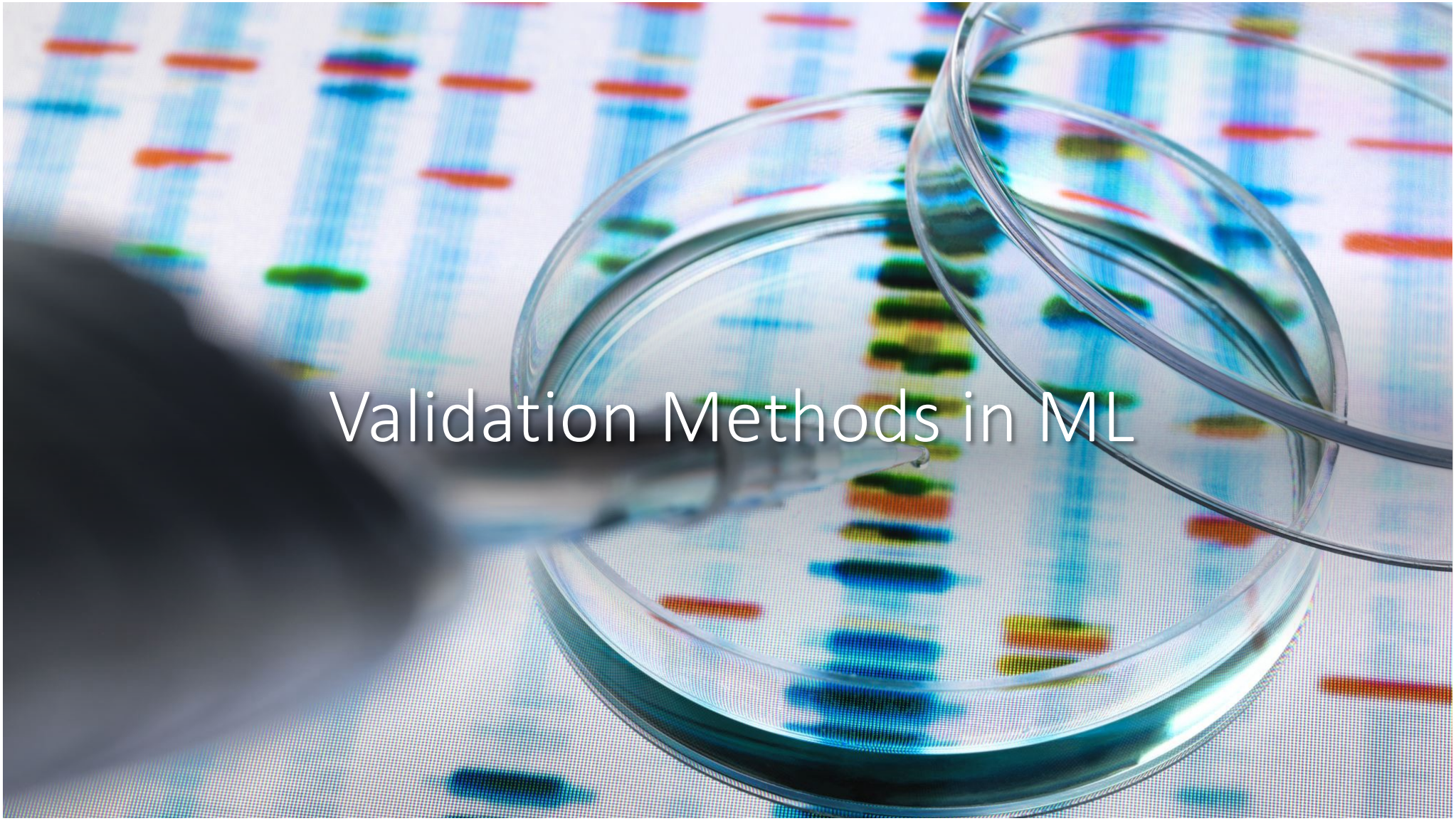

Validation Methods in ML

Goals

- Understand the basic concepts of resampling methods for machine learning model selection and assessment.
- Understand the similarities and differences between the validation set approach, leave-one-out cross-validation, and *K*-fold cross-validation.
- Understand the basic theoretical concepts of the bootstrap method for assessing statistical accuracy.
- Recognize that the performance of machine learning models depends on their prediction capability on independent test data.

Model Assessment

The *generalization* performance of a machine learning method relates to its prediction capability on independent test sets.

Assessment of this performance is extremely important in practice, since it guides the choice of the machine learning method or model.

Further, this gives us a measure of the quality of the ultimately chosen model.

