Q-1:- The residual sum of squares (RSS) is the sum of the squared distances between your actual and the predicted values.

R-squared is the absolute amount of variation as a proportion of total variation.

Q-2:- TSS:- The sum of squares is a form of regression analysis to determine the variance from data points from the mean. If there is a low sum of squares, it means there's low variation and A higher sum of squares indicates higher variance.

$$\sum_{i=1}^{n} (y_i - \bar{y})^2$$

ESS:- The explained sum of squares (ESS) is the sum of squares of the deviations of the predicted values from the mean value of a response variable.

==RSS:==- Residual sum of squares measures the level of variance in the error term, or residuals, of a regression model. The smaller the residual sum of squares, the better your model fits your data, and the greater the residual sum of squares, the poorer your model fits your data.

==Equation:-==

$$RSS = \sum_{i=1}^{n} (y^i - f(x_i))^2$$

Q-3:- Regularization refers to the techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting.

Using Regularization, we can fit our machine learning model appropriately on a given test dataset and hence reduce the errors thereof.

Q-4:- Gini Impurity is calculated by subtracting the sum of the squared probabilities of each class from one.

It favours mostly the larger partitions and are very simple to implement. Also ,it calculates the probability of a certain randomly selected feature that was classified incorrectly before.

Q-5:- Decision trees are prone to overfitting, especially when a tree is particularly deep. This is due to the amount of specificity we look at leading to smaller sample of events that meet the previous assumptions.

Q-6:- Ensemble methods are techniques that aim at improving the accuracy of results in models by combining multiple models instead of using a single model. The combined models increase the accuracy of the results significantly.

Q-7:- Bagging:- It is basically a homogeneous weak learners model that learns from each other independently in parallel and combines them for determining the model average.

It decreases variance and helps to avoid overfitting.

Boosting:- It is an ensemble modelling technique that attempts to build a strong classifier from the number of weak classifiers.

It is done by building a model by using weak models in series.

Q-8:- The Out Of Bag error is the average error for each calculated using predictions from the trees that do not contain in their respective bootstrap sample.

It allows the Random Forest Classifier to be fit and validated while being trained.

Q-9:- K-fold Cross-Validation is when the dataset is split into a K number of folds and is used to evaluate the model's ability when given the new data.

K refers to the number of groups the data sample is split into.

Q-10:- Hyperparameter tuning basically means finding a set of optimal hyperparameter values for a learning algorithm while applying this optimized algorithm to any data set.

The combination of hyperparameters maximizes the model's performance, minimizing a predefined loss function to produce better results with fewer errors.

Q-11:- A learning rate that is too high can cause the model to converge too quickly to a suboptimal solution,and also if the learning rate which is too small can cause the process to get stuck.

Q-12:- Non-linear problems will not be solved with logistic regression because it has a linear decision surface and Logestic regression requires average or no multicollinearity between independent variables.

Q-13:-

==AdaBoost:-== This method automatically adjusts its parameters to the data based on the actual performance in the current iteration. Meaning, both the weights for re-weighting the data and the weights for the final aggregation are re-computed iteratively.

==Gradient Boosting:-== This technique yields a direct interpretation of boosting methods from the perspective of numerical optimisation in a function space and generalises them by allowing optimisation of an arbitrary loss function.

Q-14:- Basically, Bias is the simplifying assumptions made by the model to make the target function easier to approximate.

Variance is the amount that the estimate of the target function will change given the different training data.

Trade-off is tension between the error introduced by the bias and the variance.

Q-15:- <mark>Linear Kernel:-</mark> It is used when the data is Linearly separable that is it can be separated using a single Line. It is mostly used when there are a Large number of Features in a particular Data Set.

<mark>RBF:-</mark> It is mostly used in kernalised learning algorithms.

It is popular because of its similarity to K-Nearest Neighbor Algorithm. It has the advantages of K-NN and overcomes the space complexity problem as RBF Kernel SVM just needs to store the support vectors during training and not the entire dataset.

Polynomial Kernel:- A polynomial kernel is a kind of SVM kernel that uses a polynomial function to map the data into a higher-dimensional space. It does this by taking the dot product of the data points in the original space and the polynomial function in the new space.