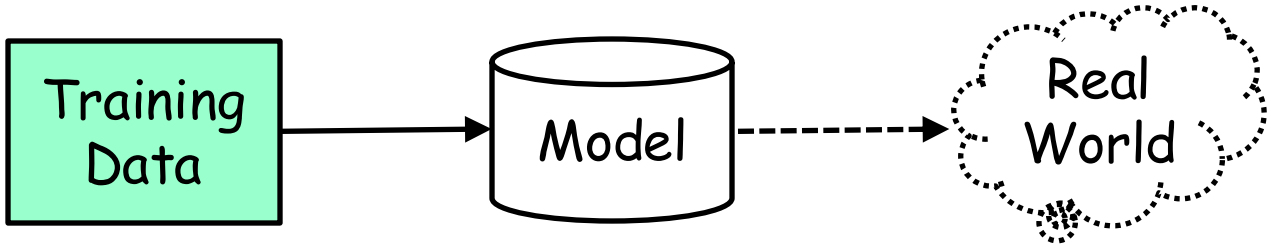
- **"True" evaluation**
  - We trained a model with given data samples
  - We need "new real world data" to evaluate the model
  - It will cost very much 비싸다!
  - Is there any simple approach?



Evaluation Method: Hold-out

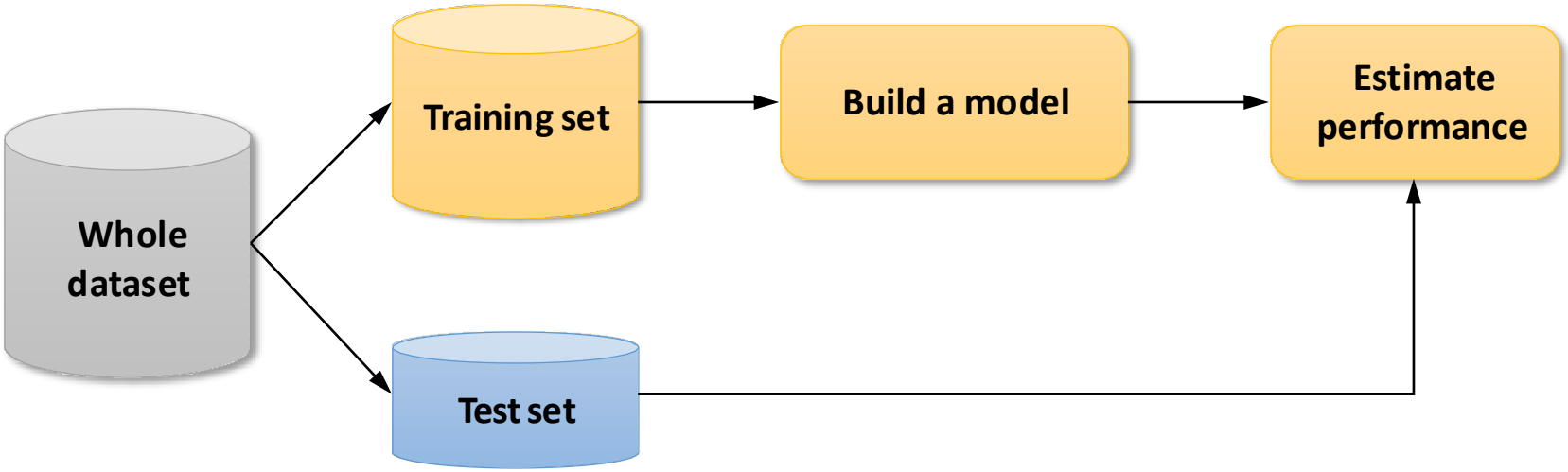
- **"True" evaluation**
  - We trained a model with given training data
  - We need "new real world data" to evaluate the model

- **Questions**
  - Do we need to collect another set of "new real world data"? No! training data ≠ real world data!
  - Is not the "given training data" from the real world?



Evaluation Method: Hold-out

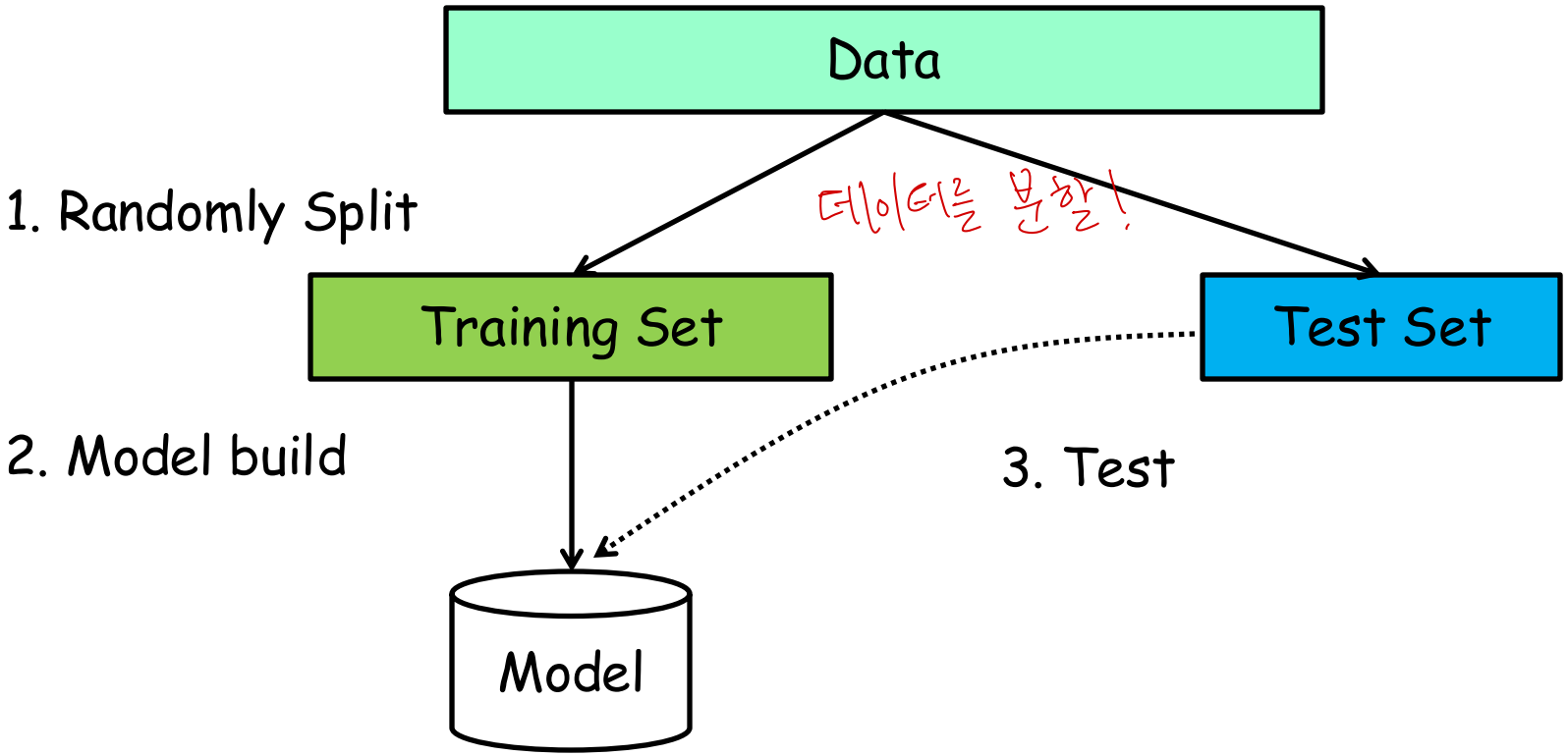
- **Hold-out**
  - Divide the given data into TRAINING set and TEST set
    - Training set and Test set should NOT overlap each other!!
    - Both need to be independent as much as possible
  - With Training set, build a model
  - With Test set, evaluate the model



