* How to handle missing or computed data in dataset?

1. Isnull() – detecting the missing values.
2. dropna() – remove columns or rows with null values
3. fillna() – used to fill values

* When should classification be used over regression?

Classification is chosen over regression when the output of the model needs to belonging of the data points in a dataset to a particular category.

* What is ROC curve and what does it represent?

ROC – Receiver Operating Characteristic. It used to graphically represent the trade off between true-false positive rates.

ROC the area under the AUC idea about the accuracy of the model.

* What is Type-I and Type-II errors?

Type-I = False positive – It is error when the outcome of a test shows the non-acceptance of a true condition.

Type-II = False negative – It is error when the outcome of a test shows the acceptance of false condition.

* What is hypothesis test?

It is statistical technique that help scientist test the validity of their claims about real-world/real-life events.

* What is Entropy in machine learning?

Entropy is machine learning measures the randomness in data that needs to be processed. The more entropy in given data, the more difficult it becomes to draw any useful insights.

* What is Epoch in machine learning?

It is used to handle to the count of passes in a given dataset where the machine learning has its job. When there is large chunk of data it grouped into several batches. When all batches go through given model and process is referred to as iteration. If the size comprises the complete training dataset, then it counts of iteration as epoch.

* What is PCA in machine learning?

1. Removing irrelevant dimension
2. Keeping only the most relevant dimension

* What is machine learning?

Machine learning – It given the ability to make business decision without any external help, using knowledge gained from the past data.

* What is deep learning?

It allows machine to make various business decision using artificial neural network, which is why it needs vast amount of data for traning.