CLASSIFICATION-BASED

MODEL EVALUATION

MODEL EVALUATION

- Performance Evaluation
- Model Comparison

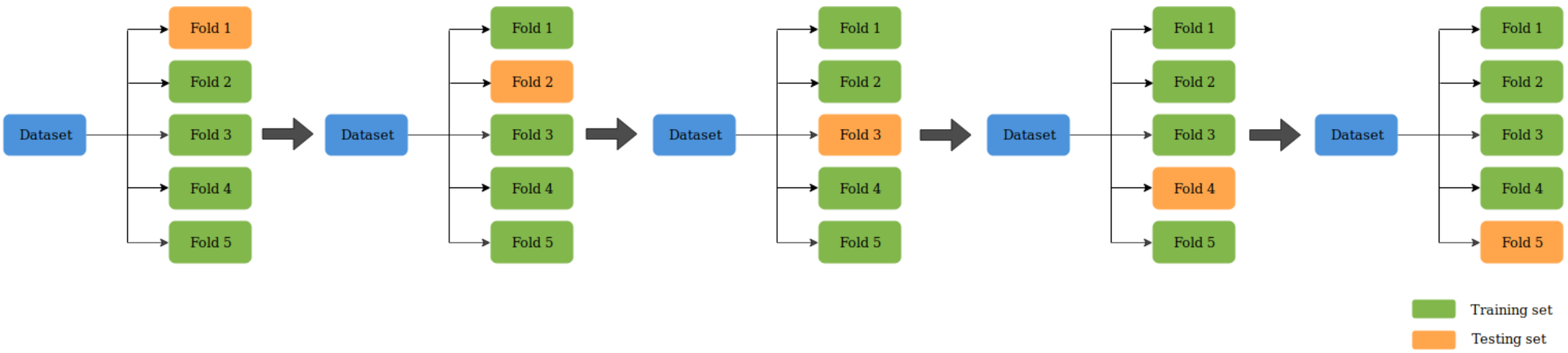
PERFORMANCE EVALUATION

- Cross-Validation

  Preferably, *k*-fold cross-validation

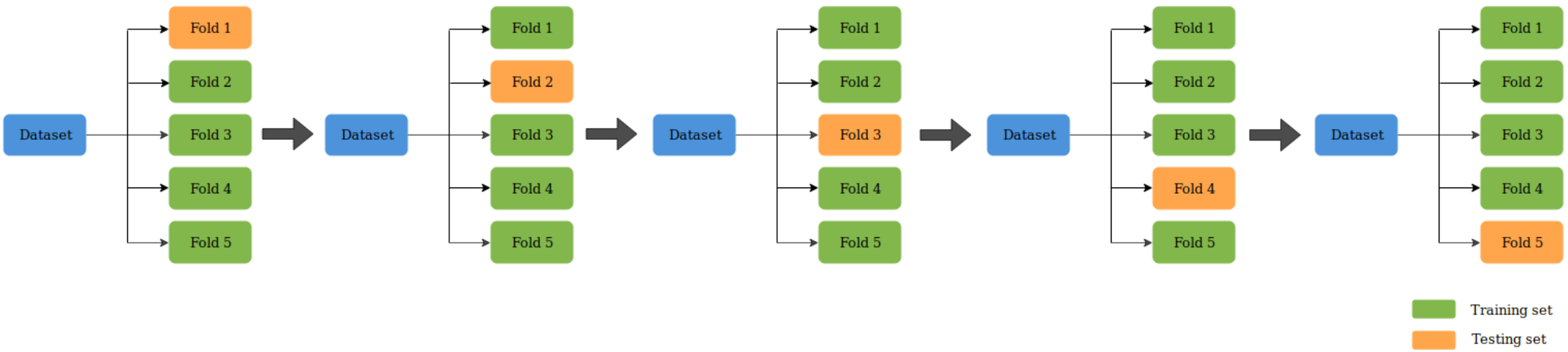
- Train-Test Accuracy

K-FOLD CROSS-VALIDATION



https://medium.com/datadriveninvestor/k-fold-cross-validation-6b8518070833

K-FOLD CROSS-VALIDATION



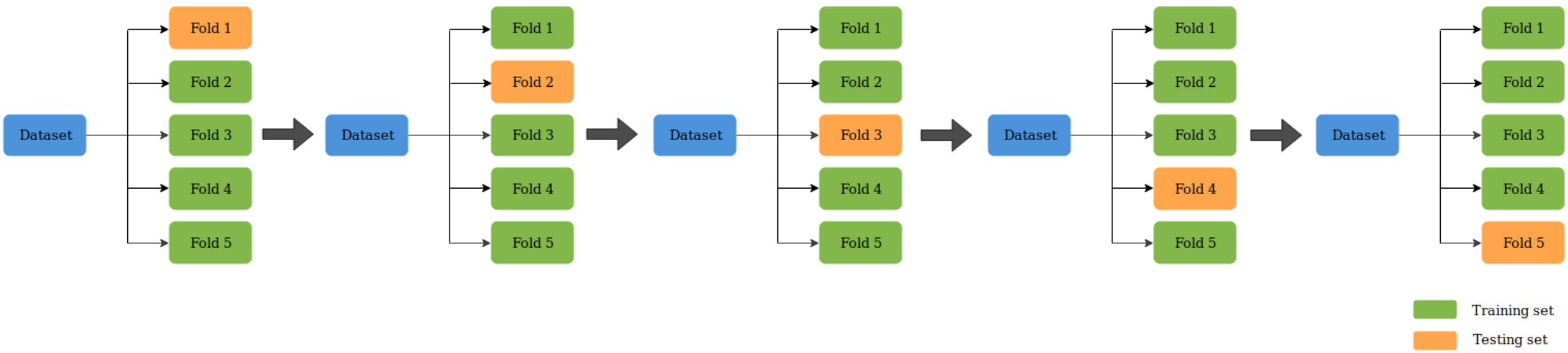
- Bias-Variance Trade-off:

  Reduces bias but increases variance as we increase the number of folds

