1. R-squared and RSS both use measurement to assess the goodness of fit of regression model but R-squared has some limitation, For example, it can be artificially inflated by adding more predictors to model, even if those predictors do not have a meaningful relationship with the dependent variable that's why RSS is good to use.
2. TSS = ESS + RSS, where TSS is total sum of square, ESS is Explai9ned sum of square and RSS is Regression sum of squares. The aim of Regression Analysis is explain the variation of dependent variable.
3. When the ML model try to capture each point of data which is not really not represent true properties this situation is called overfitting. Overfitting made our model slow. To avoid overfitting we use regularization.
4. Gini-impurity index measures how often a randomly chosen element of a set would be incorrectly labeled if it were labeled randomly and independently according to the distribution of labels in the set.
5. Decision Trees are prone to over-fitting. A decision tree will always overfit the training data if we allow it to grow to its max depth. Overfitting in decision trees occurs when the tree becomes too complex and captures the noise in the training data, rather than the underlying pattern. This can lead to poor generalization performance on new unseen data.
6. Ensemble technique is a learning paradigm where multiple model are trained to solve the same problem and combined to get better result.
7. Bagging aims to reduce the variance of the model and Boosting aims to reduce the bias from underfitting the data.
8. Out-of -bag error is a method of measuring the prediction error of random forest, boosted decision trees, and other machine learning models utilizing bootstrap aggregating.
9. K-fold cross-validation is a technique for evaluating predictive models. The dataset is divided into k subsets or folds. The model is trained and evaluated k times, using a different fold as the validation set each time. Performance metrics from each fold are averaged to estimate the model's generalization performance.
10. Hyperparameter tuning consists of finding a set of optimal hyperparameter values for a learning algorithm while applying this optimized algorithm to any data set. That combination of hyperparameters maximizes the model's performance, minimizing a predefined loss function to produce better results with fewer errors.
11. When the learning rate is too large, gradient descent can suffer from divergence. This means that weights increase exponentially, resulting in exploding gradients which can cause problems such as instabilities and overly high loss values.
12. Logistic regression is a linear classifier that produces a linear decision boundary. It can't solve non-linear problems because it assumes a linear relationship between the input features and the output. This means it can't capture the complexity and non-linearity of the data.
13. Adaboost learns from the misclassified samples; Gradient boosting learns from samples with large pseudo-residuals.
14. the bias–variance tradeoff describes the relationship between a model's complexity, the accuracy of its predictions, and how well it can make predictions on previously unseen data that were not used to train the model.
15. RBF kernels is used to preform transformation when there is no prior knowledge about data. Polynomial kernels is used to represents the similarity of vectors in the training set of data in feature space over polynomials of the original variables used in the kernels. Linear kernels is used when data is linearly spread.