Confusion Matrix

Measure of the performance of your model

Topics

What is a Confusion Matrix?

- True Positive
- True Negative
- False Positive – Type 1 Error
- False Negative – Type 2 Error

Why need a Confusion matrix?

Precision vs Recall

F1-score

Introduction

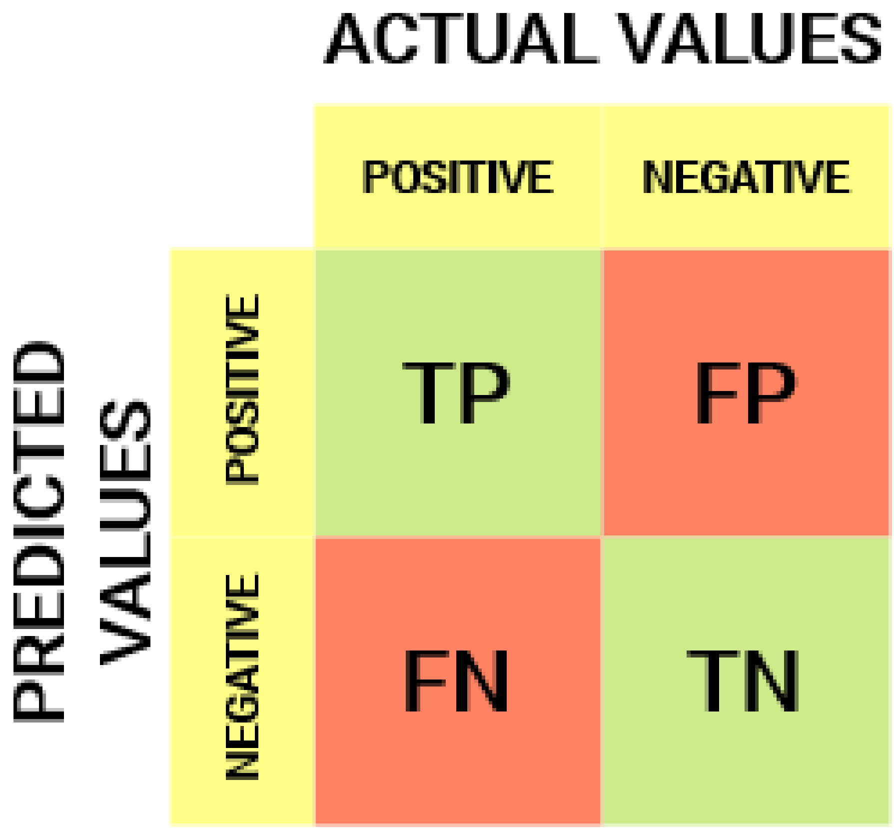
A Confusion matrix is an *N x N* matrix used for evaluating the performance of a **classification model**, where **N** is the number of **target classes**.

The matrix compares the **actual** target values with those **predicted** by the machine learning model.

This gives us an idea about how **well our classification model is performing** and **what kinds of errors it is making**.

Confusion Matrix

For a binary classification problem: 2 x 2 matrix with 4 values:



True Positive (TP)
- The **actual value is positive**, and the model predicted a positive value

True Negative (TN)
- The **actual value** is negative, and the model predicted a negative value

False Positive (FP) – Type 1 error
- The actual value is negative, but the model predicted a positive value
- Also known as the **Type 1 error**

False Negative (FN) – Type 2 error
- The actual value is positive, but the model predicted a negative value
- Also known as the **Type 2 error**

Metrics based on Confusion Matrix Data

Accuracy

- Accuracy is used to measure the performance of the model. It is the ratio of Total correct instances to the total instances.

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

Metrics based on Confusion Matrix Data

Precision

- Precision is a measure of how accurate a model's positive predictions are. It is defined as the ratio of true positive predictions to the total number of positive predictions made by the model.

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

Metrics based on Confusion Matrix Data

Recall

- Recall measures the effectiveness of a classification model in identifying all relevant instances from a dataset. It is the ratio of the number of true positive (TP) instances to the sum of true positive and false negative (FN) instances.

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

Specificity

Specificity is another important metric in the evaluation of classification models, particularly in binary classification. It measures the ability of a model to correctly identify negative instances. Specificity is also known as the True Negative Rate. Formula is given by:

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

F1-score is a harmonic mean of Precision and Recall.

F1-score is used to evaluate the overall performance of a classification model. It is the harmonic mean of precision and recall,

$$F1 - score = \frac{2}{\frac{1}{Recall} + \frac{1}{Precision}}$$

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

Actual

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

$$\text{Accuracy} = \frac{TP + TN}{\text{Total}}$$

Predicted

	Pos	Neg
Pos	TP	FP
Neg	FN	TN

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

- ***Accuracy*** is used when the ***True Positives and True Negative***s are more important. ***Accuracy*** is a better metric for ***Balanced Data.***

- Whenever **False Positive** is much more important use ***Precision.***

- Whenever **False Negative** is much more important use ***Recall.***

- ***F1-Score*** is used when the ***False Negatives and False Positives*** are important. ***F1-Score*** is a better metric for ***Imbalanced Data.***

Confusion Matrix For Multi-class Classification

	Predicted Cat	Predicted Dog	Predicted Horse
Actual Cat	True Positive (TP)	False Negative (FN)	False Negative (FN)
Actual Dog	False Negative (FN)	True Positive (TP)	False Negative (FN)
Actual Horse	False Negative (FN)	False Negative (FN)	True Positive (TP)

	Predicted Cat	Predicted Dog	Predicted Horse
Actual Cat	8	1	1
Actual Dog	2	10	0
Actual Horse	0	2	8

In this scenario:

- **Cats:** 8 were correctly identified, 1 was misidentified as a dog, and 1 was misidentified as a horse.
- **Dogs:** 10 were correctly identified, 2 were misidentified as cats.
- **Horses:** 8 were correctly identified, 2 were misidentified as dogs.

To calculate true negatives, we need to know the total number of images that were NOT cats, dogs, or horses.

Let’s assume there were 10 such images, and the model correctly classified all of them as “not cat,” “not dog,” and “not horse.”
Therefore:

True Negative (TN) Counts: 10 (for each class, as the model correctly identified each non-cat/dog/horse image as not belonging to that class)


Thank you