Machine learning 05

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

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.

3. Create a basic data collection that includes some sample records. Have at least one attribute from each of the machine learning data types.

1. Ans:- Determine What Information You Want to Collect. The first thing you need to do is choose what details you want to collect. ...
2. Set a Timeframe for Data Collection. ...
3. Determine Your Data Collection Method. ...

4. What are the various causes of machine learning data issues? What are the ramifications?

* Ans:- 1) Understanding Which Processes Need Automation. ...
* 2) Lack of Quality Data. ...
* 3) Inadequate Infrastructure. ...
* 4) Implementation. ...
* 5) Lack of Skilled Resources.

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

Ans:- Data exploration is the first step of data analysis used to explore and visualize data to uncover insights from the start or identify areas or patterns to dig into more. Using interactive dashboards and point-and-click data exploration, users can better understand the bigger picture and get to insights faster.

6. How would the learning activity be affected if certain variables have missing values? Having said that, what can be done about it?

Ans:- Missing value can bias the results of the machine learning models and/or reduce the accuracy of the model

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

Ans:- There are 2 primary ways of handling missing values: Deleting the Missing values. Imputing the Missing Values.

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

* Ans:- Data Cleaning.
* Dimensionality Reduction.
* Feature Engineering.
* Sampling Data.
* Data Transformation.
* Imbalanced Data.

9.

i. What is the IQR? What criteria are used to assess it?

Ans:- The interquartile range is calculated in much the same way as the range. All you do to find it is subtract the first quartile from the third quartile: IQR = Q3 – Q1. The interquartile range shows how the data is spread about the median.

ii. Describe the various components of a box plot in detail? When will the lower whisker surpass the upper whisker in length? How can box plots be used to identify outliers?

Ans:- A box and whisker plot—also called a box plot—displays the five-number summary of a set of data. The five-number summary is the minimum, first quartile, median, third quartile, and maximum. In a box plot, we draw a box from the first quartile to the third quartile. A vertical line goes through the box at the median.

10. Make brief notes on any two of the following:

1. Data collected at regular intervals :- Interval data, also called an integer, is defined as a data type which is measured along a scale, in which each point is placed at equal distance from one another. Interval data always appears in the form of numbers or numerical values where the distance between the two points is standardized and equal.

2. The gap between the quartiles :- The median is considered the second quartile (Q2). The interquartile range is the difference between upper and lower quartiles. The semi-interquartile range is half the interquartile range. When the data set is small, it is simple to identify the values of quartiles.

1. Make a comparison between:

1. Data with nominal and ordinal values :- Nominal data is classified without a natural order or rank, whereas ordinal data has a predetermined or natural order. On the other hand, numerical or quantitative data will always be a number that can be measured.

2. Histogram and box plot :- Histograms are a special kind of bar graph that shows a bar for a range of data values instead of a single value. A box plot is a data display that draws a box over a number line to show the interquartile range of the data. The 'whiskers' of a box plot show the least and greatest values in the data set.