1.what is data?

A: Data is information that has been translated into a form that is efficient for movement or processing.

2.what is data science?

A: Data science is the study of data to extract meaningful insights for business.it is a multidisciplinary approach that combines principles and practices from the fields of mathematics, statistics, artificial intelligence, and computer engineering to analyze large amounts of data.

3.what is data preprocessing?

A: Data preprocessing is the concept of changing the raw data into a clean data set

4.What libraries utilized to work with data preprocessing?

A: Some of the popular libraries used for data preprocessing in python include pandas, scikit-learn, and matplotlib.

5.What are roles of data scientist?

A: A data scientist is a professional who collects large amounts of data using analytical, statistical, and programmable skills.

6.what is machine learning?

A: Machine learning is the subset of artificial intelligence that focuses on building systems that learn-or improve performance-based on the data they consume.

7.why machine learning?

A: Simply put, machine learning allows the user to feed a computer algorithm an immense amount of data and have the computer analyze and make data-driven recommendations and decisions based on only the input data.

8.Purpose of machine learning?

A: The purpose of machine learning is to figure out how we can build computer systems that improve over time and with repeated use. this can be done by figuring out the fundamental laws that govern such learning processes.

9.Machine learning current examples?

A: An algorithm would be trained with pictures of dogs and other things, all labeled by humans, and the machine would learn ways to identify pictures of dogs on its own.

10.How many types of machine learning model/on what based machine learning model selected brief?

A: There are four basic types of machine learning

1.supervised learning 2. unsupervised learning 3.semisupervised learning 4.reinforcement learning.

11.What is scikit learn?

A: Scikit-learn is an open source data analysis library, and the gold standard for machine learning win the python ecosystem.

12.List out data preprocessing steps?

A: The steps involved in data preprocessing are: data collection, data transformation, data reduction, data discretization, data discretization, data normalization or data standardization, feature selection, and data representation

13.what is over \_fit?

A: Overfitting is a modeling error in statistics that occurs when a function is too closely aligned to a limited set of data points.

14.What is under-fit?

A: Underfitting is a scenario in data science where a data model is unable to capture the relationship between the input and output variables accurately, generating a high error rate on both the training set and unseen data.

15.Difference between machine learning and deep learning?

A: Machine learning and deep learning are both types of AI. In short, machine learning is AI that can automatically adapt with minimal human interference. Deep learning is a subset of machine learning that uses artificial neural networks to mimic the learning process of the human brain.

16.Brief steps involved in machine learning model approach?

A: There are seven steps involved in the machine learning model

1.collecting data 2. Preparing the data 3. Choosing a Model 4. Training the model

5.Evaluating the Model 6. parameter tuning 7. Making predictions

17.What the default ration for train/test sets?

A: The optimal spilt ratio depends on various factors. The rough standard for train-validation-test splits is 60-80% training data, 10-20% validation data, and 10-20% test data.

18.Difference between supervised/unsupervised/reinforcement learning?

A: Supervised Learning

1.Input data is labelled.

2.learn pattern of inputs and their labels

3.Finds a mapping equation on input data and its labels.

4.Model is built and training prior to testing.

Unsupervised Learning

1.Input data is not labelled.

2.Divide data into classes.

3.Finds similar features in input data to classify it into classes.

4.Model is built and trained prior to testing.

Reinforcement Learning

1.Input data is not predefined.

2.Find the best reward between a start and an end state.

3.Maximizes reward by assessing the results of state-action pairs.

4.The model is trained and tested simultaneously.

19.Decribe independent variable /dependent variable?

A: Independent variables aren’t affected by any other variables that the study measures.

Example: If someone was studying the effects of tooth-brushing on the incidence of caries, the tooth-brushing would be the independent variables.

Dependent variables: It is something that depends on other factors.

Example: A test score could be a dependent variable because it could change depending on several factors such as how much you studied.