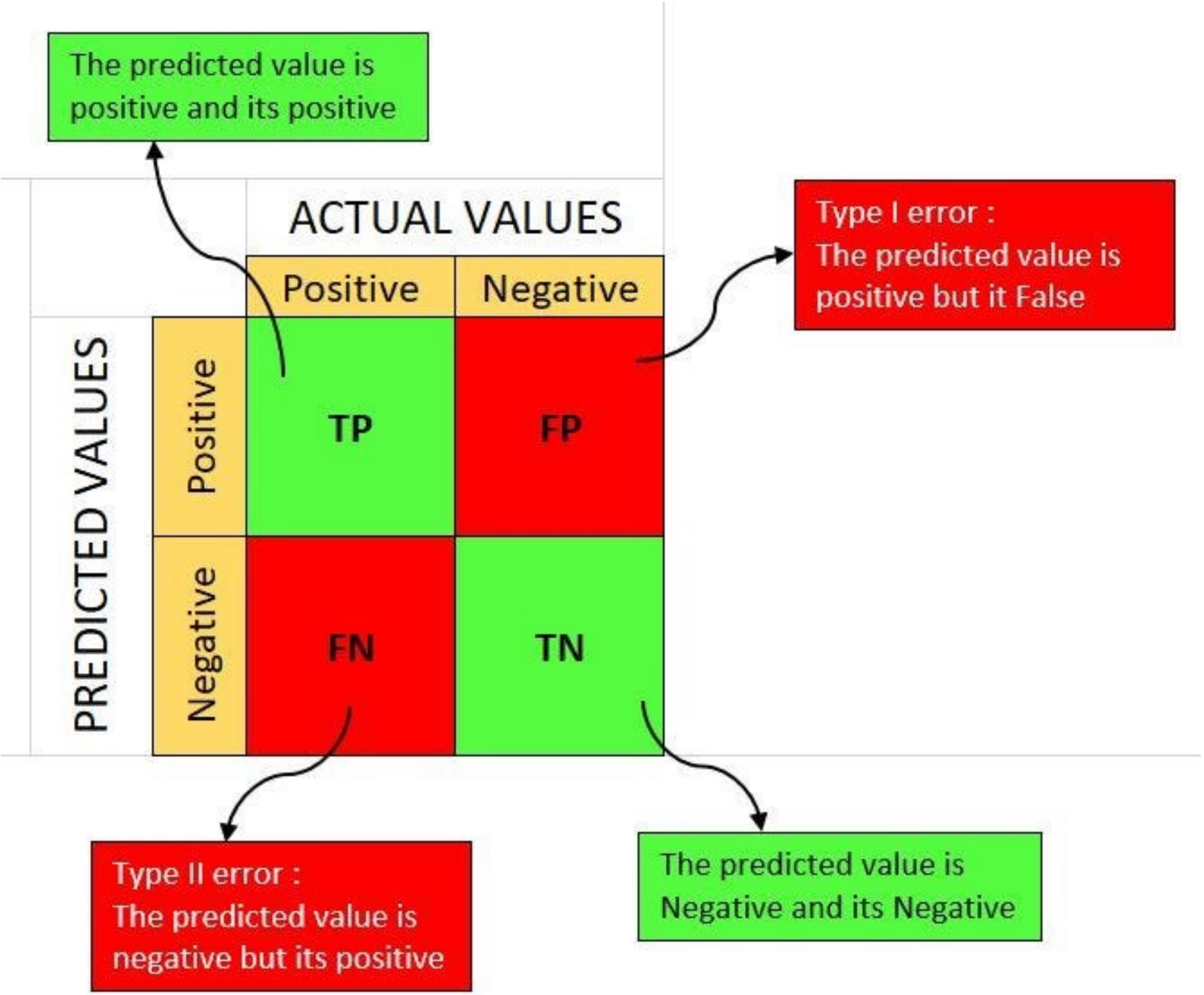
Confusion Matrix

Mustafa Coşkun

A **Confusion matrix** is an ***N x N matrix*** used for evaluating the **performance of a classification model**, where **N** is the number of ***target classes***. The matrix compares the actual target values with those predicted by the machine learning model.



