**Core Machine Learning Questions**

1. **Difference between Supervised, Unsupervised, and Reinforcement Learning**
   * ***Supervised Learning:* Model learns from labeled data. E.g., spam email detection.**
   * ***Unsupervised Learning:* Model finds patterns in unlabeled data. E.g., customer segmentation via clustering.**
   * ***Reinforcement Learning:* Model learns through rewards and penalties. E.g., training a robot to walk.**
2. **Handling Imbalanced Datasets**
   * **Use resampling methods: oversampling (SMOTE), undersampling.**
   * **Change loss function (e.g., class-weighted loss).**
   * **Use evaluation metrics like F1-score, precision, recall.**
   * **Collect more data or use ensemble techniques.**
3. **Bagging vs Boosting**
   * ***Bagging:* Trains multiple models in parallel (e.g., Random Forest). Reduces variance.**
   * ***Boosting:* Trains models sequentially (e.g., XGBoost), focusing on previous errors. Reduces bias.**
   * **Boosting generally performs better but can overfit more easily.**
4. **Bias-Variance Tradeoff**
   * **High bias = underfitting, high variance = overfitting.**
   * **Goal is to find a sweet spot where both are balanced.**
   * **Use cross-validation to tune complexity. E.g., choosing depth in decision trees.**
5. **Evaluating Classification Beyond Accuracy**
   * **Use confusion matrix, precision, recall, F1-score, ROC-AUC.**
   * **Accuracy fails on imbalanced datasets. E.g., 99% accuracy on dataset with 99% of one class is misleading.**
6. **Underfitting vs Overfitting**
   * ***Underfitting:* Model too simple, high training error. Fix with more features or complexity.**
   * ***Overfitting:* Model too complex, fits noise. Fix with regularization, pruning, or more data.**
   * **Use learning curves to diagnose.**