**Interview Questions:**

1. What is the difference between precision and recall?

**Precision** and **Recall** are two key metrics used in evaluating classification models, especially in cases like detecting fraud or diagnosing diseases.

**In simple terms:**

* **Precision** answers: "Out of all the predicted positives, how many were actually correct?"
  + It focuses on the **quality** of positive predictions.
  + High precision means fewer false positives (wrongly predicted positives).

**Example**: If a model predicts 10 people have a disease, and 8 of them actually do, the precision is 80%.

* **Recall** answers: "Out of all the actual positives, how many did we correctly predict?"
  + It focuses on **catching all actual positives**.
  + High recall means fewer false negatives (missed actual positives).

**Example**: If there are 10 people with a disease, and the model correctly identifies 8 of them, the recall is 80%.

**In summary:**

* **Precision** is about being **accurate** when you predict something positive.
* **Recall** is about **finding** as many actual positives as possible.

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

**Cross-validation** is a technique used to test how well a machine learning model will perform on unseen data. Instead of using one big chunk of data for training and another for testing, it splits the dataset into smaller pieces (called **folds**) to make better use of the data.

The most common method is **k-fold cross-validation**:

1. The data is divided into **k** equally-sized parts (folds).
2. The model is trained on **k-1** parts and tested on the remaining one.
3. This process repeats **k** times, each time using a different part as the test set.
4. The final performance is the average of the results from all folds.

**Why is it Important in Binary Classification?**

In **binary classification** (where there are only two possible outcomes, like "yes" or "no"), cross-validation is crucial because:

1. **Avoids overfitting**: Cross-validation helps ensure the model isn't just memorizing the training data but can generalize well to unseen data.
2. **Better use of data**: Instead of having a small test set, cross-validation uses the whole dataset for both training and testing across different iterations, giving a more reliable estimate of the model’s performance.
3. **Reliable evaluation**: Since the model is evaluated multiple times on different data splits, you get a more **stable and accurate** measure of how the model will perform in the real world.

**In simple terms:**

Cross-validation helps you **check** if your binary classifier is good at making predictions without getting fooled by the training data. It ensures your model will perform well when faced with new data, not just the data you used to train it.