Evaluation Method: Hold-out

- **Hold-out model**

from real world



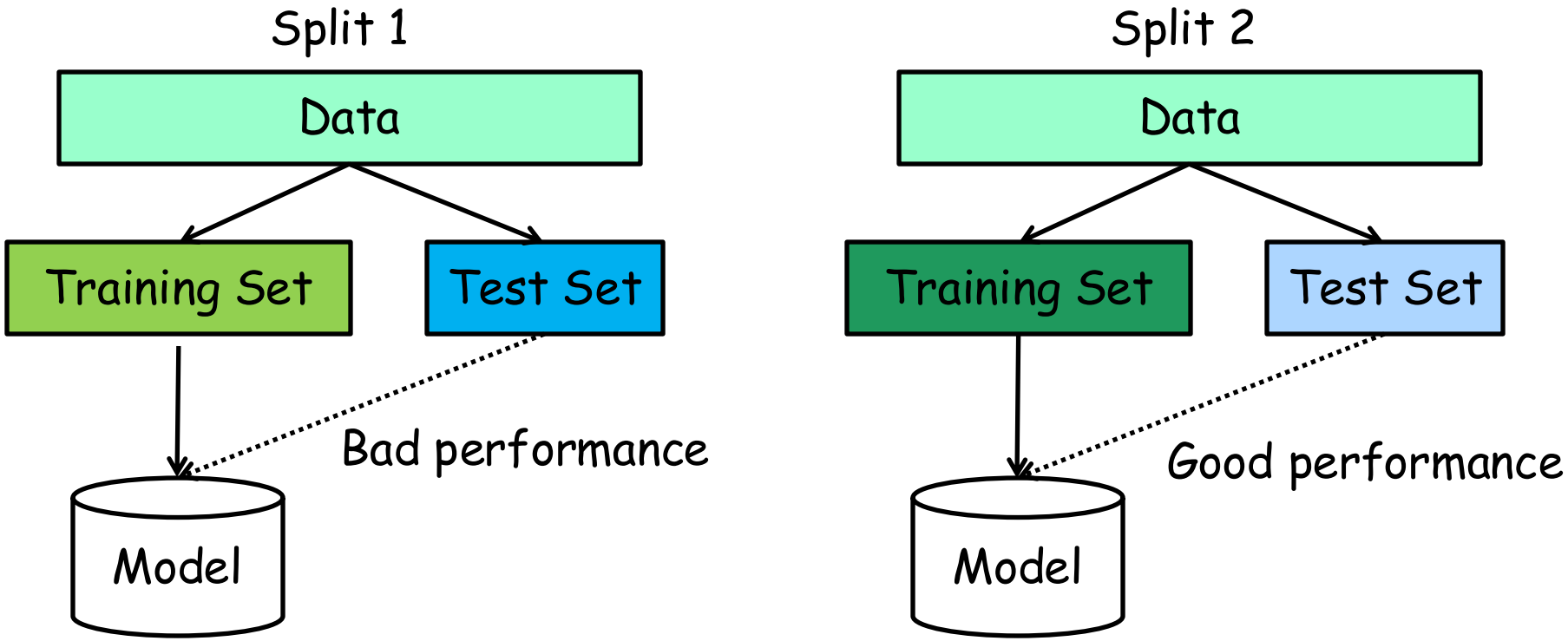
Evaluation Method: Hold-out

- **Size of Test set:**
  - 30~10% of given data
- **Advantage:**
  - Simple, easy and cheap
- **Disadvantage:**
  - Data is randomly split: Evaluation can be significantly different depending on data split

randomly하게 스플릿!

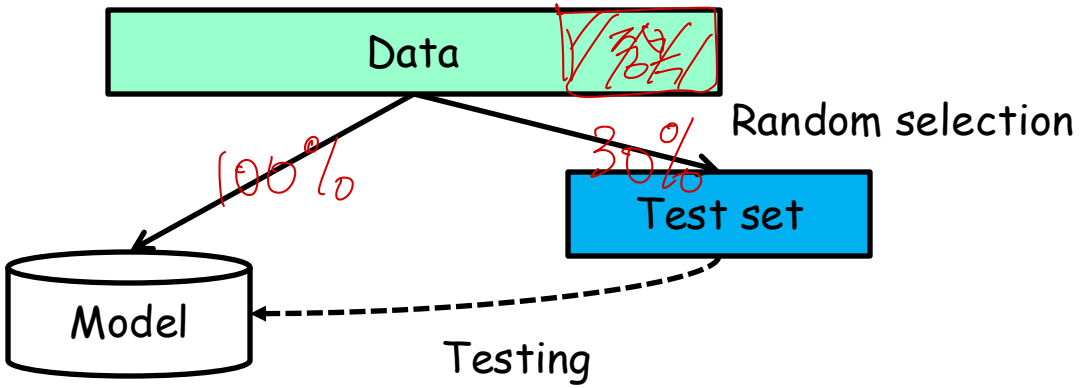
Evaluation Method: Hold-out

- **Disadvantage**
  - Waste of data (?)
  - Random Split: Performance may depend on the split



Evaluation Method: Hold-out

- **Question**
  - I use only 70% of training data for model building
    - It is waste of data :(
  - Can I use the whole data for training, and test the model with some part of training data

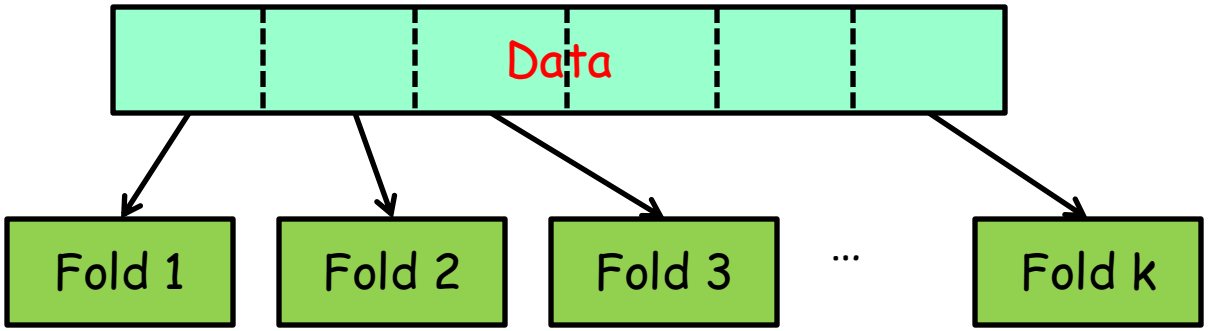


  - Absolutely NO!
  - Training set and test set should be independent

Evaluation Method: Cross Validation

- **In order to reduce statistical variance**
  - Usually, k-fold cross validation is widely used

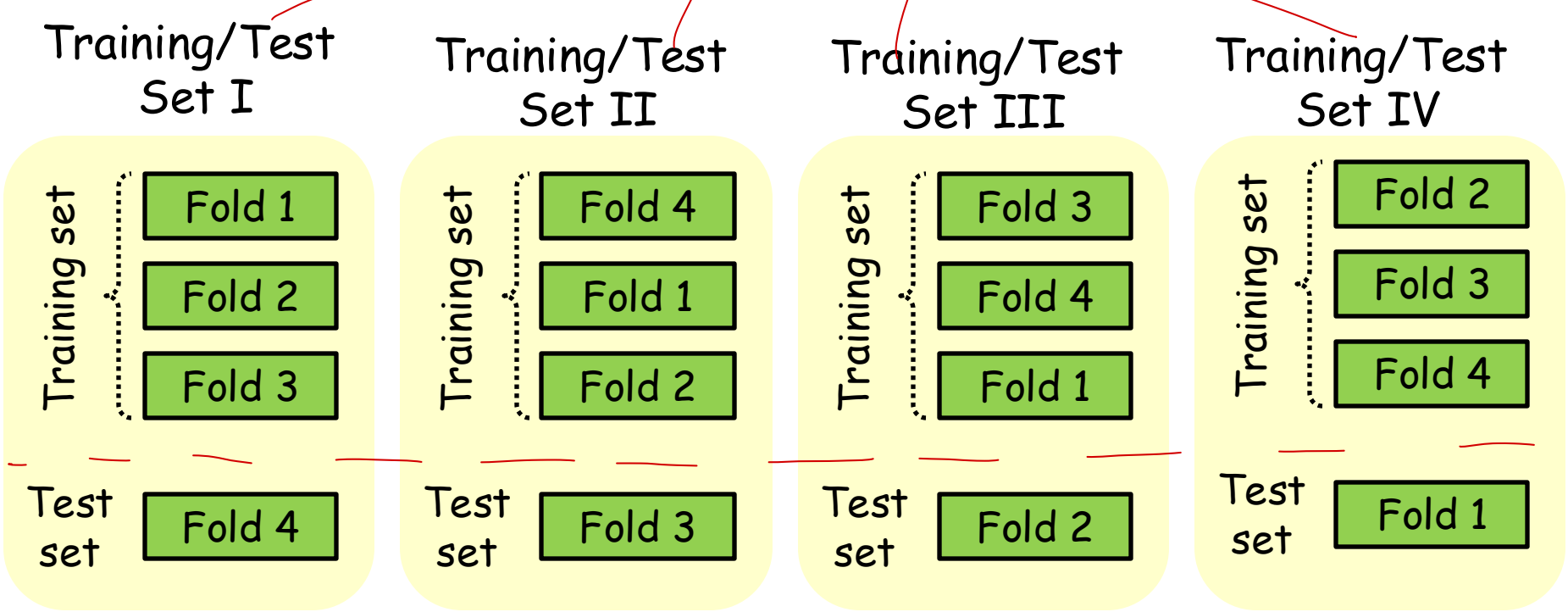
- **k-fold Cross Validation**
  - Split given data into *K* folds
  - Folds should not overlap with each other



