📅 DAY 28 — Model Evaluation & Validation

Goal :

Learn to answer **professional ML questions**:

- *Is my model good or just lucky?*
- *Will it perform on unseen data?*
- *Is it overfitting or underfitting?*
- *Which model should I deploy?*

1️⃣ Why Accuracy Alone Is Dangerous ❌

Accuracy hides problems.

Example:

- 95% data = class 0
- Model predicts always 0
- Accuracy = 95% ❌ but useless

We need **better metrics**.

2️⃣ Confusion Matrix

A table showing **what the model got right/wrong**.

	Predicted 0	**Predicted 1**
Actual 0	TN	FP
Actual 1	FN	TP

Why it matters

- Shows error types
- Critical for ML decisions

3️⃣ Precision, Recall, F1-Score

Precision

> Of predicted positives, how many were correct?

```
Precision = TP / (TP + FP)
```

Important when **false positives are costly**

F1 Score

> Balance between precision and recall

```
F1 = 2 * (precision * recall ) / (precision + recall)
```

Best for **imbalanced datasets**.

4️⃣ Classification Report

Provides:

- Precision
- Recall
- F1-score
- Support

One line of code gives everything.

5️⃣ ROC Curve & AUC

ROC Curve

- Ture Positive Rate vs False Positive Rate

AUC

- Area under curve
- Higher = better

AUC	Meaning
0.5	Random

0.7+	Acceptable
0.9+	Excellent

6️⃣ Cross-Validation

> Instead of 1 train-test split, use **multiple splits.**

K-Fold CV:

- Split data into K parts
- Train K times
- Average performance

Reduces randomness

More reliable evaluation

7️⃣ Bias vs Variance

Problem	Meaning
High Bias	Underfitting
High Variance	Overfitting

Symptoms

- High train score + low test score → Overfitting
- Low train & test score → Underfitting

8️⃣ Validation Set vs Test Set

Dataset	Purpose
Train	Learn
Validation	Tune
Test	Final evaluation

Test set must be **touched once only.**

9️⃣ When to Use Which Metric?

Use case	Metric
Balanced data	Accuracy
Medical	Recall
Fraud	Precision
General ML	F1
Model comparison	Cross-val score
