1. What does one mean by the term "machine learning"?

A. Machine Learning is defined as the science of programming computers using various statistical methods, so that they can learn from the data and provide useful insights of the data.

2.Can you think of 4 distinct types of issues where it shines?

1. It helps in understanding the patterns in complex datasets and make predictions.

2. It is used in anomaly detection, where models are trained to detect any unusual activity.

3. It is used in aggregating data by forming clusters, which play a vital role in data segmentation.

4. It can be used in sentiment analysis problems, where it can understand text and detect the sentiment of the text without any human intervention.

3.What is a labeled training set, and how does it work?

A. A labelled training set is defined as the dataset which consists of a tag for each record which describes about the data.

If we take the example of an sentiment analysis modelling the dataset provided is labelled since for each review a tag is assigned to specify the sentiment.

4.What are the two most important tasks that are supervised?

1. Classification.

2.Regression.

5.Can you think of four examples of unsupervised tasks?

1. Fraud detection in financial transactions or insurance claims

2. Recommending products based on past purchase history.

3. Customer segmentation for targeted marketing campaigns

4. Improving the efficiency of machine learning algorithms by reducing the number of features.

6.State the machine learning model that would be best to make a robot walk through various unfamiliar terrains?

A. Reinforcement Learning model. Reinforcement learning is a type of machine learning that involves an agent interacting with an environment to learn how to take actions that maximize a cumulative reward signal. In this case, the robot would be the agent, and the terrain would be the environment.

7.Which algorithm will you use to divide your customers into different groups?

A. I will use clustering algorithms to divide customers into different groups.

8.Will you consider the problem of spam detection to be a supervised or unsupervised learning problem?

A. Spam classification is an example of supervised learning. While creating a spam classification model we input the model with data which consists of emails which are classified as spam or ham. The model learns based on the labelled data.

9.What is the concept of an online learning system?

A. Unlike batch learning, in online learning system we train the model incrementally by feeding it with new data. When a new data arrives the model need not be trained again on the entire dataset, It can be updated incrementally with new data.

10.What is out-of-core learning, and how does it differ from core learning?

A. Out-of-core algorithms can handle vast quantities of data that cannot fit in a computer's main memory. An out-of-core learning algorithm chops the data into mini-batches and uses online learning techniques to learn from these mini-batches.

11.What kind of learning algorithm makes predictions using a similarity measure?

A. The algorithm like K-Nearest Neigbors makes predictions using a similarity measure.

12.What's the difference between a model parameter and a hyperparameter in a learning algorithm?

A. Model parameters are the values that the model learns through training, and they are used to make predictions on new data. Hyperparameters control the behavior of the learning algorithm and can have a significant impact on the performance of the model.

13.What are the criteria that model-based learning algorithms look for? What is the most popular method they use to achieve success? What method do they use to make predictions?

A. Model-based learning algorithms search for an optimal value for the model parameters such that the model will generalize well to new instances. We usually train such systems by minimizing a cost function that measures how bad the system is at making predictions on the training data, plus a penalty for model complexity if the model is regularized. To make predictions, we feed the new instance's features into the model's prediction function, using the parameter values found by the learning algorithm.

14.Can you name four of the most important Machine Learning challenges?

A. Poor quality data, 2. Sampling bias, 3. Irrelevant features, 4. Overfitting the trining data to the model.

15.What happens if the model performs well on the training data but fails to generalize the results to new situations? Can you think of three different options?

A. Such a situation is called overfitiing.

1. Reduce the number of features, by eliminating irrelevant features.

2. Gather more data and train the model so that it can generalize well.

3. Perform data cleaning and remove noisy data or handle the noisy data.

16.What exactly is a test set, and why would you need one?

A. A test set is a subset of the dataset which is kept aside to evaluate the model. It is used in evaluating the model to determine its accuracy.

17.What is a validation set's purpose?

A. A validation set is used to fine tune the hyperparameters and it is also used to compare different models.

18.What precisely is the train-dev kit, when will you need it, how do you put it to use?

A. The train-dev set is used when there is a risk of mismatch between the training data and the data used in the validation and test datasets (which should always be as close as possible to the data used once the model is in production). The train-dev set is a part of the training set that's held out (the model is not trained on it). The model is trained on the rest of the training set, and evaluated on both the train-dev set and the validation set. If the model performs well on the training set but not on the train-dev set, then the model is likely overfitting the training set. If it performs well on both the training set and the train-dev set, but not on the validation set, then there is probably a

significant data mismatch between the training data and the validation + test data, and you should try to improve the training data to make it look more like the validation + test data.

19.What could go wrong if you use the test set to tune hyperparameters?

A. We are overfitting the test data set and our model is being biased towards the test dataset. It will fail to generalize unseen data.