**What is Classification?**

Classification is technique to categorize our data into a desired and distinct number of classes where we can assign label to each class.

***Applications of Classification are:***speech recognition, handwriting recognition, biometric identification, document classification etc.

**Binary classifiers:**Classification with only 2 distinct classes or with 2 possible outcomes

example: Male and Female

example: classification of spam email and non spam email

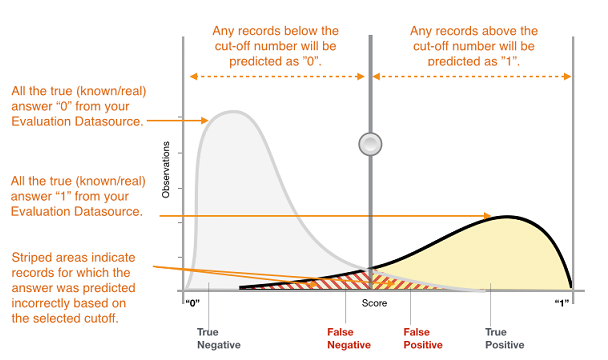
**Binary** or **binomial classification** is the task of classifying  the elements of a given set into two groups (predicting which group each one belongs to) on the basis of a classification rule.

## **Evaluation of binary classifiers**

There are many metrics that can be used to measure the performance of a classifier or predictor; different fields have different preferences for specific metrics due to different goals.

Given a classification of a specific data set, there are four basic combinations of actual data category and assigned category: **true positive** TP (correct positive assignments), **true negatives** TN (correct negative assignments), **false positives** FP (incorrect positive assignments), and **false negatives** FN (incorrect negative assignments).

|  |  |  |
| --- | --- | --- |
|  | **Condition Positive (CP)** | **Condition Negative (CN)** |
| **Test Outcome Positive (OP)** | True Positive | False Positive |
| **Test Outcome Negative (ON)** | False Negative | True Negative |



As the target variable is not continuous, binary classification model predicts the probability of a target variable to be Yes/No. To evaluate such a model, a metric called the confusion matrix is used, also called the classification or co-incidence matrix. With the help of a confusion matrix, we can calculate important performance measures:

1. True Positive Rate (TPR) or Hit Rate or Recall or Sensitivity = TP / (TP + FN)
2. False Positive Rate(FPR) or False Alarm Rate = 1 - Specificity = 1 - (TN / (TN + FP))
3. Accuracy = (TP + TN) / (TP + TN + FP + FN)
4. Error Rate = 1 – accuracy or (FP + FN) / (TP + TN + FP + FN)
5. Precision = TP / (TP + FP)
6. F-measure: 2 / ( (1 / Precision) + (1 / Recall) )
7. ROC (Receiver Operating Characteristics) = plot of FPR vs TPR
8. AUC (Area Under the Curve)
9. Kappa statistics

<https://www.digitalocean.com/community/tutorials/how-to-build-a-machine-learning-classifier-in-python-with-scikit-learn>