  - Compose *k* different set of training and test datasets

Evaluation Method: Cross Validation

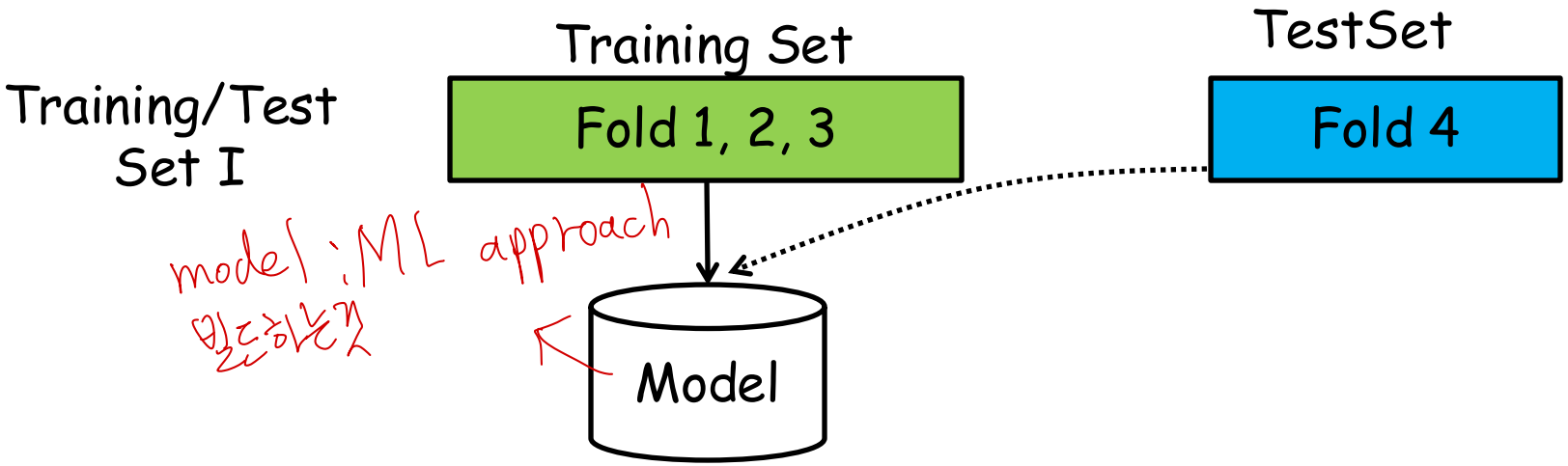
- **Example: 4 fold cross validation**



Evaluate a model by the average performance of 4 sets

Evaluation Method: Cross Validation

- **Example: 4 fold cross validation**



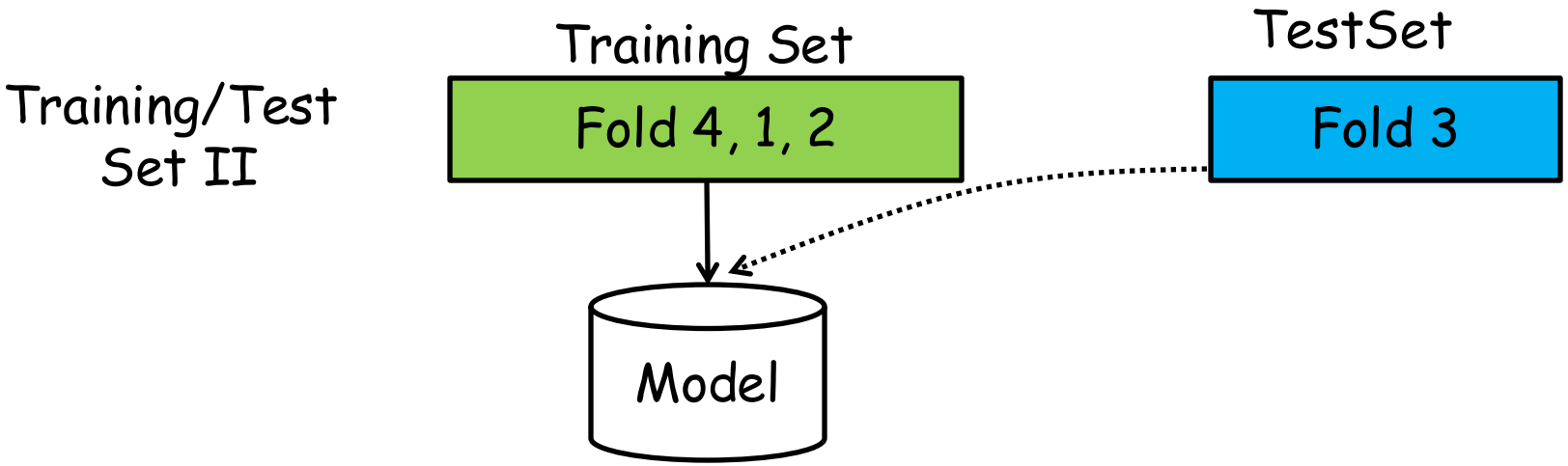
Set I	90%
Set II	
Set III	
Set IV	

성능 계산

Test set evaluation

Evaluation Method: Cross Validation

- **Example: 4 fold cross validation**

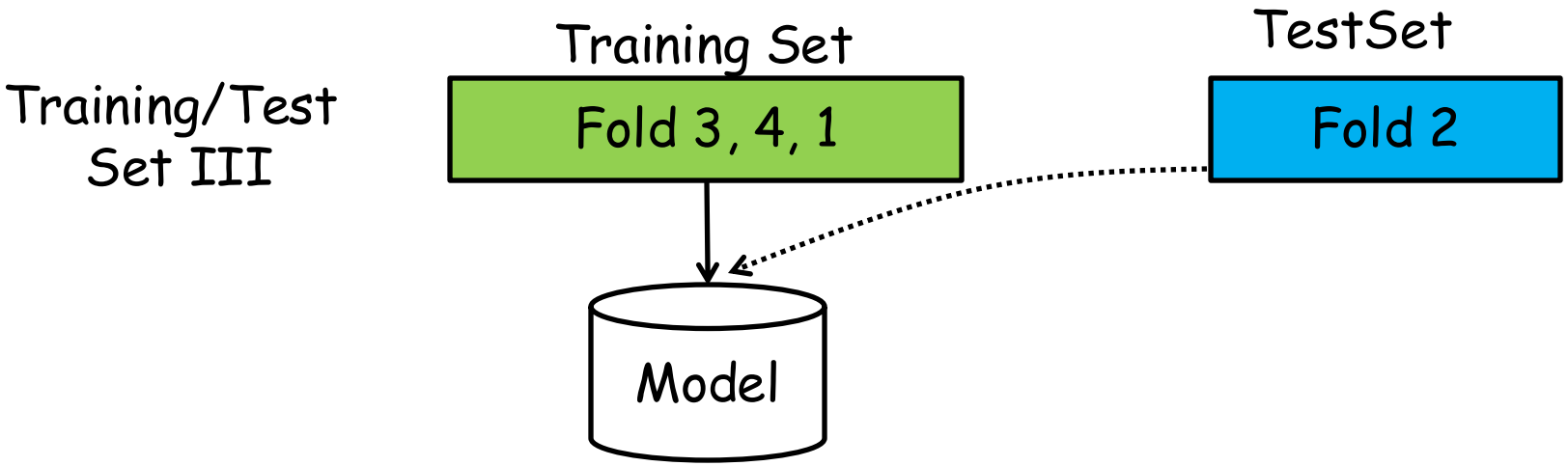


Set I	90%
Set II	95%
Set III	
Set IV	

Test set evaluation

Evaluation Method: Cross Validation

- **Example: 4 fold cross validation**

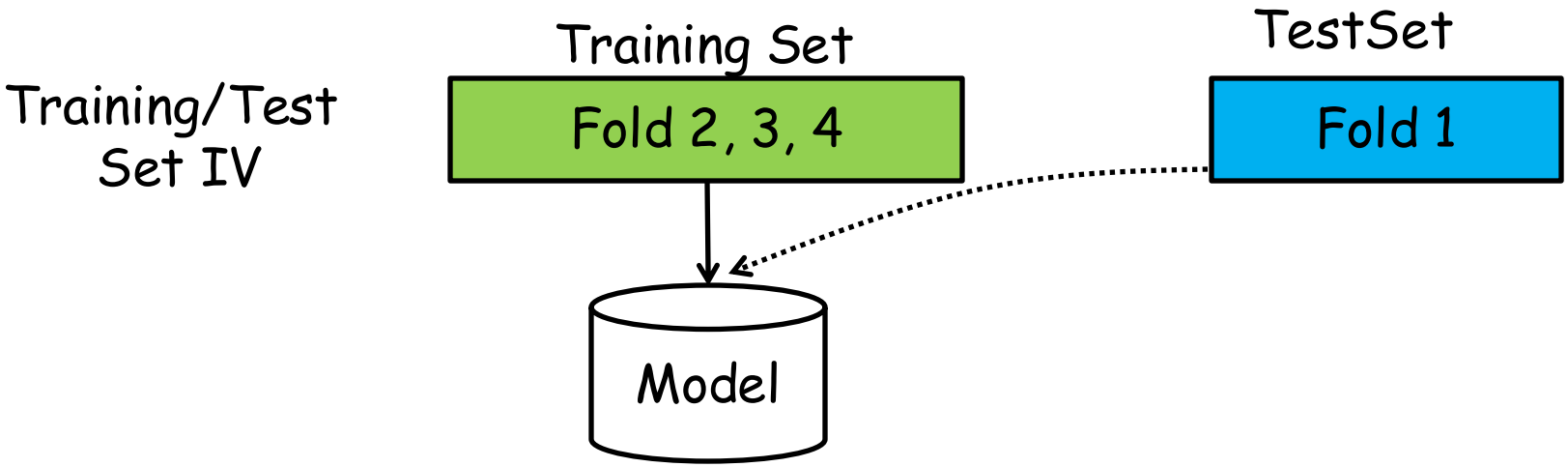


Set I	90%
Set II	95%
Set III	85%
Set IV	

Test set evaluation

Evaluation Method: Cross Validation

- **Example: 4 fold cross validation**



Set I	90%
Set II	95%
Set III	85%
Set IV	90%

avg = 90%

Test set evaluation

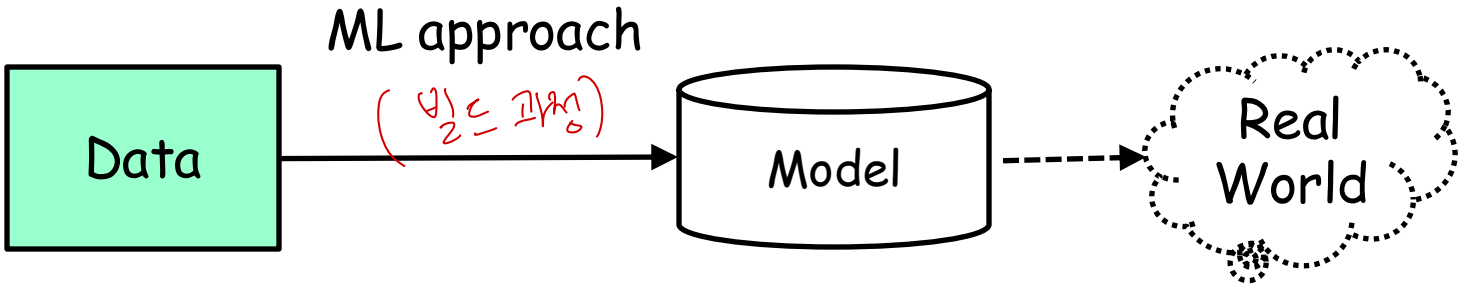
Evaluation Method: Cross Validation

- **Summary**
  - The data set is divided into *k* subsets, and the hold-out method is repeated *k* times.
  - Then the average error across all *k* trials is computed.
  - The variance is reduced as *k* is increased.
- **Advantage & Disadvantage**
  - Less dependent on how the data gets divided.
  - Every data point gets to be in a test set exactly once, and gets to be in a training set k-1 times.
  - But, time consuming
- **Variation**
  - No split into folds and Randomly dividing the data into a test and training set k different times.

