Confusion Matrix

Confusion Matrix in Machine Learning

A **Confusion Matrix** is a table used to evaluate the **performance of a classification model** by comparing the actual vs. predicted values. It helps understand how well a model distinguishes between classes, especially in **binary and multiclass classification problems**.

Structure of a Confusion Matrix (Binary Classification)

Actual / Predicted	Predicted: 0	Predicted: 1
Actual: 0 (Negative Class)	**True Negative (TN)**	**False Positive (FP)**
Actual: 1 (Positive Class)	**False Negative (FN)**	**True Positive (TP)**

Explanation of Terms:

True Positive (TP) → Model correctly predicted **1** (Actual = 1, Predicted = 1)

True Negative (TN) → Model correctly predicted **0** (Actual = 0, Predicted = 0)

False Positive (FP) → Model incorrectly predicted **1** (Actual = 0, Predicted = 1) (**Type I Error**)

False Negative (FN) → Model incorrectly predicted **0** (Actual = 1, Predicted = 0) (**Type II Error**)

Key Metrics Derived from Confusion Matrix

1️⃣ **Accuracy** → How many predictions were correct?

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

2️⃣ **Precision (Positive Predictive Value)** → Out of predicted positives, how many were actually positive?

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

3️⃣ **Recall (Sensitivity, True Positive Rate)** → Out of actual positives, how many did we correctly predict?

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

4️⃣ **F1-Score** → Harmonic mean of Precision & Recall (useful for imbalanced datasets).

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

5️⃣ **Specificity (True Negative Rate)** → Out of actual negatives, how many were correctly identified?

$$Specificity = \frac{TN}{TN + FP}$$

6️⃣ **False Positive Rate (FPR)** → How often does the model predict positive when it's actually negative?

$$FPR = \frac{FP}{FP + TN}$$

Example Calculation

Assume we have **100 test samples** and the confusion matrix:

Actual / Predicted	Predicted: 0	Predicted: 1
Actual: 0	**50 (TN)**	**10 (FP)**
Actual: 1	**5 (FN)**	**35 (TP)**

🔷 **Accuracy** = $\frac{50+35}{100} = 0.85$ (85%)

🔷 **Precision** = $\frac{35}{35+10} = 0.78$ (78%)

🔷 **Recall** = $\frac{35}{35+5} = 0.875$ (87.5%)

🔷 **F1-Score** = $2 \times \frac{0.78 \times 0.875}{0.78 + 0.875} = 0.825$ (82.5%)

Why Use a Confusion Matrix?

✅ **Gives a Detailed View of Model Performance** – Unlike accuracy, it shows the distribution of errors.

✅ **Helps Handle Imbalanced Datasets** – Accuracy alone can be misleading in skewed datasets.

✅ **Used in Medical, Fraud Detection, and Spam Classification Models** – Helps identify misclassifications.

Support

In a **confusion matrix**, **support** refers to the **number of actual true instances** of each class in the **dataset (usually the test set)**. It's the same as what you saw in the classification report.

Support gives context to other metrics like **precision**, **recall**, and **f1-score**. If support is very low (e.g., only 5 samples), then a small change in predictions can cause a big change in those metrics.