Model assessment

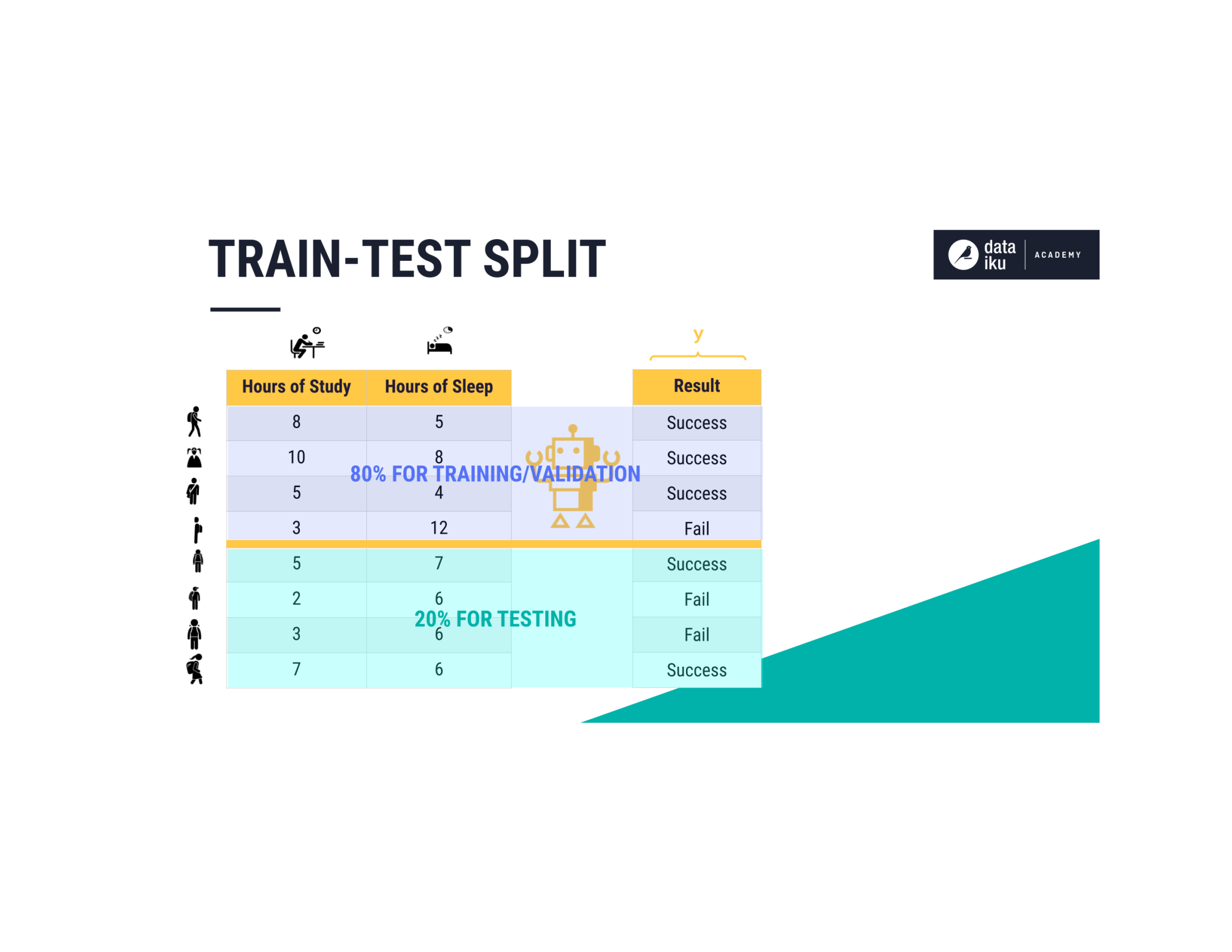
- *Test Error*
  - The average error that results from using a machine learning method to predict the response on a new observation.
  - The prediction error over an independent test sample.

- *Training Error*
  - The average loss over the training sample: $\overline{\text{err}} = \frac{1}{N}\sum_{i=1}^{N} L(y_i, \hat{f}(x_i))$

- **Note:** The training error rate can dramatically *underestimate* the test error rate