Model Selection & Evaluation

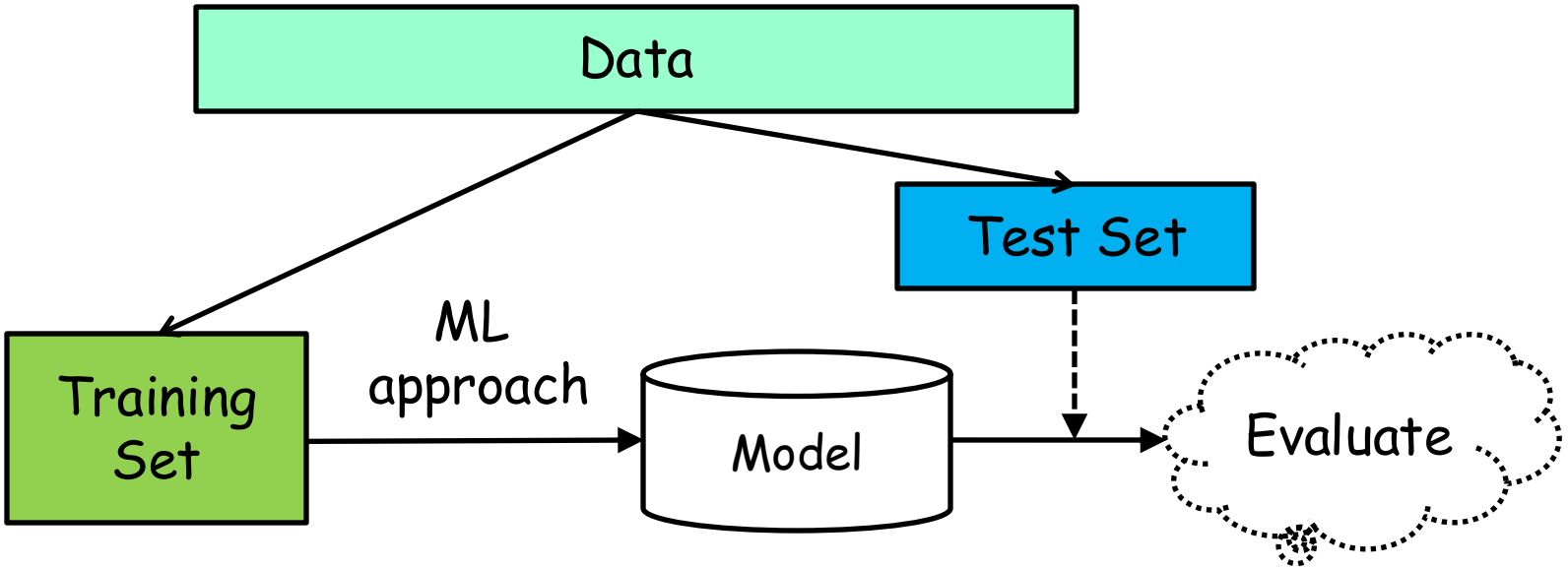
- **How different "Model Evaluation" and "Model Selection & Evaluation"?**

- **Recap: Model Evaluation**
  - I want to build a model by an ML approach
  - I wonder how good model the approach can build?



Model Selection & Evaluation

- **Recap: Model Evaluation**
  - Step1: Split the data into Training Set and Test Set
  - Step2: Build a model by the approach using Training Set
  - Step3: Evaluate the model using Test Set

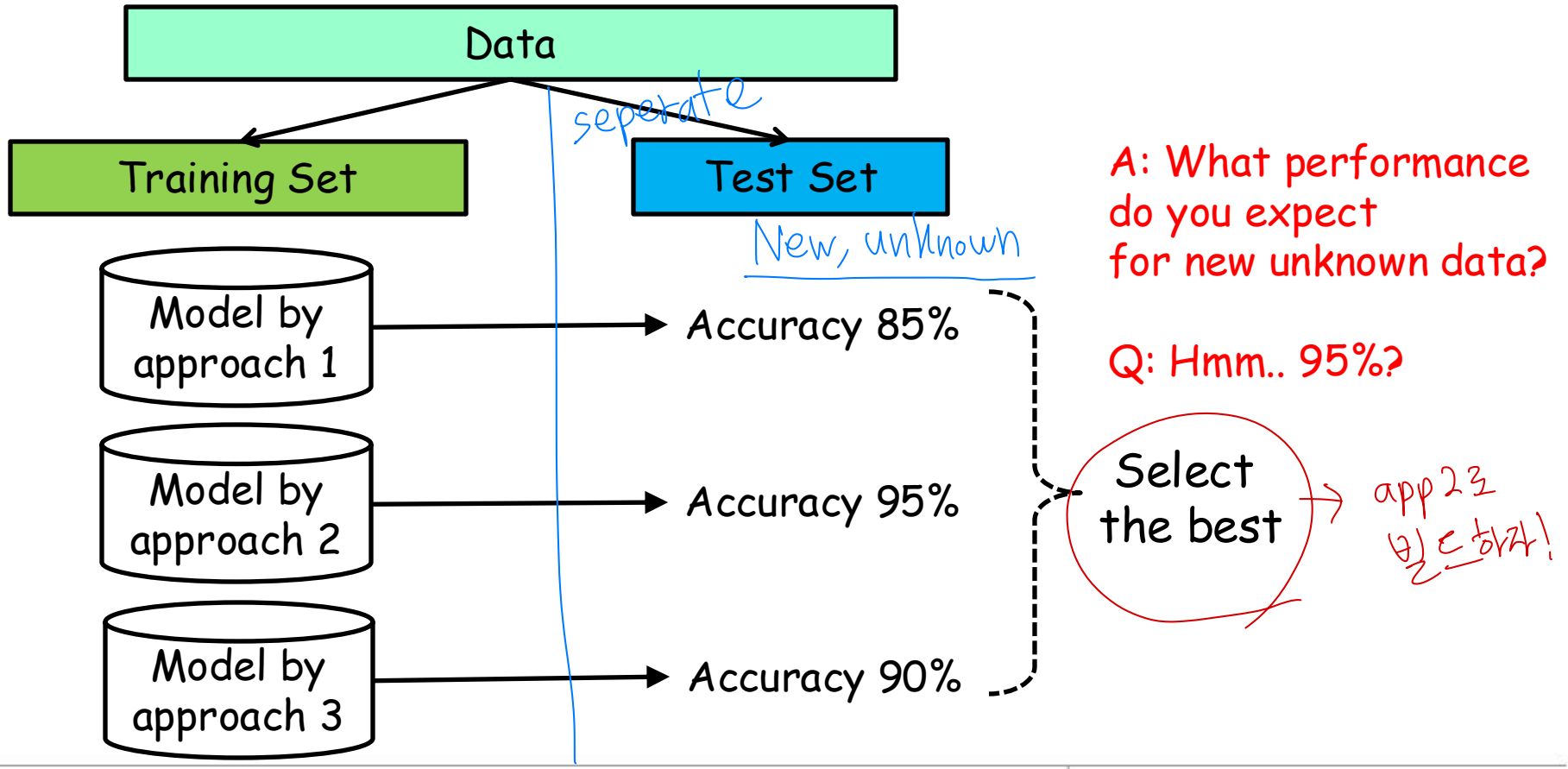


Model Selection & Evaluation

- **Model Selection & Evaluation**
  - I want to choose the best among several ML approaches
  - I want to build a model by it
  - I wonder how good model the best approach can build?
- **Contents**
  - Data Split
  - Model Selection & Evaluation by Hold-out
  - Model Selection & Evaluation by CV

