1. What is the difference between precision and recall?

Ans:

**Precision** and **recall** are two key metrics used to evaluate the performance of a classification model, particularly in scenarios with imbalanced datasets. They offer distinct perspectives on a model's accuracy:

**Precision:**

* **Focus: The proportion of true positive predictions among all predicted positive instances.**
* **High precision means the model rarely predicts a class as positive when it's actually negative.**

**Recall:**

* **Focus: The proportion of true positive predictions among all actual positive instances.**
* **High recall means the model finds most of the actual positive instances.**

**2. What is cross-validation, and why is it important in binary classification?**

Ans:

**Cross-validation** is a statistical technique used to evaluate the performance of a machine learning model by partitioning the dataset into subsets and repeatedly training and testing the model on these subsets. It provides a reliable estimate of model performance on unseen data, ensuring the model's generalizability.

**Why is it important in binary classification ?**

i. Prevents Overfitting

Cross-validation helps assess whether the model is overfitting the training data.

A model that performs well on the training set but poorly on validation folds indicates overfitting.

ii. Reliable Performance Estimate

It provides a more robust measure of a model's performance on unseen data compared to a simple train-test split. This is crucial in binary classification, where small imbalances in class distribution can significantly affect performance metrics.

iii. Handles Limited Data

When the dataset is small, cross-validation makes efficient use of all available data by using different subsets for training and testing in each iteration.

iv. Balances Class Distribution

Stratified k-fold cross-validation ensures that both positive and negative classes are proportionally represented in each fold, mitigating the effects of class imbalance.

v. Compares Models

Cross-validation provides a consistent framework for comparing multiple models or hyperparameter settings, enabling the selection of the best-performing model.

vi. Supports Hyperparameter Tuning:

Techniques like grid search or random search use cross-validation to evaluate combinations of hyperparameters, ensuring the chosen configuration generalizes well.