Test Train Validation

TRAIN-TEST SPLIT

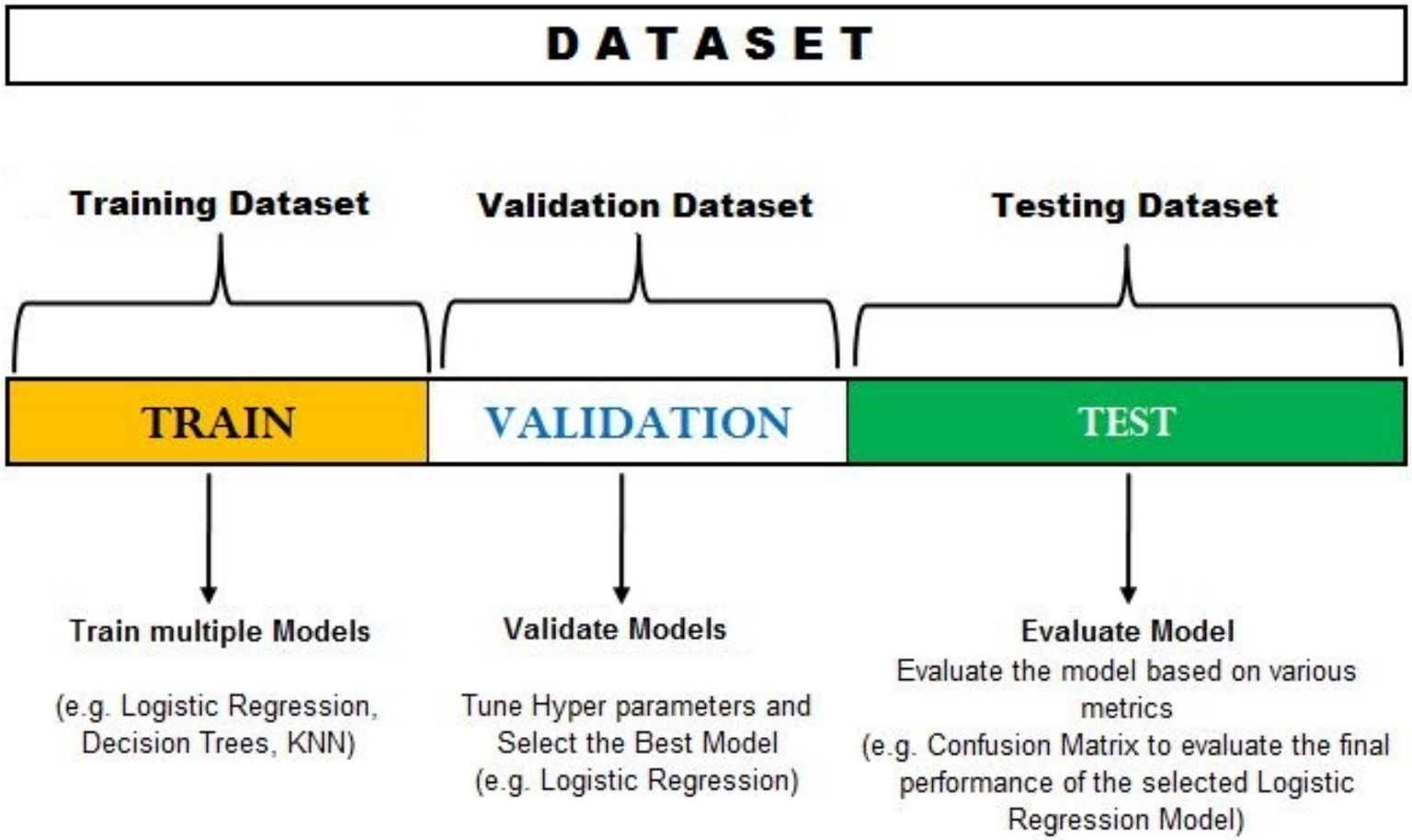


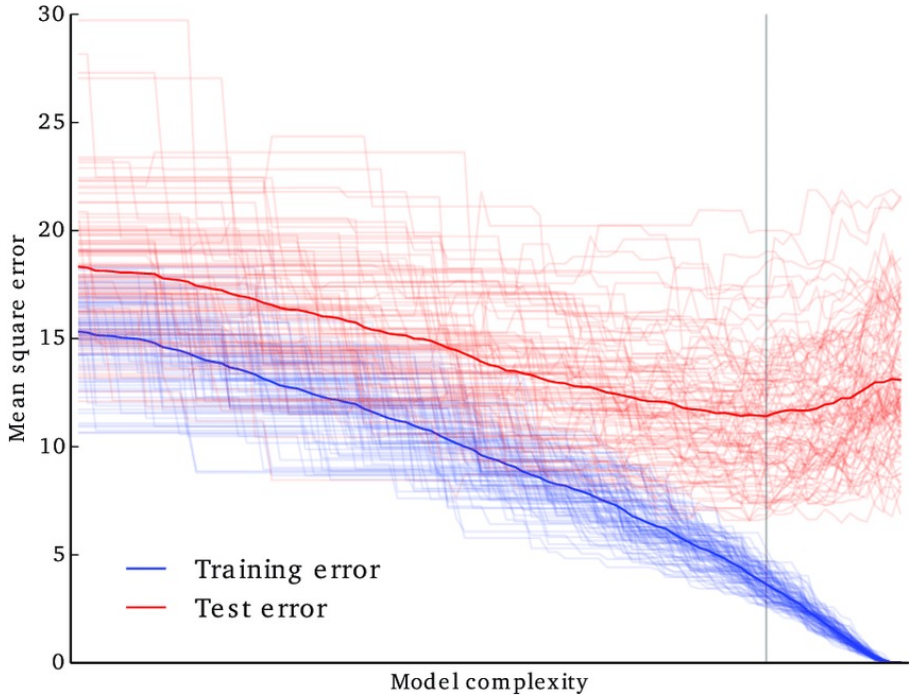
y