Model Selection & Evaluation

- **Easy(?): Repeat Model Evaluation Process**

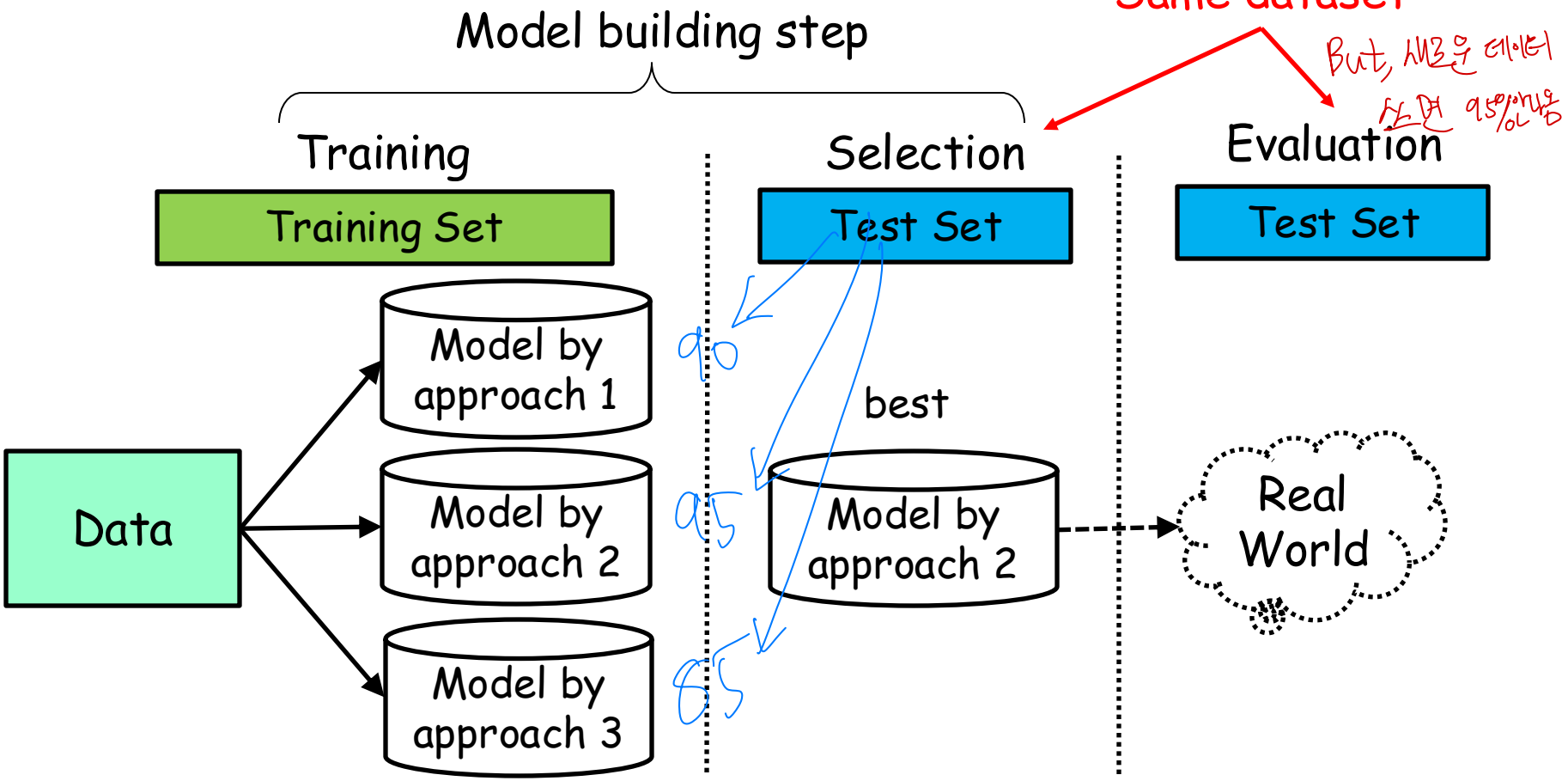


Model Selection & Evaluation

- **Questions on the easy way**
  - Q1: Can we say that Model 2 is the best?
  - Q2: Can we say that Model 2's expected performance on new unknown data is 95% in average?
  - Q3: Can we say that the best model's expected performance on new unknown data is 95% in average?
- **Answers**
  - A1: Yes, in the viewpoint of "test set"
  - A2: Yes, because it is evaluated with a randomly split dataset
  - A3: No.. Why?
    - The best model is not randomly chosen
    - It is chosen in the viewpoint of given "test set"
    - Its result may be biased on the given set

