Evaluation Metrics in ML

1. For Regression (continuous outputs)

- **MSE (Mean Squared Error):**
  (\text{MSE} = \frac{1}{n}\sum (y_i - \hat{y}_i)^2)
  - Penalizes large errors heavily.
- **RMSE (Root MSE):**
  - Square root of MSE → easier to interpret in same units as data.
- **MAE (Mean Absolute Error):**
  - Less sensitive to outliers than MSE.
- **R^2 (Coefficient of Determination):**
  - Proportion of variance explained by model (0–1).

#Used in: Linear Regression, forecasting, price prediction.

2. For Classification (discrete outputs)

We use **confusion matrix** terms:

- TP = True Positive
- TN = True Negative
- FP = False Positive
- FN = False Negative

From this:

- **Accuracy = (TP+TN)/(Total)**
  → Good when classes are balanced.
- **Precision = TP / (TP+FP)**
  → Of the predicted positives, how many were actually positive?
  (Example: Precision in spam detection = how many flagged emails were actually spam).
- **Recall = TP / (TP+FN)**
  → Of the actual positives, how many did we catch?
  (Example: Recall in spam detection = how many spam emails we caught).
- **F1-score = 2 * (Precision * Recall) / (Precision + Recall)**
  → Balances precision and recall.
- **ROC-AUC:** Measures performance across different thresholds.
  - AUC close to 1 = good model.
  - AUC = 0.5 = random guessing.

#Used in: Logistic Regression, Decision Trees, Neural Nets.

3. For Unsupervised Learning

- **Silhouette Score:**
  Measures how well clusters are separated (-1 to 1).
- **Davies–Bouldin Index:**
  Lower = better clustering.

#Used in: k-Means, Hierarchical clustering.

Mini Example (Classification)

```
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, confusion_matrix

y_true = [1, 0, 1, 1, 0, 1, 0, 0, 1]

y_pred = [1, 0, 1, 0, 0, 1, 1, 0, 1]


print("Confusion Matrix:\n", confusion_matrix(y_true, y_pred))

print("Accuracy:", accuracy_score(y_true, y_pred))

print("Precision:", precision_score(y_true, y_pred))

print("Recall:", recall_score(y_true, y_pred))

print("F1-score:", f1_score(y_true, y_pred))
```

#**Key Insight:**

- Use **accuracy** only if classes are balanced.
- Use **precision/recall** when cost of FP/FN is high (medical, fraud).
- Use **AUC** when you care about ranking predictions.