  A good empirical value for $k = 10$

https://medium.com/datadriveninvestor/k-fold-cross-validation-6b8518070833

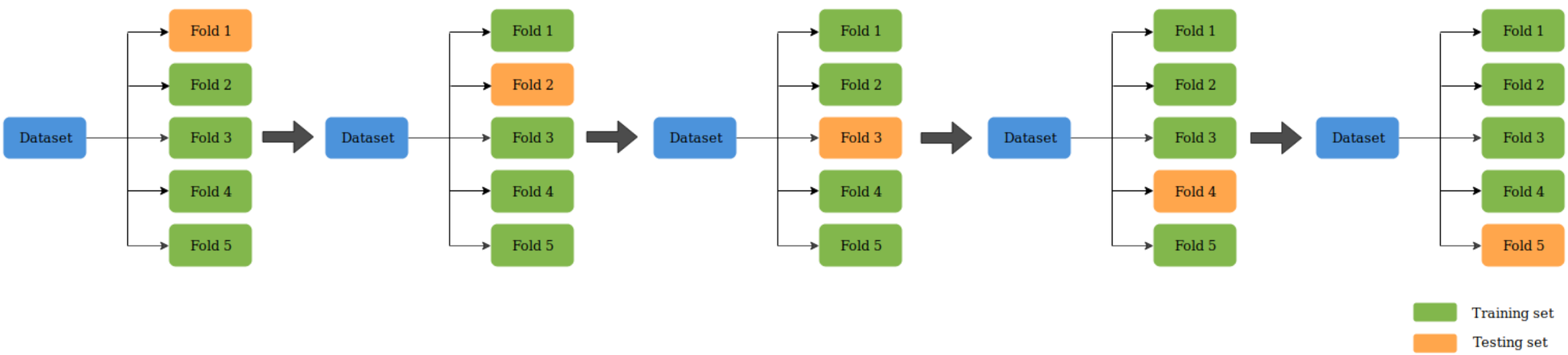
OTHER TYPES OF CROSS-VALIDATION



- Repeated *k*-fold
- Stratified *k*-fold
- Leave-One-Out

https://medium.com/datadriveninvestor/k-fold-cross-validation-6b8518070833

OTHER TYPES OF CROSS-VALIDATION



- Repeated *k*-fold: Reduces variance by averaging
- Stratified *k*-fold: Good for imbalanced classes
- Leave-One-Out: Better for smaller datasets

