* What is imputation

Imputation is a technique used for replacing the missing data with some substitute value to retain most of the data/information of the dataset

* How can you define mode?

The mode is known to be most frequently occurring value, commonly used for working on categorical data.

* Mention a disadvantage of mean?

The biggest disadvantage is that the mean gets influenced by the outliers (also known as extreme values).

* Mention all the measure used to analyze dispersion?

The measures used for the analysis of dispersion of data are variance, standard deviation and range.

* What does Right Skewed mean?

Right Skewness is a measure that represents the mass of the data distributed towards the right side of the center. A histogram on the data will also enable us to understand the existence of extreme value in the dataset.

* Does Standard Deviation get influenced by Outliers?

The distance of the data points from the center would be affected by outliers. Yes, the Standard Deviation would get effected.

* What is Confusion matrix with respect to machine learning algorithms?

A confusion matrix (or error matrix) is a specific table that is used to measure the performance of an algorithm. It is mostly used in supervised learning; in unsupervised learning, it’s called the matching matrix.

* When will you used classification over Regression?

Classification is used when your target is categorical, while regression is used when your target variable is continuous. Both classification and regression belong to the category of supervised machine learning algorithms.

Examples of classification problems include:

Predicting yes or no

Estimating gender

Breed of an animal

Type of color

Examples of regression problems include:

Predicting the score of a team

Predicting the amount of rainfall

* What is Random Forest?

A ‘Random Forest’ is a supervised machine learning that is generally used for classification problems. It operates by constructing multiple decision trees during the training phase. The random forest chooses the decision of the majority of the trees as the final decision.

* What do you understand by the F1 score?

The F1 score is a metric that combines both Precision and Recall. It is also the weighted average of precision and recall.

The F1 score can be calculated using the below formula:

F1=2\*(P\*R)/(P+R)

The F1 score in one when both Precision and Recall scores are one.

* What do you understand by Type I vs Type II error?

Type I Error: Type I error occurs when the null hypothesis is true and we reject it.

Type II Error: Type II error occurs when the null hypothesis is false and we accept it.