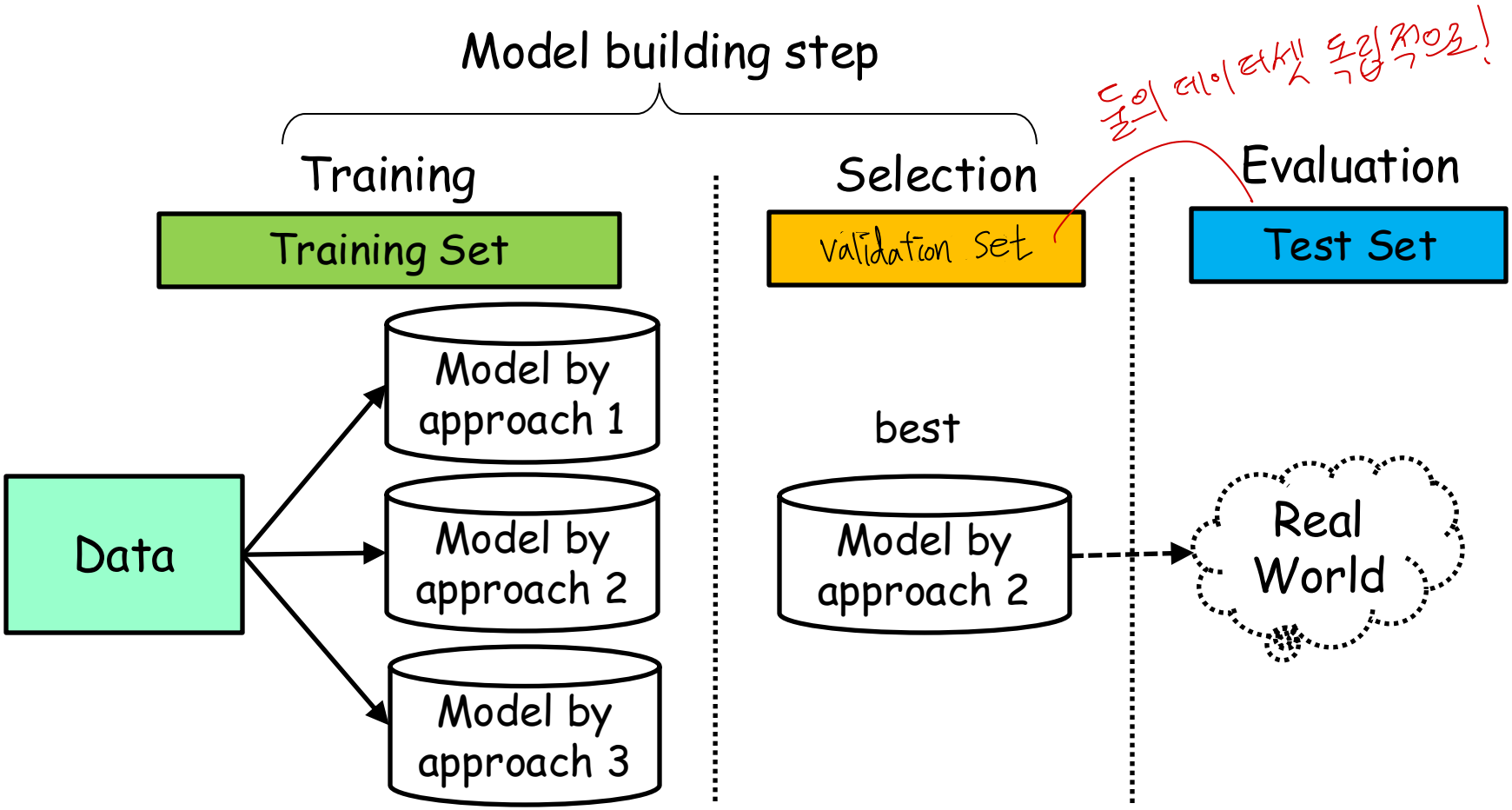
Model Selection & Evaluation

- **The problem of the easy way**



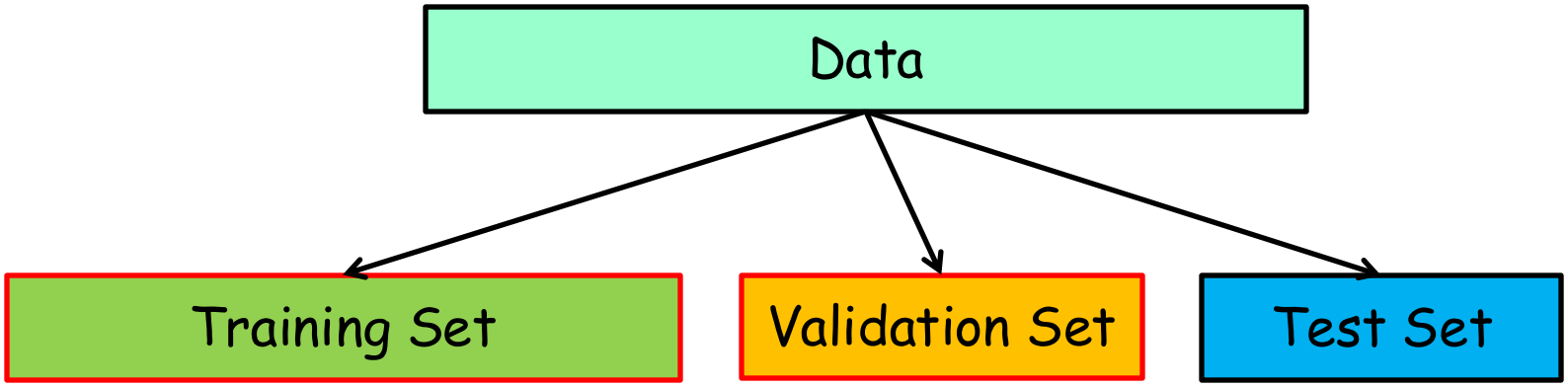
Model Selection & Evaluation

- **We have to make each step independent**



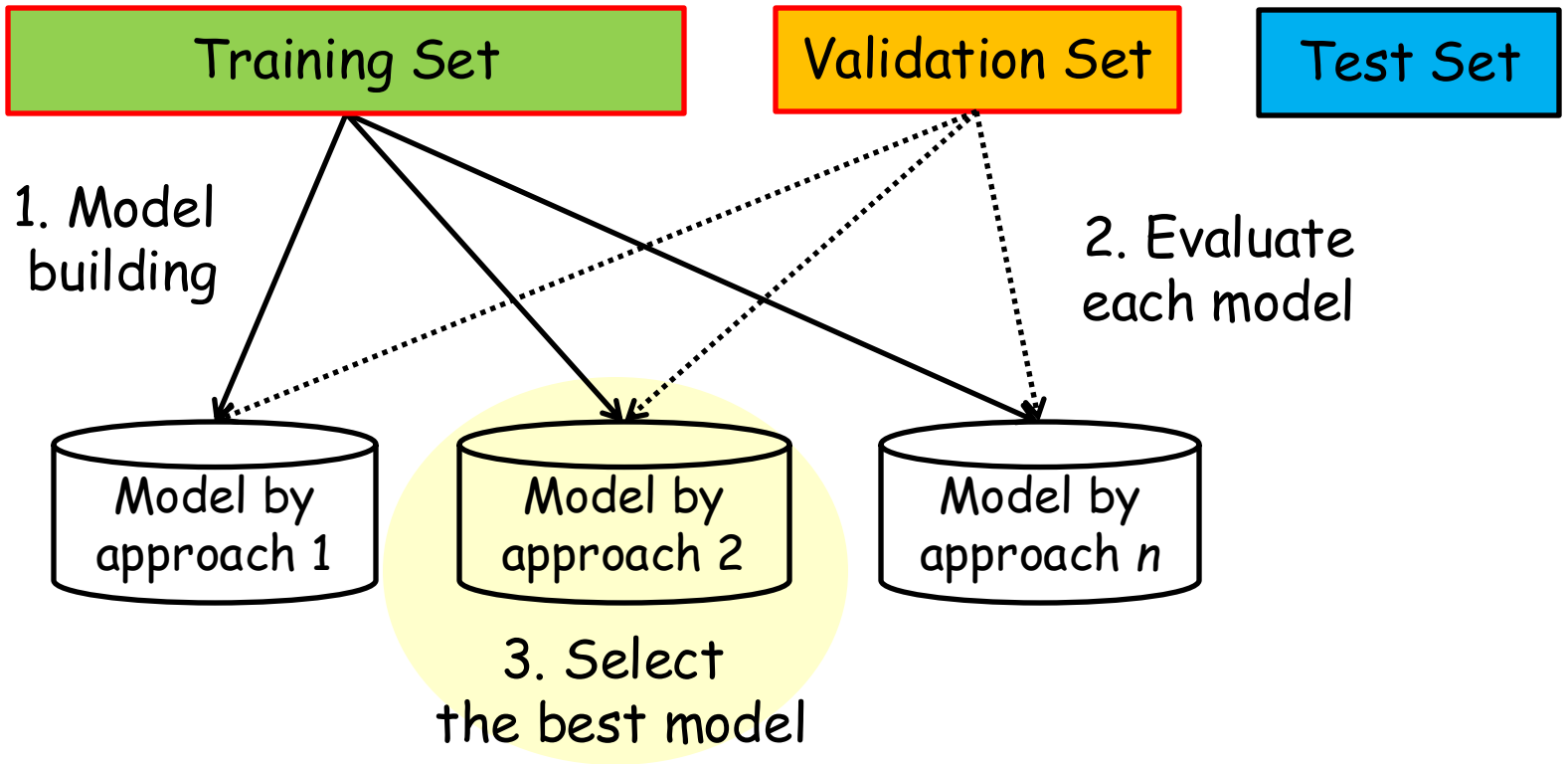
Model Selection & Evaluation

- **Data Split for Model Selection & Evalutioan**
  - We need to separate training set, selection set for the best, and evaluation set



