# Evaluation Metrics

As it was established that the **cross-validation** is used to estimate the test performance of a model, and for that there is many evaluation metrics to determine the effectiveness of a model

# Classification Metrics

## Threshold Based Metrics

These metrics are derived from the confusion matrix which changes depends on a set threshold :

- **Accuracy**
  - Measures how correct the predictions of our model (total correct predictions)

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Instances}}$$

- **Precision**
  - Measures the accuracy of **positive predictions**, (how many predicted were actually positive..**True** )

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

- **Recall (Sensitivity)**
  - Measures the model's ability to find all the positive instances (Of all the positives ,how many positives the model got correctly )

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

- **Specificity**
  - The proportion of actual negatives correctly identified

$$\text{Specificty} = \frac{\text{True Negative}}{\text{True Negative} + \text{False Positive}}$$

  - **Negative** class has high value
  - **False Positive** cost is high
  - Quality Control
  - Security and Fraud Detection
  - Rare Disease
- $\text{F-1}$ **Score**
  - Useful for a balance between **Precision** and **Recall**, works best on **imbalanced** class distributions

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

*Why Accuracy isn't always enough?*

- In many cases where the class distributions isn't **balanced** if a class $A$ is $99\%$ of the data and only $1\%$ belong to class $B$
- A model can have $99\%$ accuracy but fails to detect class $B$

# Confusion Matrix

| | **Actually Positive** | **Actually Negative** |
|---|---|---|
| Predicted Positive $1$ | $\text{True Positive (TP)}$ | $\text{False Positive (FP)}$ |
| Predicted Negative $0$ | $\text{False Negative (FN)}$ | $\text{True Negative(TN)}$ |

- $\text{False Positive}$ is a **Type I** error where the model incorrectly predicted something **Negative** as **Positive**
- $\text{False Negative}$ is a **Type II** error where the model miss the **Positive** and predict it as **Negative**

# Threshold Independent Metrics

These metrics evaluate the model's quality without depending on a fixed **Threshold** value :

- **Area Under the ROC Curve** ROC Curve : visual representation of the classification model across all the possible **Threshold**, it plot the **True Positive** (Sensitivity) against **False Positive**(Specificity)
- **Area Under the Precision-Recall Curve** : As the **ROC Curve** it's a graphical representation of a Classifier Model , It's often used in situations where the classes are heavily imbalanced