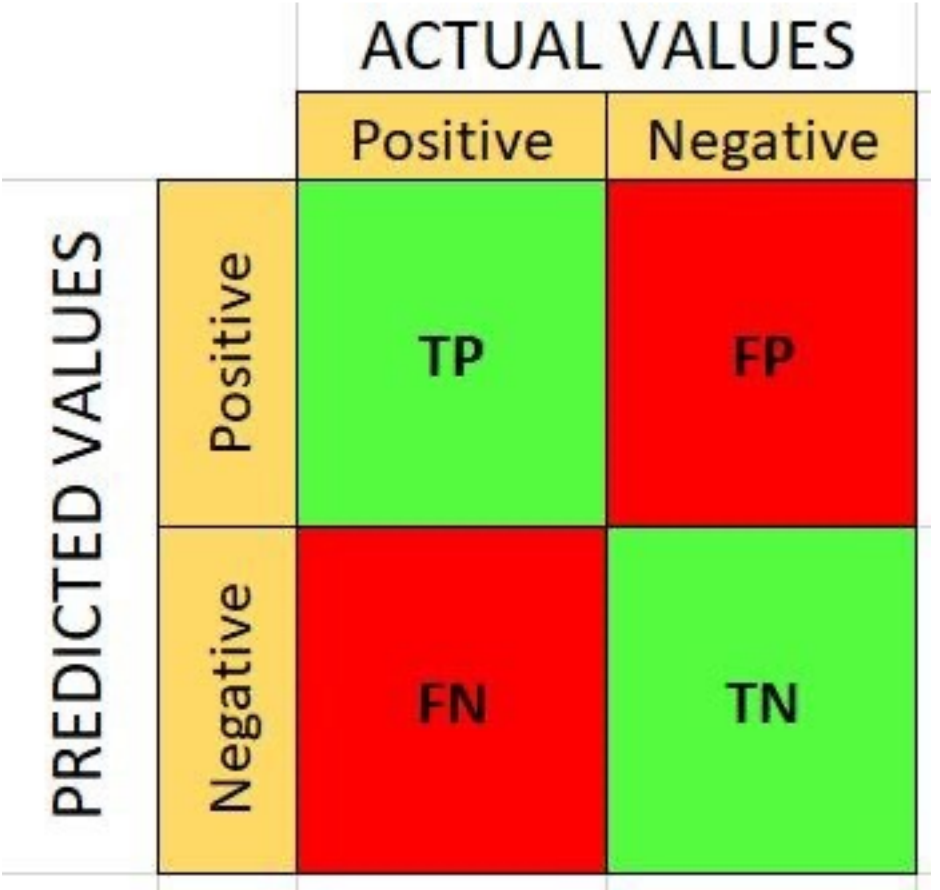
1. A **good model** is one which has ***high TP and TN rates***, while ***low FP and FN rates***.
2. If you have an ***imbalanced dataset*** to work with, it's always better to use ***confusion matrix*** as your evaluation criteria for your machine learning model.