https://medium.com/datadriveninvestor/k-fold-cross-validation-6b8518070833

CROSS-VALIDATION BY SKLEARN

https://scikit-learn.org/stable/modules/cross_validation.html

GRID SEARCH IS YOUR NEW BEST FRIEND

- Grid Search is a wonderful way to find the right model and also tune all parameters
- Randomized Grid Search is better for a first-pass model
- Complete Grid Search can then be used to find the right configurations

GRID SEARCH IS YOUR NEW BEST FRIEND

- Grid Search is a wonderful way to find the right model and also tune all parameters
- Randomized Grid Search is better for a first-pass model
- Complete Grid Search can then be used to find the right configurations
- Of course, *sklearn* is here for our rescue! https://scikit-learn.org/stable/modules/grid_search.html

MODEL COMPARISON

- Gauging the Predictive Capability of a Model

		PREDICTED CLASS	
		CLASS = YES	CLASS = NO
ACTUAL CLASS	CLASS = YES	**a**	**b**
	CLASS = NO	**c**	**d**

MODEL COMPARISON

- Gauging the Predictive Capability of a Model

		PREDICTED CLASS	
		CLASS = YES	CLASS = NO
ACTUAL CLASS	CLASS = YES	**a** (True Positive)	**b** (False Negative)
	CLASS = NO	**c** (False Positive)	**d** (True Negative)

ACCURACY

- Total number of instances correctly classified

$$\text{Accuracy} = \frac{TP + FP}{TP + TN + FP + FN}$$

ACCURACY

- Total number of instances correctly classified

$$\text{Accuracy} = \frac{TP + FP}{TP + TN + FP + FN}$$

- The higher the accuracy, the better

ACCURACY

		PREDICTED CLASS	
		CLASS = YES	CLASS = NO
ACTUAL CLASS	CLASS = YES	**a** (True Positive)	**b** (False Negative)
	CLASS = NO	**c** (False Positive)	**d** (True Negative)

$$\text{Accuracy} = \frac{a + d}{a + b + c + d} = \frac{TP + FP}{TP + FN + FP + TN}$$

ACCURACY

		PREDICTED CLASS	
		CLASS = YES	CLASS = NO
ACTUAL CLASS	CLASS = YES	**a** (True Positive)	**b** (False Negative)
	CLASS = NO	**c** (False Positive)	**d** (True Negative)

$$\text{Accuracy} = \frac{a + d}{a + b + c + d} = \frac{TP + FP}{TP + FN + FP + TN}$$

IS ACCURACY ALWAYS GOOD?

- What is the accuracy of the model?

A Two-Class Problem

- Number of Class = Yes : 9990
- Number of Class = No : 10

IS ACCURACY ALWAYS GOOD?

- What is the accuracy of the model?

A Two-Class Problem

- Number of Class = Yes : 9990
- Number of Class = No : 10

- If the model predicts ALL instances to be “Yes”:

Accuracy = 99.90%

IS ACCURACY ALWAYS GOOD?

- What is the accuracy of the model?

  A Two-Class Problem

  - Number of Class = Yes : 9990
  - Number of Class = No : 10

- It is possible to get high accuracy if the model does not detect any instance of an imbalanced class

IMBALANCED-CLASS PROBLEM

- Precision
- Recall
- F-1 Score

PRECISION

- Total number of correctly classified instances out of all **Predicted** instances

$$\text{Precision} = \frac{TP}{TP + FP}$$

PRECISION

		PREDICTED CLASS	
		CLASS = YES	CLASS = NO
ACTUAL CLASS	CLASS = YES	**a** (True Positive)	**b** (False Negative)
	CLASS = NO	**c** (False Positive)	**d** (True Negative)

$$\text{Precision } (p) = \frac{a}{a + c} = \frac{TP}{TP + FP}$$

RECALL

- Total number of correctly classified instances out of all **Relevant** instances