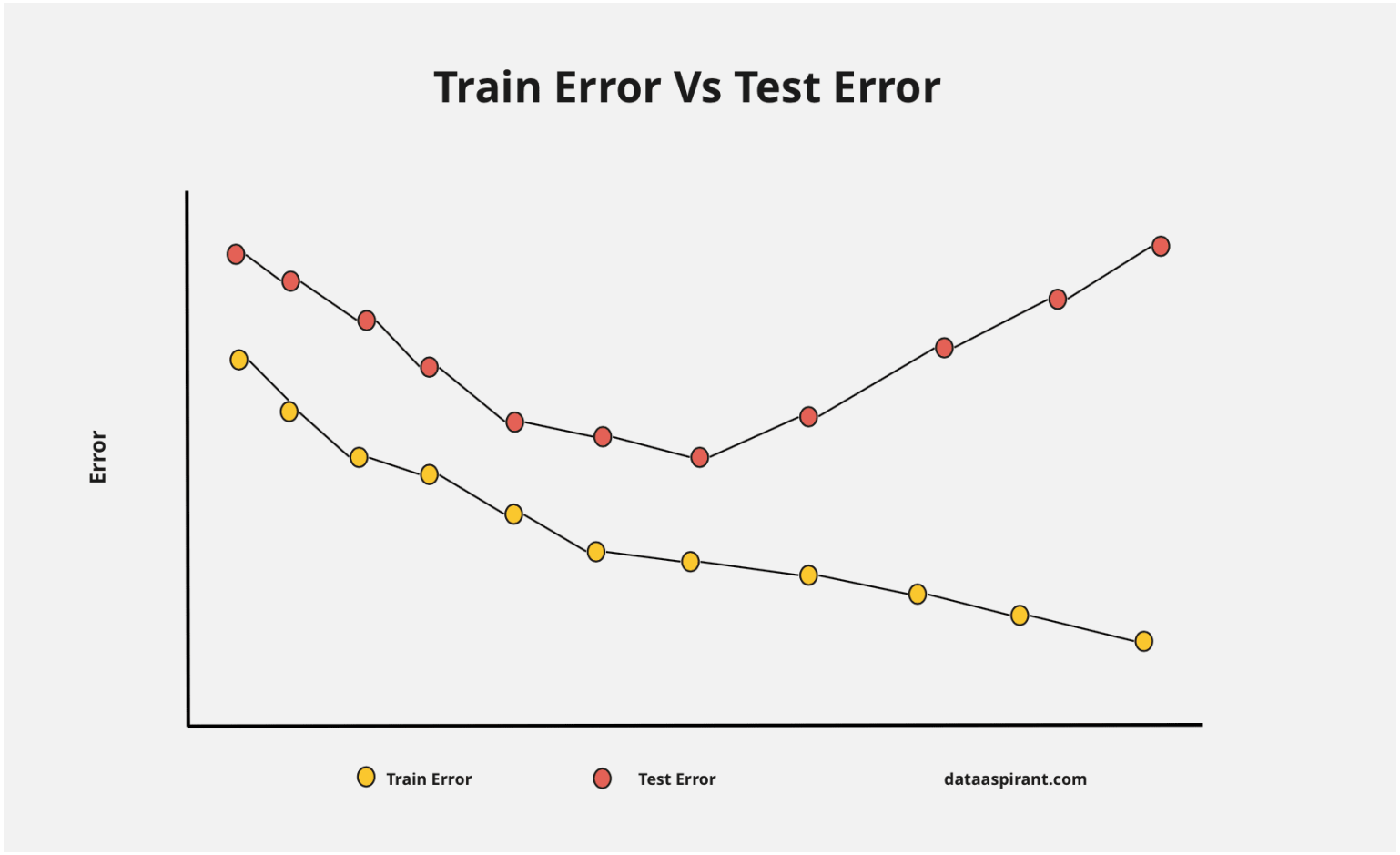
Hours of Study	Hours of Sleep	Result
8	5	Success
10	8	Success
5	4	Success
3	12	Fail
5	7	Success
2	6	Fail
3	6	Fail
7	6	Success