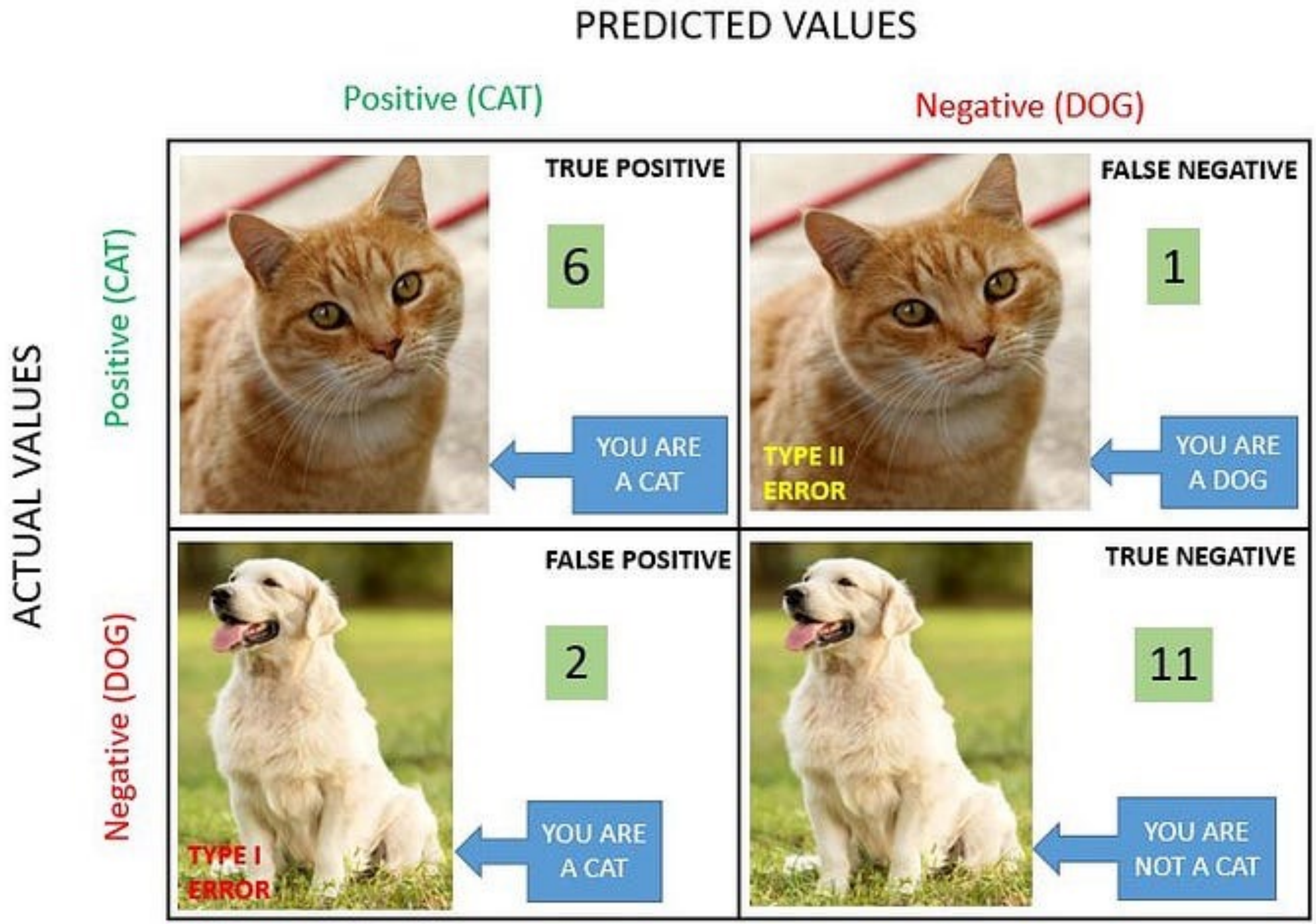
		Predicted	
		Negative **(N)** **-**	Positive **(P)** **+**
Actual	Negative **-**	True Negative **(TN)**	**False Positive (FP)** **Type I Error**
	Positive **+**	**False Negative (FN)** **Type II Error**	True Positive **(TP)**

Binary Classification Problem (2x2 matrix)

