Evaluation Metrics for Classification Problems

📌 1. Confusion Matrix

The foundation of all classification metrics.

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

- **TP (True Positive):** Correctly predicted positives
- **TN (True Negative):** Correctly predicted negatives
- **FP (False Positive):** Incorrectly predicted positives (Type I error)
- **FN (False Negative):** Incorrectly predicted negatives (Type II error)

📌 2. Accuracy

> Measures the overall correctness of the model.

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

✅ When to Use:

- When classes are **balanced**
- When **every class matters equally**

❌ Not Reliable:

- When the data is **imbalanced**
  (e.g., 95% class A and 5% class B — model can have 95% accuracy by always predicting A)

📌 3. **Precision**

> Of all predicted positives, how many are actually positive?

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

✅ Use when:

- **False positives are costly**
  e.g., Spam classification — don't want to mark important emails as spam

📌 4. **Recall (Sensitivity / True Positive Rate)**

> Of all actual positives, how many did we correctly predict?

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

✅ Use when:

- **Missing positives is costly**
  e.g., Medical test — failing to detect disease is dangerous

📌 5. **F1-Score**

> Harmonic mean of precision and recall
> Gives a **balanced measure** when both are important

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

- F1 Score is high only when **both** precision and recall are high.
- Useful when data is **imbalanced**.