80% FOR TRAINING/VALIDATION

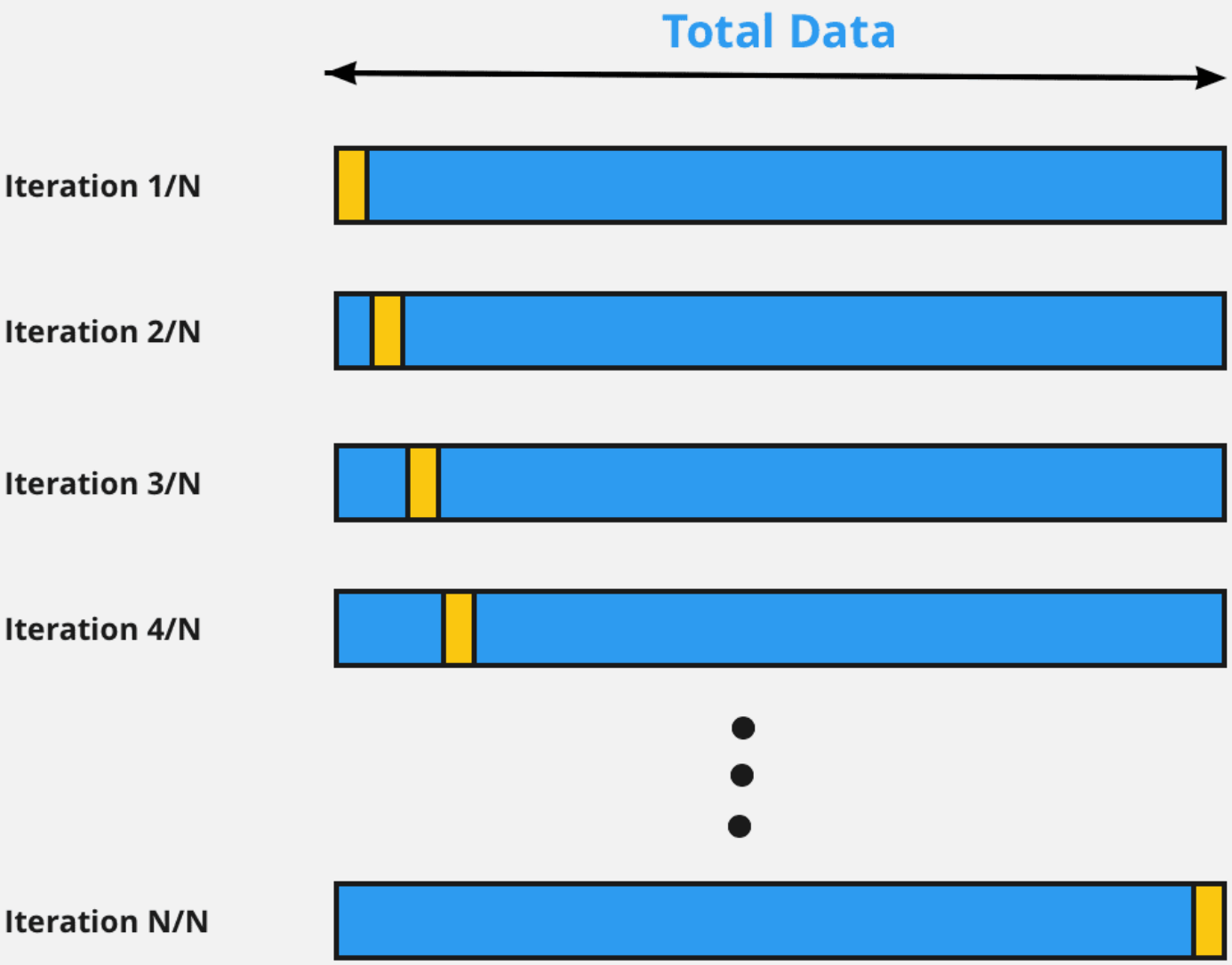
20% FOR TESTING


DATASET
Training Dataset
Validation Dataset
Testing Dataset
TRAIN
VALIDATION
TEST
Train multiple Models
(e.g. Logistic Regression, Decision Trees, KNN)
Validate Models
Tune Hyper parameters and Select the Best Model
(e.g. Logistic Regression)
Evaluate Model
Evaluate the model based on various metrics
(e.g. Confusion Matrix to evaluate the final performance of the selected Logistic Regression Model)