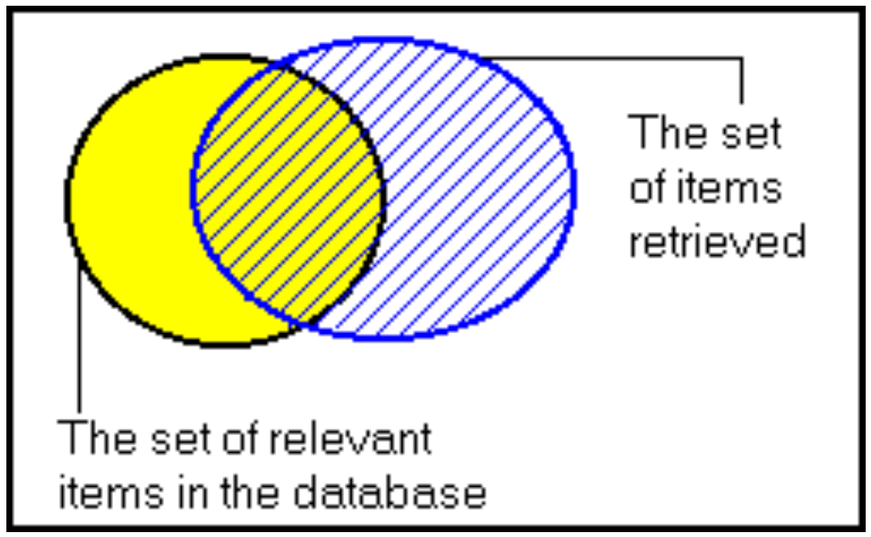
$$\text{Recall} = \frac{TP}{TP + FN}$$

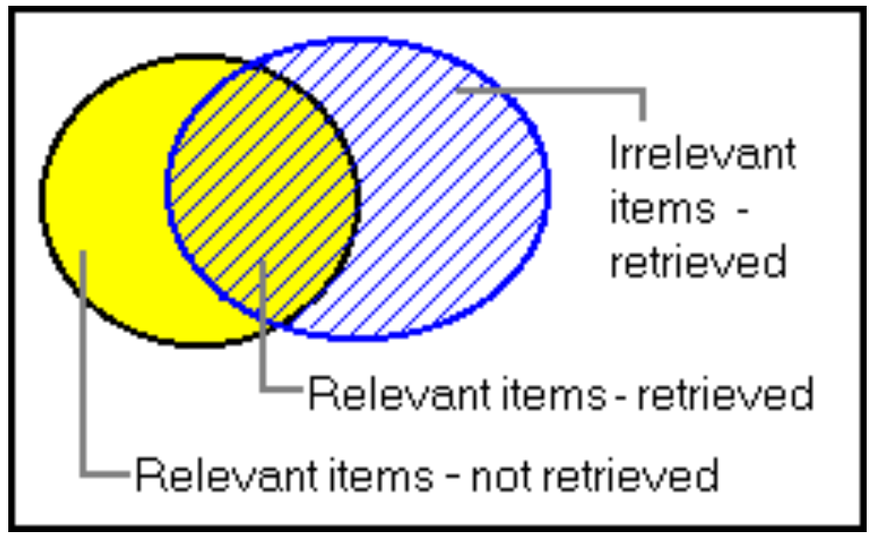
RECALL

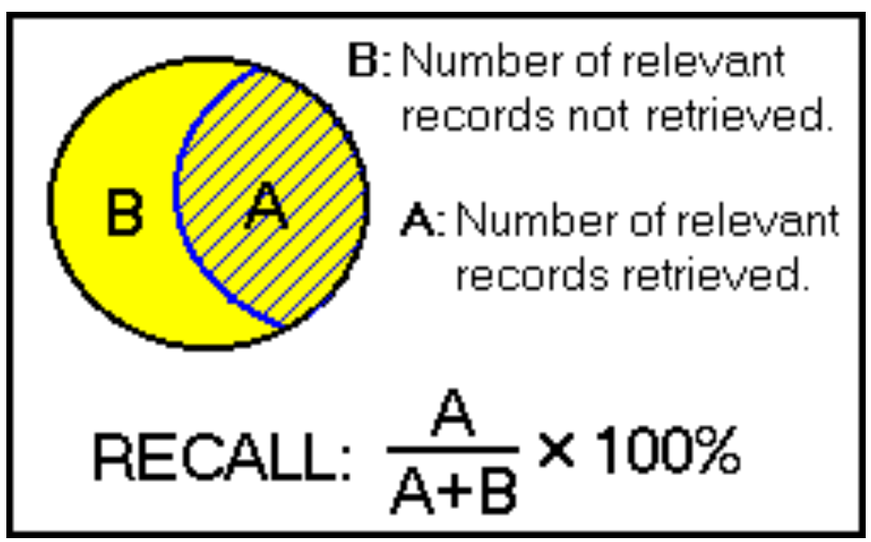
		PREDICTED CLASS	
		CLASS = YES	CLASS = NO
ACTUAL CLASS	CLASS = YES	**a** (True Positive)	**b** (False Negative)
	CLASS = NO	**c** (False Positive)	**d** (True Negative)

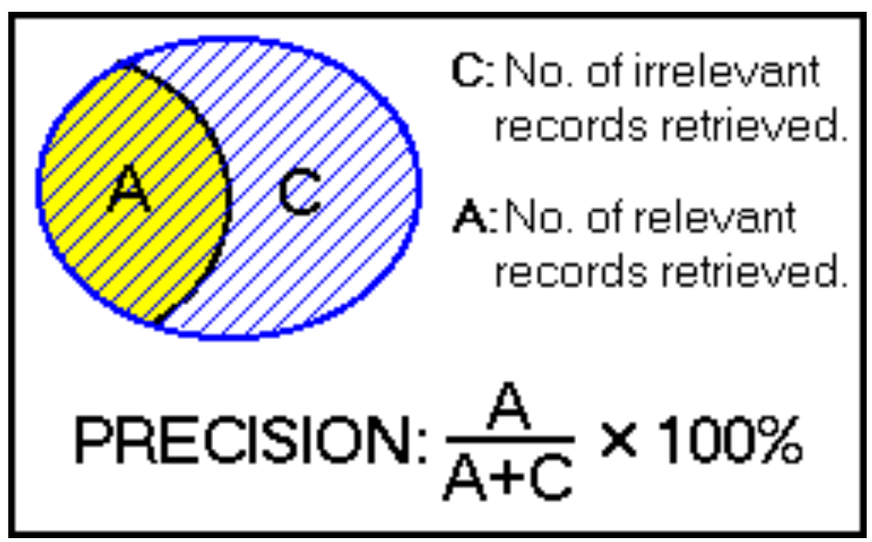
$$\textbf{Recall } (r) = \frac{a}{a+b} = \frac{TP}{TP+FN}$$

PRECISION AND RECALL









F-1 SCORE

- The harmonic mean between precision and recall

$$\text{F-1 Score} = \frac{2TP}{2TP + FP + FN}$$

F-1 SCORE

		PREDICTED CLASS	
		CLASS = YES	CLASS = NO
ACTUAL CLASS	CLASS = YES	**a** (True Positive)	**b** (False Negative)
	CLASS = NO	**c** (False Positive)	**d** (True Negative)

