Accuracy & Confusion Matrix | Type 1 and Type 2 Errors

1. Accuracy

- ***Definition:***
  Accuracy = (Number of correct predictions) ÷ (Total predictions).
- ***Purpose:*** Measures overall correctness of the model.
- ***Example:*** If your model made 100 predictions and 90 were correct, accuracy is 90%.
- ***Limitation:*** High accuracy doesn't always mean good performance, especially with imbalanced datasets (e.g., 95% accuracy predicting all "No Disease" when only 5% have the disease).

2. Accuracy in Multi-Class Problems

- In multi-class classification, accuracy is the proportion of correct predictions across all classes.
- Works the same way as binary classification but checks for correctness in all categories.

3. How Much Accuracy is Good?

- There's no fixed "good" accuracy — it depends on the domain.
  1. For spam detection, 99% may be required.
  2. For movie recommendations, even 70% might be acceptable.

4. Problem with Accuracy

- Can be misleading if classes are **imbalanced** (one class dominates).
- In such cases, the model might predict only the majority class and still show high accuracy.

5. Confusion Matrix

- **Definition:** A table showing correct and incorrect predictions, broken down by class.
- **Structure for Binary Classification:**

	Predicted Positive	Predicted Negative
Actual Positive	TP	FN
Actual Negative	FP	TN

  1. **TP** (True Positive) – Correctly predicted positive cases.

2. **TN** (True Negative) – Correctly predicted negative cases.
3. **FP** (False Positive) – Predicted positive but actually negative.
4. **FN** (False Negative) – Predicted negative but actually positive.

6. Type 1 & Type 2 Errors

- **Type 1 Error (False Positive):** Predict positive when it's actually negative.
- *Example:* Detecting a disease in a healthy patient.
- **Type 2 Error (False Negative):** Predict negative when it's actually positive.
- *Example:* Missing a disease in a sick patient.
- The trade-off between these depends on the problem's criticality.

7. Confusion Matrix for Multi-Classification

- Expanded version of confusion matrix with rows and columns for each class.
- Each diagonal cell shows correct predictions; off-diagonal cells show misclassifications.

8. When Accuracy is Misleading

- In cases of severe class imbalance.
- Better metrics to use: **Precision, Recall, F1-Score**, these give more insight into model performance for minority classes.