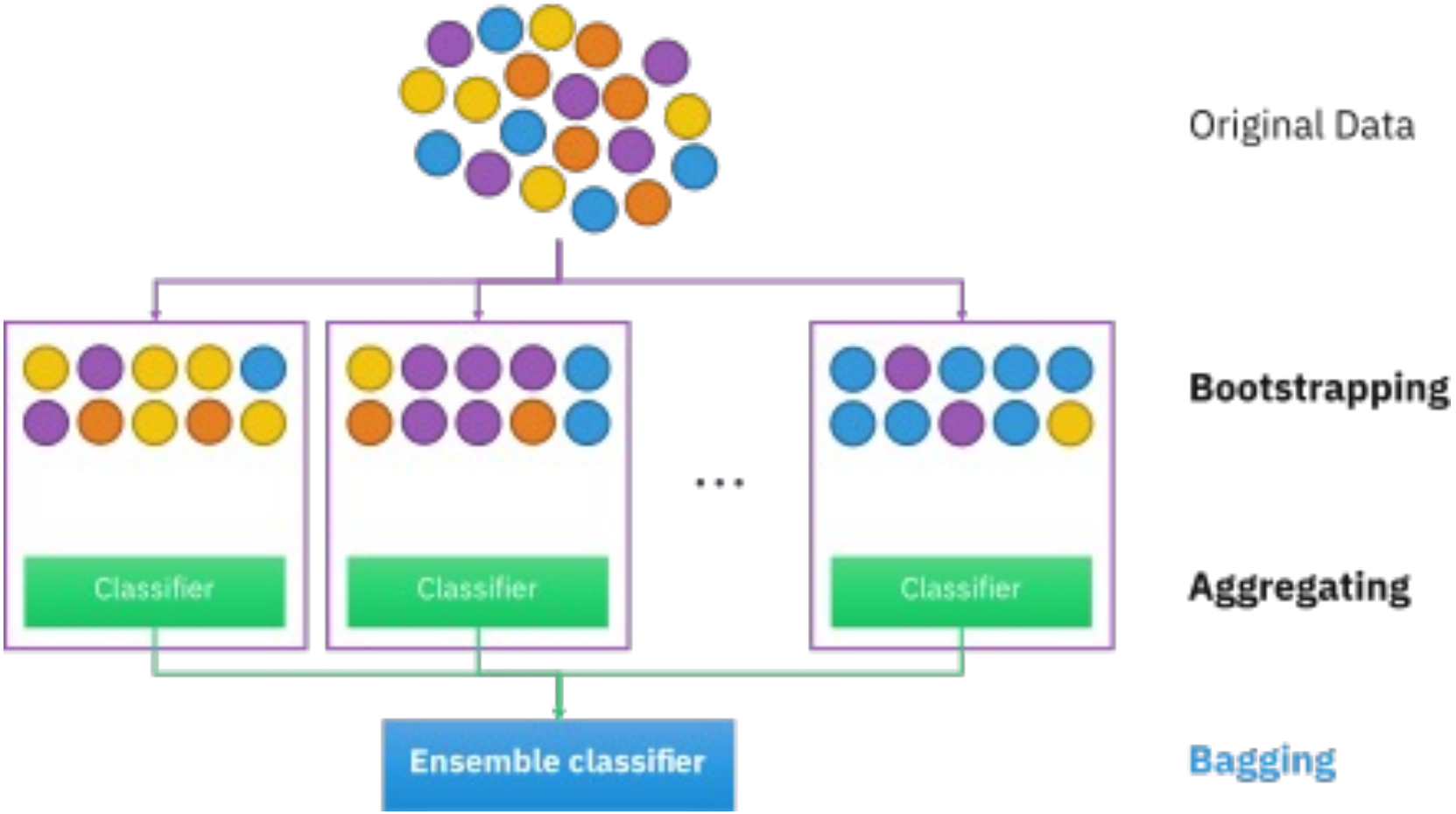

30
25
20
15
10
5
0
Mean square error
Training error
Test error
Model complexity


Train Error Vs Test Error
Error
Train Error
Test Error
dataaspirant.com


Leave One Out Cross Validation

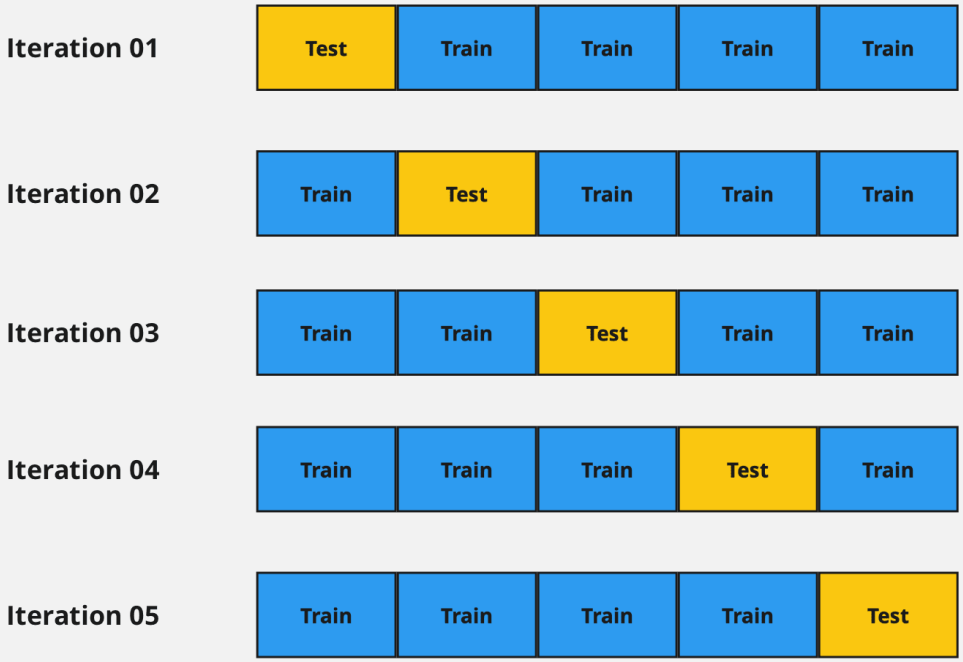

Total Data
Iteration 1/N
Iteration 2/N
Iteration 3/N
Iteration 4/N
Iteration N/N


Bootstrap


Original Data
Bootstrapping
Classifier
Classifier
Classifier
Aggregating
Ensemble classifier
Bagging

