



DAY 16 — Model Evaluation Metrics

Goal : Evaluate ML models correctly (not blindly)

1 Why Accuracy is Dangerous (VERY IMPORTANT)

Accuracy =

| Correct predictions / Total predictions

Sounds good... but **can lie badly**.

Example (Real world)

- 95% of emails are NOT spam
- Model predicts "not spam" always

Accuracy = 95%

Model usefulness = 0%

⚠ This is why accuracy alone is **not enough**.

2 Confusion Matrix (FOUNDATION)

This is the **base of all metrics**.

| | Predicted YES | Predicted NO |
|------------|---------------|--------------|
| Actual YES | TP | FN |
| Actual NO | FP | TN |

Terms

- **TP** (True Positive): correct positive
- **TN** (True Negative) : correct negative
- **FP** (False Positive) : false alarm
- **FN** (False Negative) : missed detection

🧠 Everything else comes from this.

3 Precision

Formula

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

Meaning

“When the model predicts YES, how often is it right

Use when:

- False positives are expensive
(spam, fraud alerts, medical tests)

4 Recall

Formula

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

Meaning

“Out of all actual YES cases, how many did we catch?”

Use when:

- Missing a positive case is dangerous
(cancer detection, fraud detection)

5 Precision vs Recall Tradeoff (CRITICAL)

| Metric | Focus |
|-----------|--------------------|
| Precision | Fewer false alarms |
| Recall | Fewer missed cases |

You **cannot maximize both at once**.

This tradeoff is controlled by **thresholds**.

6 F1 Score (BALANCED METRIC)

Formula

$$F1 = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

Meaning

- Balance between Precision & Recall
- Good for **Imbalanced datasets**

If unsure → use **F1 score**

7 Accuracy vs F1 (When to Use What)

| Scenario | Metric |
|-------------------|-----------|
| Balanced data | Accuracy |
| Imbalanced data | F1 |
| Medical diagnosis | Recall |
| Spam detection | Precision |

8 ROC Curve & AUC (Advanced but IMPORTANT)

ROC Curve

- Plots **TPR vs FPR**
- Shows performance across thresholds

AUC

- Area under ROC curve
- Measures overall discrimination ability

AUC = 0.5 → random guessing

AUC = 1.0 → perfect model

9 Classification Report (ONE-LINE SUMMARY)

```
from sklearn.metrics import classification_report
```

```
print(classification_report(y_test, y_pred))
```

Includes:

- Precision
- Recall
- F1-score
- Support

Use this instead of accuracy alone.
