1.R squared is the better as it gives us better measure of fitness as it is relative.

2.total sum of squares=measures deviation of observed values from their mean.

Explained sum of squares=measures the amount of variability in the dependent variable.

Residual sum of squares=the difference between actual values and predicted values.

Their relation=

TSS=ESS+RSS.

3.we need REGULARIZATION incase of

1.high dimensional data.

2.handling complex models.

3.Avoid risk of overfitting.

4.GINI IMPURITY INDEX

To measure the degree of uncertainity in the classification problem.

5.unregularised decision trees are highly prone to overfitting when dataset is complex and due to this the decision tree can grow very deep and complex.

6.ensemble technique in machine learning is the use of multiple model for predictions.

7.bagging is building multiple models and then aggregating them to make predictions.

Whereas boosting technique uses sequential operations to correct the error in the previous models.

8. **Out-of-Bag error** is a built-in method of evaluating the performance of a **Random Forest** without the need for a separate validation or test set. It serves as an internal cross-validation mechanism for Random Forests.

9.k fold cross validation is the method used to aceess the performance of machine learning model.

It helps ensure that the model generalizes well to unseen data by systematically dividing the dataset into multiple subsets and evaluating the model on different subsets.

10. **Hyperparameter tuning** is the process of optimizing the hyperparameters of a machine learning model to improve its performance.