① Explain Bagging and Boosting methods. How is it different from each other.

Bagging vs Boosting understanding different.

Bagging & Boosting are two popular ensemble learning techniques used to improve the performance and accuracy of machine learning model. While both methods combine multiple weak learners to create a strong predictive

1). Bagging:-

Bagging aim reduce varience by training multiple model independently on different subsets of the datasets and then averaging their predictions.

Key features:-

* user Bootstrap to create different subsets of data
* Trains multiple independent models in parallel.

Ex:-

*Random forest is an extension of bagging applied to decision Trees.

*Each tree is trained on a different Bootstrap sample.

2) Boosting:-

Boosting aims to reduces bias by training model learns from the errors of the privious ones.

Key features:-

* Model's are trained sequentially, each correcting the mistakes of it's predecessor.

* Can overfit not carefully tuned.

Ex:-

1) Ada Boost :-

* Assign higher weight to misclassified point.

2) Gradient Boosting :-

* Models learn sequentially by minimizing a loss Function.

② Explain how to handle imbalance in the data.

Handling imbalanced Data in ML

Imbalanced data occurs when one class significantly outcomes another in a classification problem.

1) Data - Entry Techniques

There data imbalance class distribution.

A) Over Sampling.

* SMOTE :- Generate synthetic sample for the minority class by interpolating existing instance

* ADASYN :- Similar to SMOTE but focuses more on harder to learn minority instance.

* Random oversampling :- Duplicates existing minority samples.

B) Under Sampling.

* Random undersampling :-

Remove random samples from the majority class.

* Nearmiss Algorithm :-
Select majority class samples.

2) Algorithm level Techniques.
Instead of modifying data, these methods fits objective function.

A) Cost - Sensitive learning.
* Assign better misclassification costs to the minority class. The model more sensitive error.

```
from sklearn.ensemble import *
RandomForestClassifier
RFC = RandomForestClassifier(class_
                             weight =
                             "balanced"
```

B) Ensemble Techniques.
* Balanced Random Forest :-
uses bootstrap sample with equals class representation
* Easy Ensemble :-
Combine under sampling & boosting.

3) Evaluation metrics for Imbalanced data.

* Precision & Recall
* F1-Score
* ROC_AUC
* PR_AUC

From sklearn.metrics import Classification report.

classification_report(y_Pred, y_test)

4) Alternative Approaches :-

* Anomaly Detection :-

The minority very rare, treat anomaly detection problem.

Choosing the Right methods.

* Small dataset - SMOTE-Xgboost sensitive learning
* large dataset - under sampling + ensemble method.
* Extreme imbalance :- consider anomaly detection techniques.