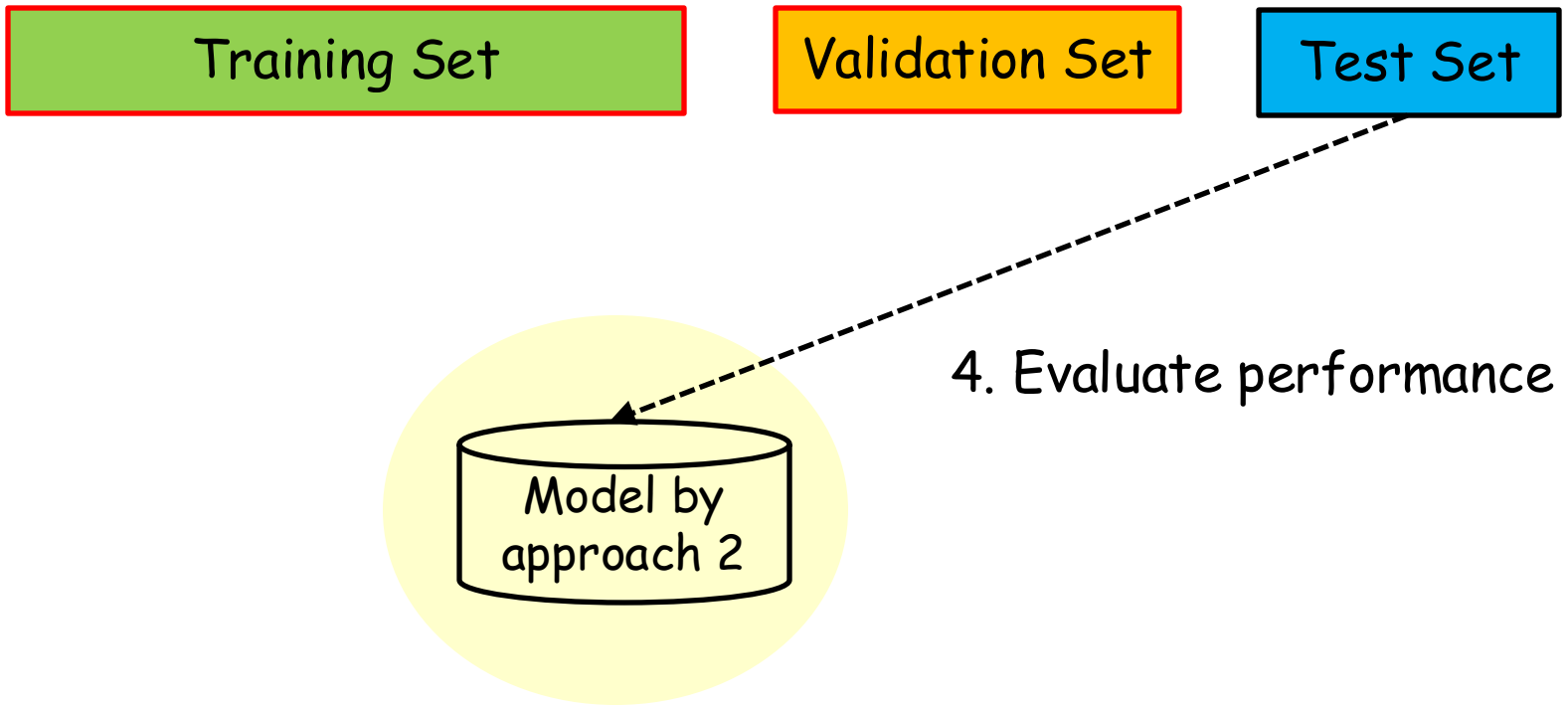
Model Selection & Evaluation

- **Steps using Hold-out**
  - Step1: Model Building & Selection



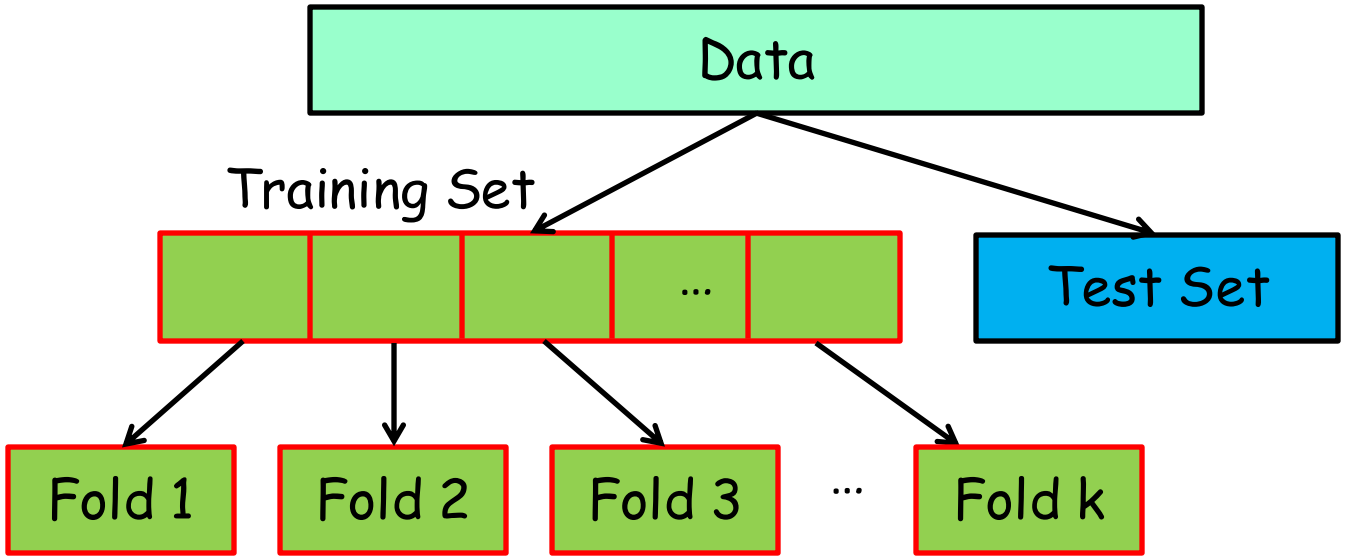
Model Selection & Evaluation

- **Steps using Hold-out**
  - Step2: Evaluating the chosen model



Model Selection & Evaluation

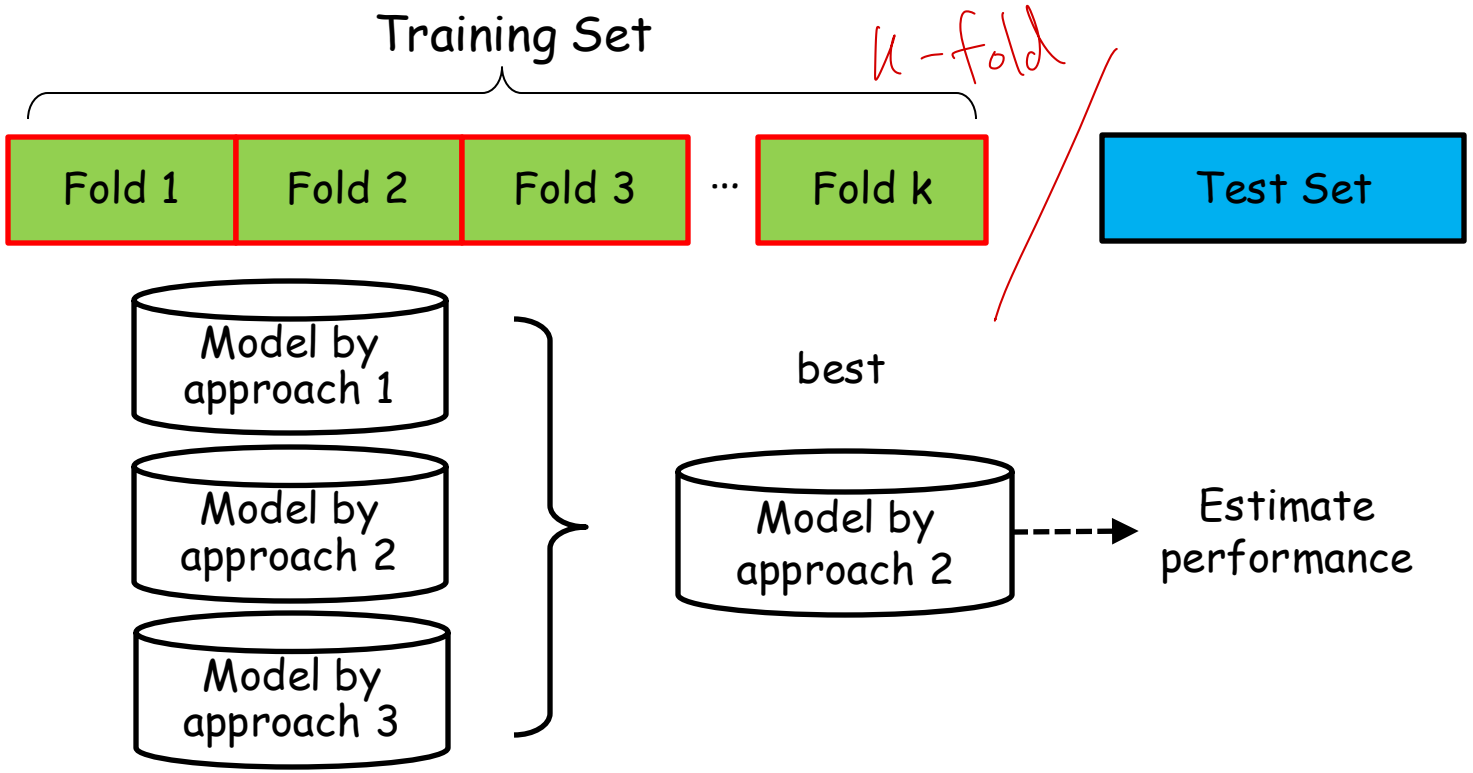
- **Steps using Cross Validation**
  - Step 1: Data split into Training and Test set



We will use one fold as the validation set

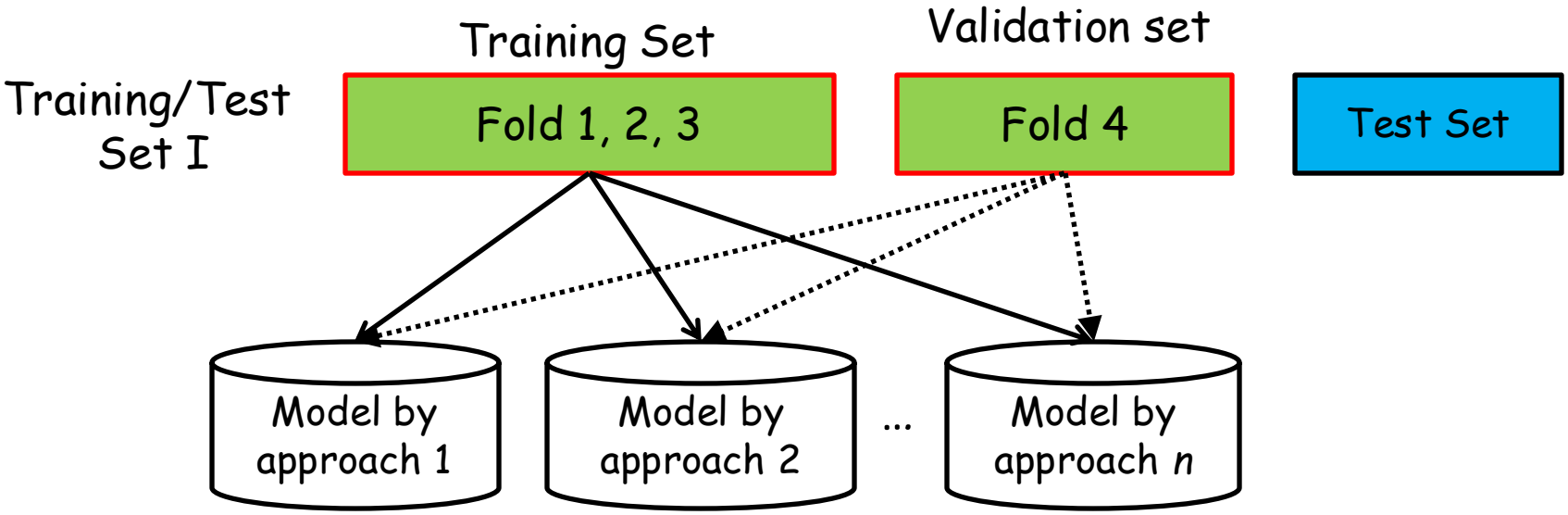
Model Selection & Evaluation

- **Steps using Cross Validation**
  - Step 2: Build models and Choose the best (by CV)
  - Step 3: Evaluate the chosen model