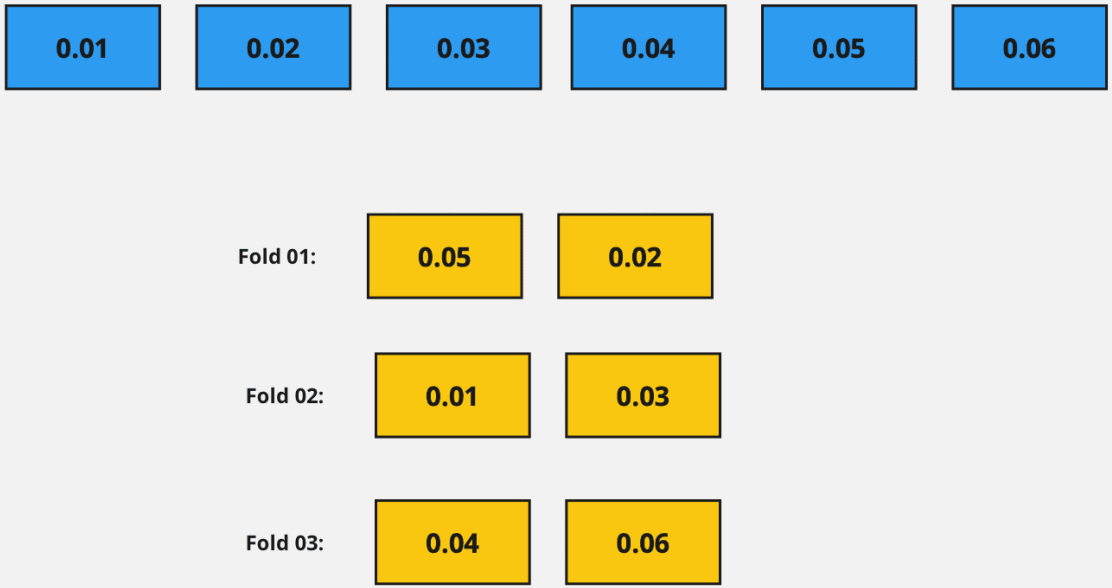

K fold Cross-Validation

K-Fold Cross Validation

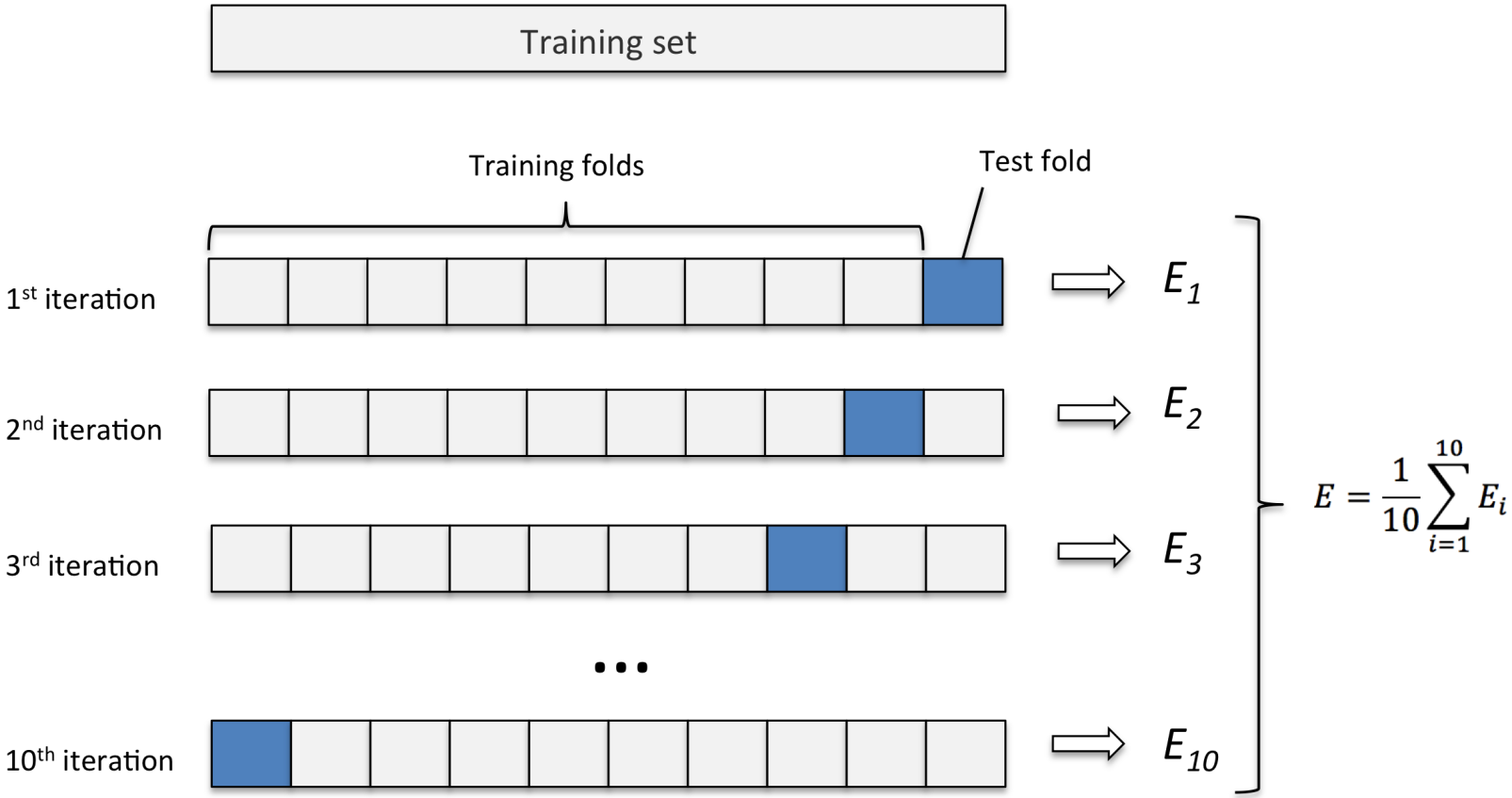


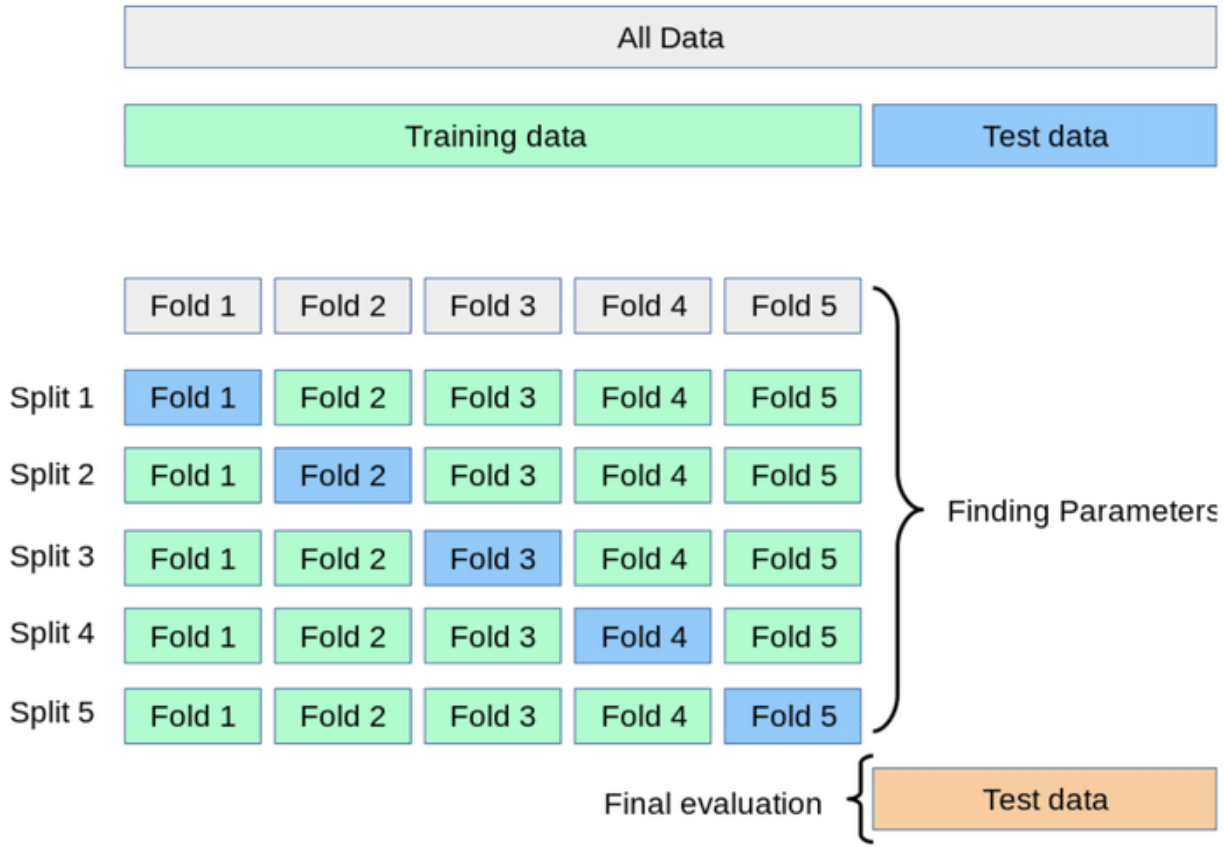
dataaspirant.com

K-Fold Cross Validation Example



dataaspirant.com

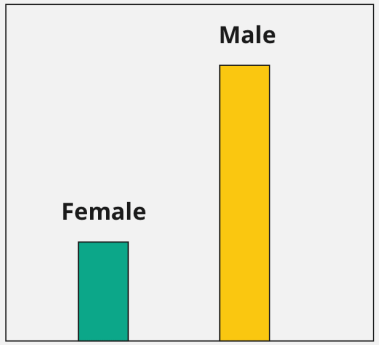

Training set
Training folds
Test fold
1st iteration
2nd iteration
3rd iteration
10th iteration
E1
E2
E3
E10