Confusion Matrix

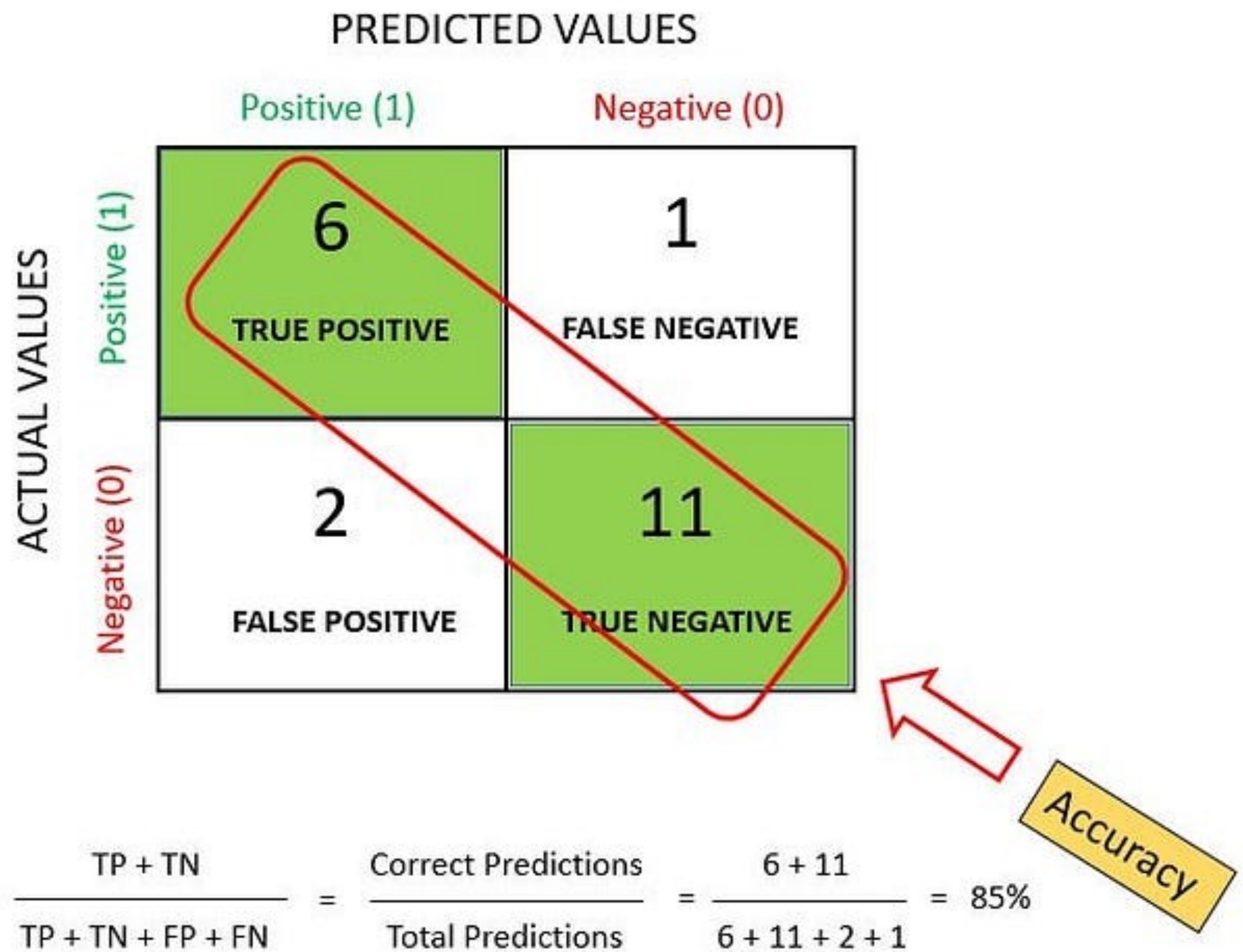
- **True Positives (TP)**: when the actual value is Positive and predicted is also Positive.
- **True negatives (TN)**: when the actual value is Negative and prediction is also Negative.
- **False positives (FP)**: When the actual is negative but prediction is Positive. Also known as the **Type 1 error**
- **False negatives (FN)**: When the actual is Positive but the prediction is Negative. Also known as the **Type 2 error**




PREDICTED VALUES
Positive (CAT)
Negative (DOG)
ACTUAL VALUES
Positive (CAT)
Negative (DOG)
TRUE POSITIVE
6
YOU ARE A CAT
FALSE NEGATIVE
1
TYPE II ERROR
YOU ARE A DOG
FALSE POSITIVE
2
TYPE I ERROR
YOU ARE A CAT
TRUE NEGATIVE
11
YOU ARE NOT A CAT