Model Selection & Evaluation

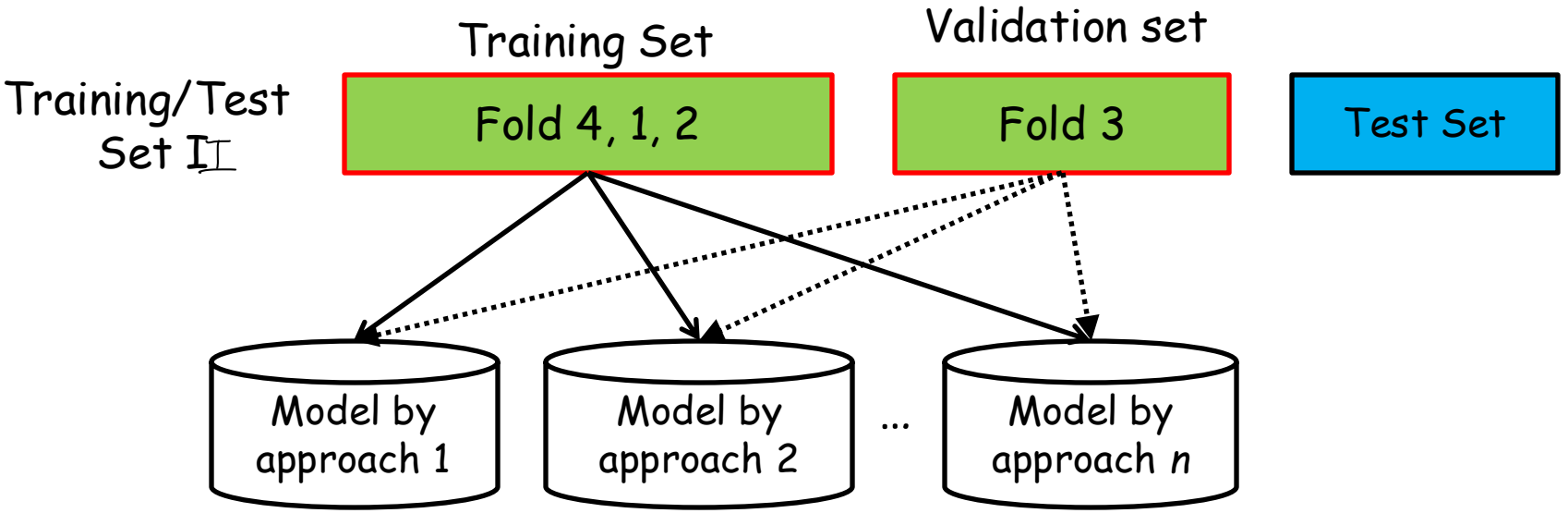
- **Example: 4 fold cross validation**



	Model 1	Model 2		Model n
Set I	90%	90%	....	85%
Set II				
Set III				
Set IV				

Model Selection & Evaluation

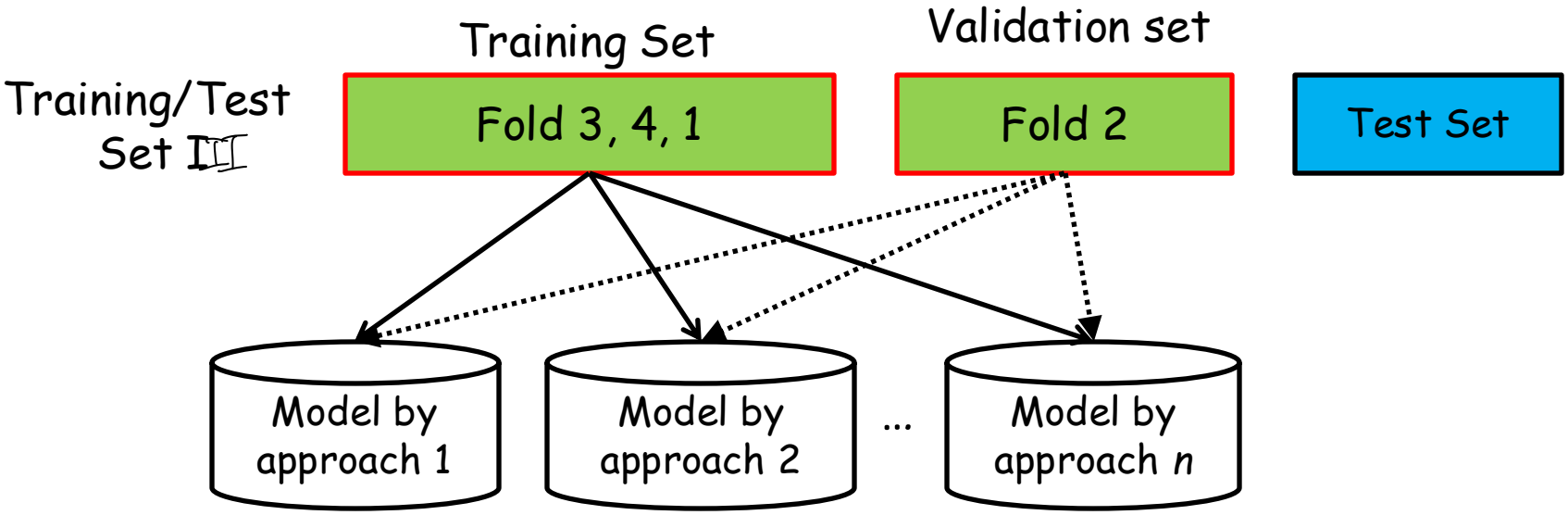
- **Example: 4 fold cross validation**



	Model 1	Model 2		Model n
Set I	90%	90%	....	85%
Set II	90%	95%	....	90%
Set III				
Set IV				

Model Selection & Evaluation

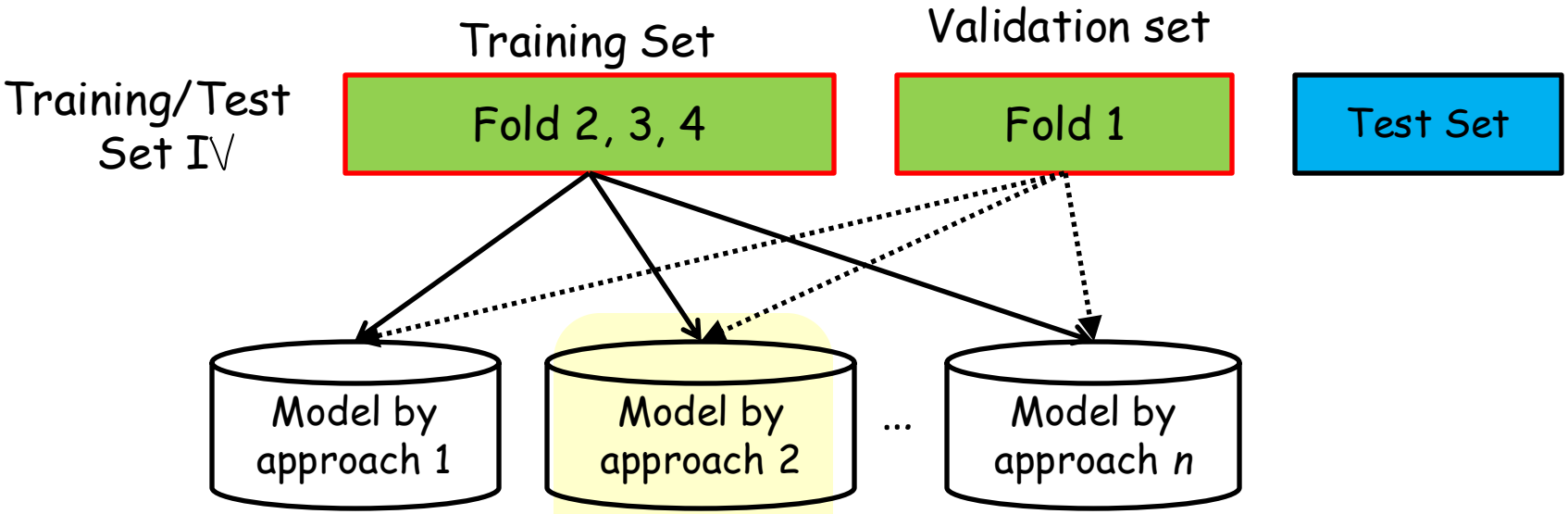
- **Example: 4 fold cross validation**



	Model 1	Model 2		Model n
Set I	90%	90%	....	85%
Set II	90%	95%	....	90%
Set III	95%	95%	....	90%
Set IV				

Model Selection & Evaluation

- **Example: 4 fold cross validation**

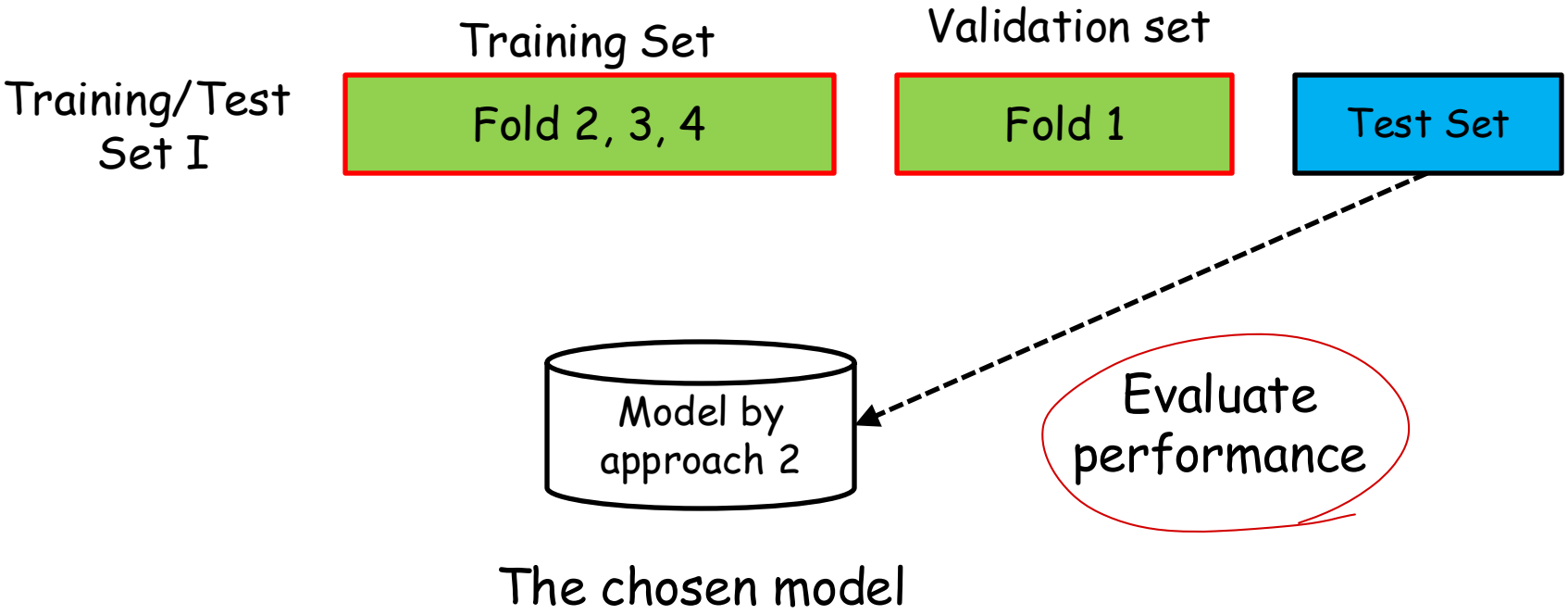


	Model 1	Model 2		Model n
Set I	90%	90%	....	85%
Set II	90%	95%	....	90%
Set III	95%	95%	....	90%
Set IV	90%	90%	....	80%

Choose the best

Model Selection & Evaluation

- **Example: 4 fold cross validation**



Summary

- **Model Building and Evaluation**
  - Training set: Build a model
  - Test set: Evaluate the model

- **Model Selection and Evaluation**
  - Training set: Build several models
  - Validation set: Choose the best
  - Test set: Evaluate the best model