**INTERVIEW QUESTIONS**

1. **Difference Between Precision and Recall:**

* **Precision**: Precision measures the proportion of correctly predicted positive instances out of all instances predicted as positive. In simpler terms, it answers the question, “Out of all the positive results I predicted, how many were actually correct?”

**Formula**:

Precision=True PositivesTrue Positives+False Positives\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}Precision=True Positives+False PositivesTrue Positives​

Precision is crucial when the cost of false positives is high (e.g., in medical diagnoses where a false positive might lead to unnecessary treatment).

* **Recall (or Sensitivity)**: Recall measures the proportion of actual positive instances that were correctly predicted. It answers the question, “Out of all the actual positives, how many did I predict correctly?”

**Formula**:

Recall=True PositivesTrue Positives+False Negatives\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}Recall=True Positives+False NegativesTrue Positives​

Recall is important when missing actual positives (false negatives) is costly (e.g., detecting cancer).

* **Key Difference**: Precision focuses on the accuracy of positive predictions, while recall focuses on capturing all true positive instances. There is often a trade-off between the two.

1. **What is Cross-Validation, and Why is It Important in Binary Classification?**

* **Cross-Validation**: Cross-validation is a technique for assessing how a machine learning model generalizes to an independent dataset. The most common form is **k-fold cross-validation**, where the dataset is split into k subsets or "folds." The model is trained on k-1 folds and tested on the remaining fold. This process is repeated k times, each time using a different fold as the test set, and the average performance is reported.
* **Importance in Binary Classification**:
  + **Prevents Overfitting**: Cross-validation helps prevent overfitting, which occurs when the model performs well on training data but poorly on unseen data.
  + **Improves Model Generalization**: It provides a more reliable estimate of the model’s performance on new, unseen data, especially in imbalanced datasets (a common challenge in binary classification).
  + **Efficient Use of Data**: Cross-validation allows you to use the entire dataset for both training and validation, which is particularly important when the dataset is small.