1. What are the key tasks that machine learning entails? What does data pre-processing imply?
Ans:- A machine learning task is the type of prediction or inference being made, based on the problem or question that is being asked, and the available data. For example, the classification task assigns data to categories, and the clustering task groups data according to similarity. Data preprocessing is the concept of changing the raw data into a clean data set. The dataset is preprocessed in order to check missing values, noisy data, and other inconsistencies before executing it to the algorithm. Data must be in a format appropriate for ML.

2. Describe quantitative and qualitative data in depth. Make a distinction between the Two.
Ans-Quantitative data is numbers-based, countable, or measurable. Qualitative data is interpretation-based, descriptive, and relating to language. Quantitative data tells us how many, how much, or how often in calculations. Qualitative data can help us to understand why, how, or what happened behind certain behaviors.
Quantitative researchQuantitative research is expressed in numbers and graphs. It is used to test or confirm theories and assumptions. This type of research can be used to establish generalizable facts about a topic. Common quantitative methods include experiments, observations recorded as numbers, and surveys with closed-ended questions.

Quantitative research is at risk for research biases including information bias, omitted variable bias, sampling bias, or selection bias.

Qualitative researchQualitative research is expressed in words. It is used to understand concepts, thoughts or experiences. This type of research

enables you to gather in-depth insights on topics that are not well understood.
Common qualitative methods include interviews with open-ended questions, observations described in words, and literature reviews that explore concepts and theories.

Qualitative research is also at risk for certain research biases including the Hawthorne effect, observer bias, recall bias, and social desirability bias.
4. What are the various causes of machine learning data issues? What are the ramifications?

Ans:-Noisy data, incomplete data, inaccurate data, and unclean data lead to less accuracy in classification and low-quality results. Hence, data quality can also be considered as a major common problem while processing machine learning algorithms.

5. Demonstrate various approaches to categorical data exploration with appropriate examples.

Categorical Variable : Such variables take on a fixed and limited number of possible values. For example – grades, gender, blood group type, etc. Also, in the case of categorical variables, the logical order is not the same as categorical data e.g. “one”, “two”, “three”. But the sorting of these variables uses logical order. For example, gender is a categorical variable and has categories – male and female and there is no intrinsic ordering to the categories. A purely categorical variable is one that simply allows you to assign categories, but you cannot clearly order the variables.

6. How would the learning activity be affected if certain variables have missing values? What can be done about it?

Ans:- Many machine learning algorithms fail if the dataset contains missing values. However, algorithms like K-nearest and Naive Bayes support data with missing values. You may end up building a biassed machine learning model, leading to incorrect results if the missing values are not handled properly.

7. Describe the various methods for dealing with missing data values in depth.

When dealing with missing data, data scientists can use two primary methods to solve the error: imputation or the removal of data. The imputation method develops reasonable guesses for missing data. It's most useful when the percentage of missing data is low.

8. What are the various data pre-processing techniques? Explain dimensionality reduction and function selection in a few words.

Ans:- Important Data Preprocessing Techniques

- Data Cleaning.
- Dimensionality Reduction.
- Feature Engineering.
- Sampling Data.
- Data Transformation.

- Imbalanced Data.

Dimensionality reduction is a machine learning (ML) or statistical technique of reducing the amount of random variables in a problem by obtaining a set of principal variables.
Dimensionality reduction is a technique used in data mining to map high-dimensional data into a low-dimensional representation in order to visualise data and find patterns that are otherwise not apparent using traditional methods.