Classification Measures: Accuracy

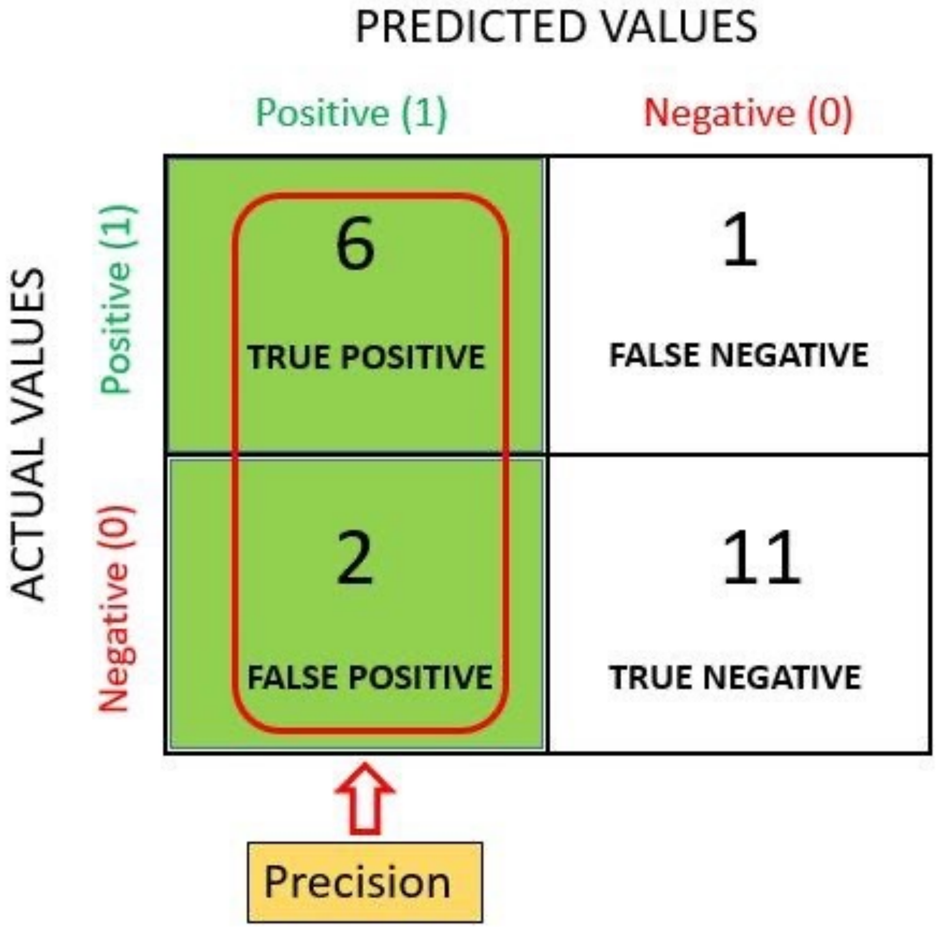
- **Accuracy** simply measures how often the classifier makes the correct prediction. It's the ratio between the number of correct predictions and the total number of predictions.
- The **accuracy metric** is ***not suited*** for **imbalanced classes**. **Accuracy** has its own *disadvantages*, for **imbalanced data**, when the model predicts that each point belongs to the majority class label, the accuracy will be high. But, the model is not accurate.



$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{\text{Correct Predictions}}{\text{Total Predictions}} = \frac{6 + 11}{6 + 11 + 2 + 1} = 85\%$$

