**Interview Questions**

**1. What is the difference between Precision and Recall?**

Precision and Recall are evaluation metrics used in classification problems, especially in imbalanced datasets.

**Precision (Positive Predictive Value):**

* Measures how many of the predicted positive cases were actually positive.
* Formula:

Precision

* High precision means fewer false positives.
* Useful when **false positives are costly** (spam detection, fraud detection).

**Recall (Sensitivity or True Positive Rate):**

* Measures how many actual positive cases were correctly identified.
* Formula:

Recall

* High recall means fewer false negatives.
* Useful when **false negatives are costly** (medical diagnosis, safety systems).

**F1-score:** A balance between precision and recall:

* Used when both false positives and false negatives matter.

2. What is Cross-Validation, and Why is it Important in Binary Classification?

**Cross-Validation** is a technique to assess a model’s performance by splitting the dataset into multiple training and testing sets.

**Cross-Validation Important**

1. **Prevents Overfitting:** Ensures the model generalizes well by testing it on different subsets of data.
2. **Reduces Bias & Variance:** Helps avoid a model that is too simplistic (high bias) or too complex (high variance).
3. **Provides Reliable Performance Estimates:** More accurate than a single train-test split.

**Common Types of Cross-Validation:**

* **K-Fold Cross-Validation:**
  + Splits data into **K subsets** (folds), trains on K-1 folds, and tests on the remaining fold.
  + Repeats K times, each fold being used for testing once.
  + Example: 5-Fold CV (80% train, 20% test in each round).
* **Stratified K-Fold:**
  + Ensures class distribution is maintained in each fold (useful for imbalanced datasets).
* **Leave-One-Out Cross-Validation (LOOCV):**
  + Uses **one data point** for testing and the rest for training.
  + Computationally expensive but effective for small datasets.