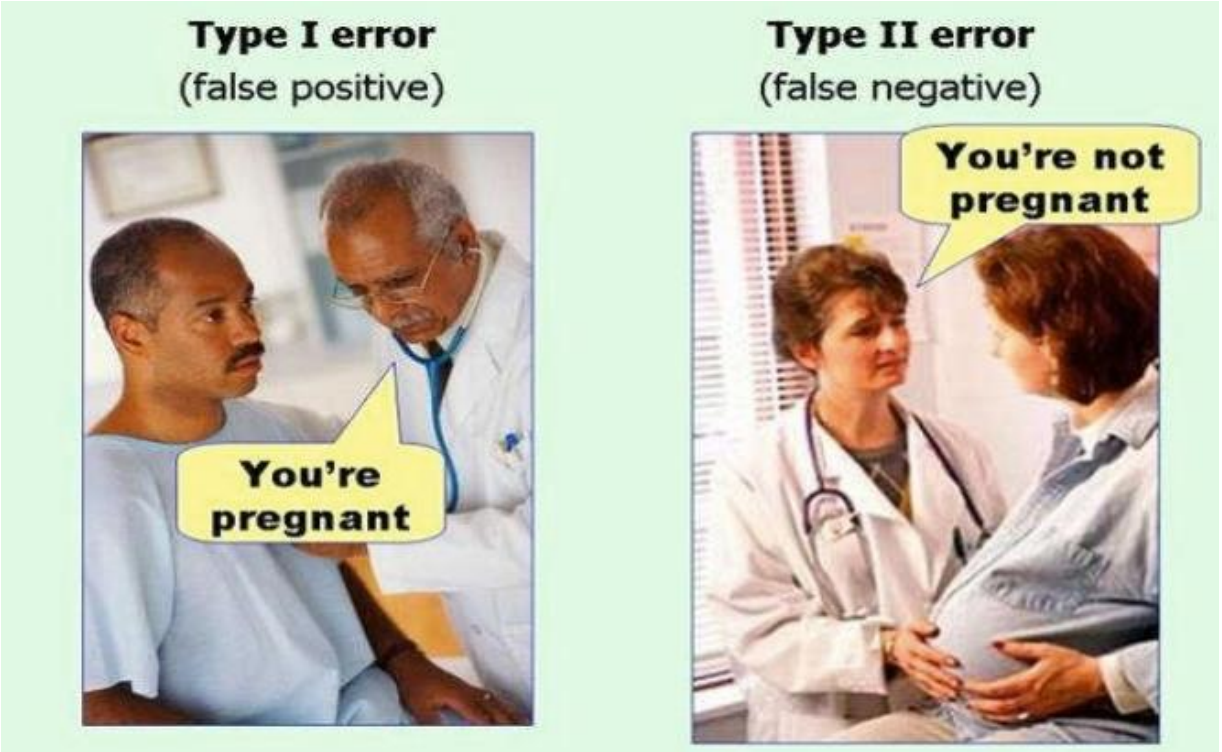
Confusion Matrix



DATA SCIENCE

Confusion Matrix

The confusion matrix is the primary method used to validate a classifier. Most of the model quality and accuracy metrics are based on the values of the confusion matrix. This matrix is a table that contains information about the actual and predicted values for a classifier.

- **True positives (TP):** These are cases in which we predicted yes (they have the disease), and they do have the disease.
- **True negatives (TN):** We predicted no, and they don't have the disease.
- **False positives (FP):** We predicted yes, but they don't actually have the disease. (Also known as a "Type I error.")
- **False negatives (FN):** We predicted no, but they actually do have the disease. (Also known as a "Type II error.")