Classification Measures: Precision

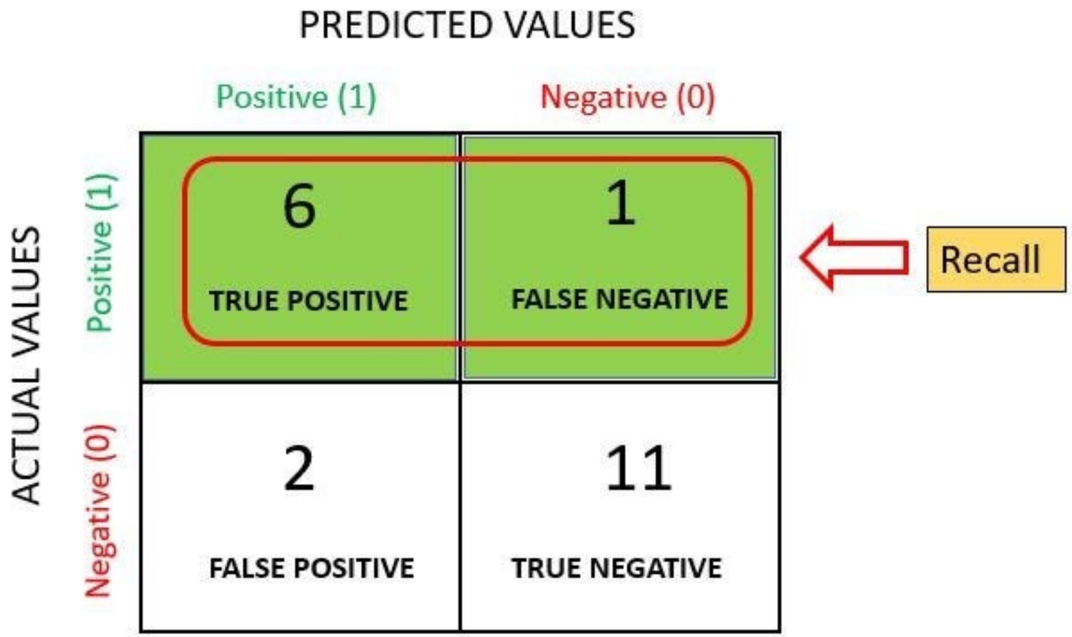
- It is a measure of **correctness** that is achieved in **true prediction**. In simple words, it tells us how many predictions are ***actually positive*** out of all the ***total positive predicted***.
- Precision is defined as the ratio of the total number of *correctly classified positive classes* divided by the *total number of predicted positive classes*. Or, out of all the predictive positive classes, how much we predicted correctly. **Precision should be high(ideally 1).**
- *"**Precision** is a useful metric in cases where **False Positive** is a higher concern than **False Negatives**"*



$$\text{Precision} = \frac{TP}{TP + FP} = \frac{\text{Predictions Actually Positive}}{\text{Total Predicted positive}} = \frac{6}{6+2} = 0.75$$

Classification Measures: Recall

- It is a measure of **actual observations** which are predicted **correctly**, i.e. how many observations of positive class are actually predicted as positive. It is also known as **Sensitivity**. ***Recall*** is a valid choice of evaluation metric when we want to capture ***as many positives*** as possible.
- Recall is defined as the ratio of the total number of *correctly classified positive classes* divide by the *total number of positive classes*. Or, out of all the positive classes, how much we have predicted correctly. **Recall should be high(ideally 1).**
- *"**Recall** is a useful metric in cases where **False Negative** trumps **False Positive**"*



$$\text{Recall} = \frac{TP}{TP + FN} = \frac{\text{Predictions Actually Positive}}{\text{Total Actual positive}} = \frac{6}{6+1} = 0.85$$

Classification Measures: F1-Score

- The **F1 score** is a number between **0 and 1** and is the ***harmonic mean of precision and recall***. We use harmonic mean because it is not sensitive to extremely large values, unlike simple averages.
- **F1 score** sort of maintains a **balance** between the ***precision and recall*** for your classifier. If your ***precision is low***, the ***F1 is low*** and if the ***recall is low*** again your ***F1 score is low***.
- There will be cases where there is no clear distinction between whether *Precision is more important or Recall*.
- F1 score is a ***harmonic mean*** of Precision and Recall. As compared to Arithmetic Mean, Harmonic Mean punishes the extreme values more. **F-score should be high(ideally 1).**

$$\text{F1-Score} = 2^* \frac{(\text{Recall}^*\text{Precision})}{(\text{Recall} + \text{Precision})} = 2^* \frac{(0.85^*0.75)}{(0.85 + 0.75)} = 0.79$$

When to use Accuracy / Precision / Recall / F1-Score

- a. ***Accuracy*** is used when the ***True Positives and True Negative***s are more important. ***Accuracy*** is a better metric for ***Balanced Data.***
- b. Whenever **False Positive** is much more important use ***Precision.***
- c. Whenever **False Negative** is much more important use ***Recall.***
- d. ***F1-Score*** is used when the ***False Negatives and False Positives*** are important. ***F1-Score*** is a better metric for ***Imbalanced Data.***

https://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html,
https://scikit-learn.org/stable/auto_examples/model_selection/plot_confusion_matrix.html
https://medium.com/analytics-vidhya/what-is-a-confusion-matrix-d1c0f8feda5