**Interview Questions**

**1. Explain Bagging and Boosting methods. How are they different from each other?**

**Bagging (Bootstrap Aggregating):**

An ensemble learning method that reduces variance and avoids overfitting.

Creates multiple subsets of the original dataset using bootstrap sampling (random sampling with replacement).

Trains a separate model on each subset independently.

Aggregates (e.g., majority voting for classification or averaging for regression) to make the final prediction.

Example: Random Forest is a popular bagging algorithm using decision trees.

**Boosting:**

An ensemble learning method that reduces bias and improves weak models.

Trains models sequentially, where each new model focuses on correcting the mistakes of previous models.

Assigns higher weights to misclassified data points to improve accuracy.

Combines all models to make a strong prediction.

Example: AdaBoost, Gradient Boosting, XGBoost, LightGBM

**2. Explain How to Handle Imbalance in Data**

**Data-Level Methods (Resampling Techniques):**

Oversampling (SMOTE - Synthetic Minority Over-sampling Technique):

Generates synthetic samples for the minority class to balance the dataset.

Prevents overfitting by creating diverse samples instead of duplicating data.

**Undersampling:**

Removes samples from the majority class to balance the dataset.

Risk: May lead to information loss if too many samples are removed.

**Hybrid Sampling (SMOTE + Tomek Links):**

A combination of oversampling and undersampling.

SMOTE adds synthetic data, and Tomek Links removes noisy majority samples.