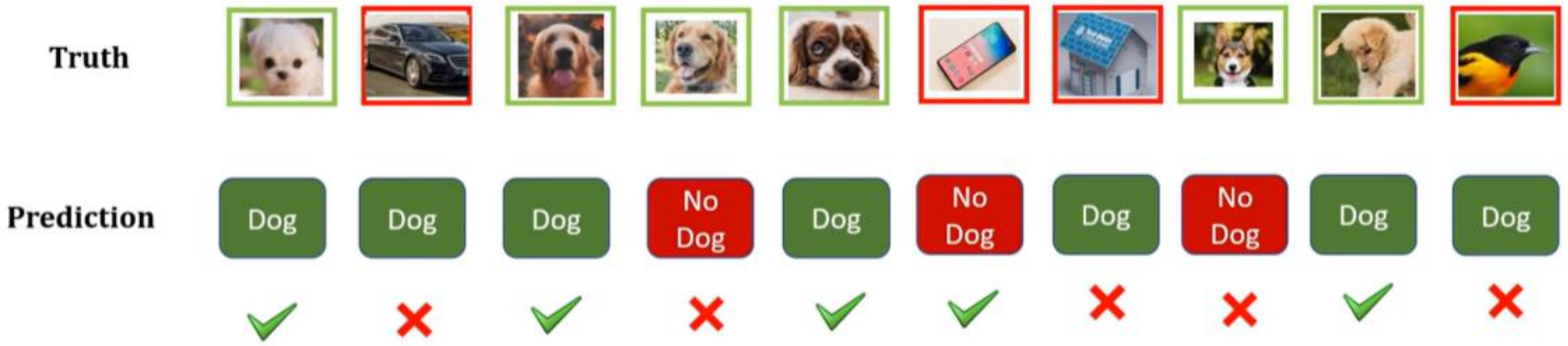
n=165	**Predicted: NO**	**Predicted: YES**	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

DATA SCIENCE

Confusion Matrix in Machine Learning

Truth										
Prediction	Dog	Dog	Dog	No Dog	Dog	No Dog	Dog	No Dog	Dog	Dog

Confusion Matrix in Machine Learning



How many we got right? → 5

Accuracy → 5/10 → **0.5**

DATA SCIENCE

Confusion Matrix in Machine Learning

True Positive = 4

False Positive = 3

Confusion Matrix in Machine Learning

True Negative = 1

False Negative = 2

Confusion Matrix in Machine Learning

True Positive = 4

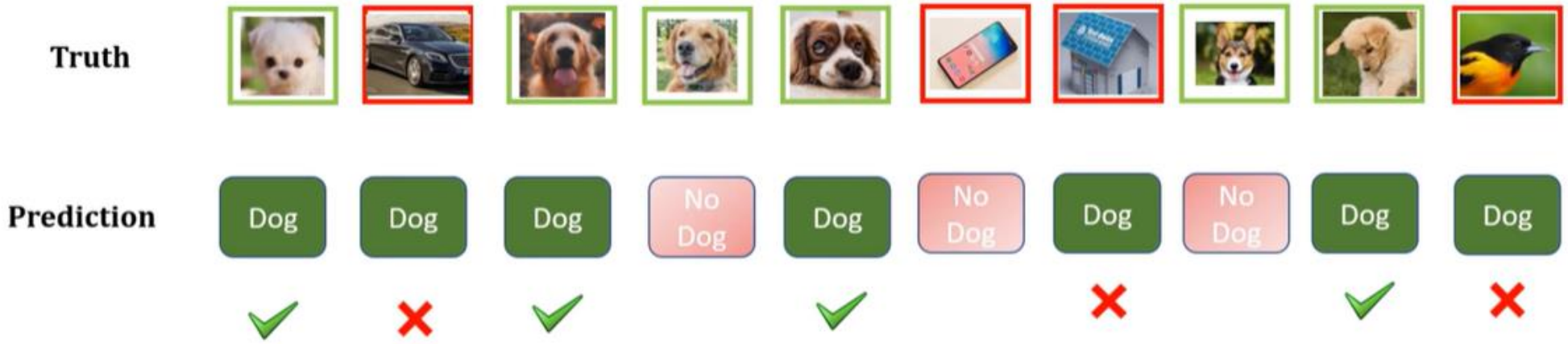
False Positive = 3

Precision is out of all dog predictions how many you got it right?

Precision = 4 / 7 = 0.57

Precision = TP / (TP + FP)

Confusion Matrix in Machine Learning



Recall is out of all dog truth how many you got it right?

Total Dog truth samples = 6

True Positive = 4

Recall = 4 / 6 = 0.67

Recall = TP / (TP + FN)

Confusion Matrix in Machine Learning

Precision = 1 / 3 = 0.33

Recall = 1 / 4 = 0.25

Confusion Matrix in Machine Learning

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} = 2*(0.57*0.67/(0.57+0.67)) = 0.6159677419354839$$