$$\text{F-1 Score} = \frac{2rp}{r + p} = \frac{2TP}{2TP + FP + FN}$$

F-1 SCORE

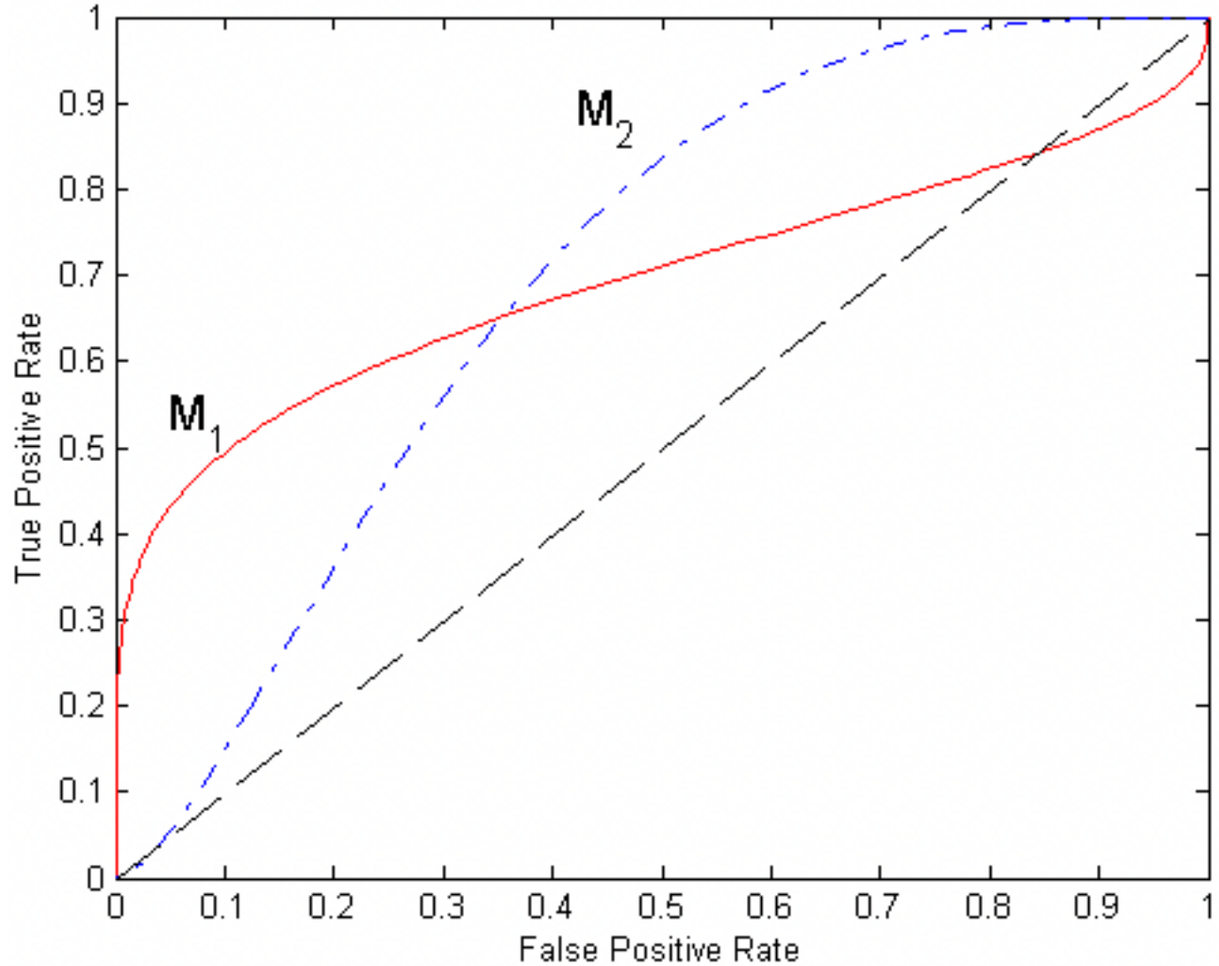
- The harmonic mean between precision and recall

$$\text{F-1 Score} = \frac{2TP}{2TP + FP + FN}$$

- Related Measure - Balanced Accuracy
  It is the arithmetic mean of sensitivity and specificity

ROC CURVE

- Receiver Operating Characteristics (ROC) curve
- M_1 is better for lower FPR
  M_2 is better for greater FPR
- Related Measure:
  Area Under the Curve (AUC)



FOR MORE METRICS

What better resource than a practical one?

https://scikit-learn.org/stable/modules/model_evaluation.html

EXAMPLE CODE

Examples (along with measures to tackle imbalanced-class problems) can be found here:

https://github.com/learn-co-students/dc-ds-071519/blob/master/Module-5/week-3/Class_imbalance_model_eval.ipynb