$E = \frac{1}{10}\sum_{i=1}^{10} E_i$

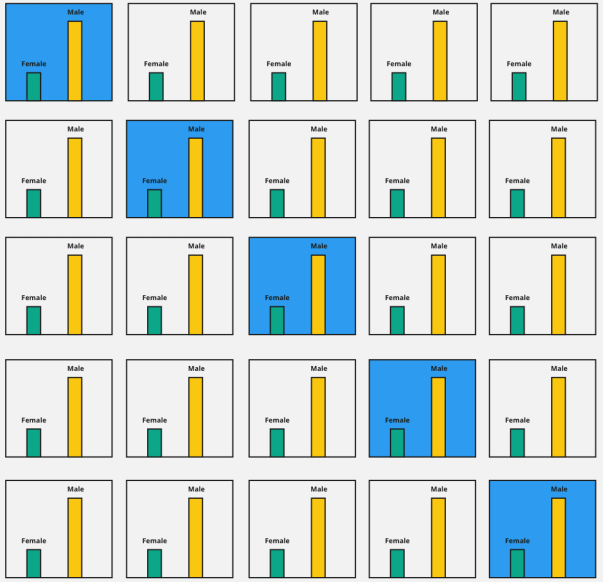


K-Fold Cross Validation

Stratified K-Fold Cross Validation

Target Class Distribution





dataaspirant.com

Summary

Resampling methods for machine learning model selection and assessment.

The validation set approach, leave-one-out cross-validation, and *K*-fold cross-validation.

The bootstrap method for assessing statistical