i) **Define about any data with an unknown form is called:** Answer: b) Unstructured

ii) **Demonstrate about veracity:** Answer: c) Manage data

iii) **The primary aspect of Big Data is to provide:** Answer: a) Demanding data rapidly

iv) **Define about Big Data approaches which are reducing the:** Answer: b) Identify the data we need

v) **State about master job manager to oversee:** Answer: d) All of these

vi) **Demonstrate about categories of clustering methods:** Answer: d) All of these (Partitioning methods, Hierarchical methods, Grid-based methods)

vii) **Describe about semi-structured data, represented in an:** Answer: a) XML file

viii) **Justify which of the following function is used for k-means clustering:** Answer: a) k-means

Short Answer Type Questions:

2. **How big data analysis helps to predict the stock market:** Big data analysis can help predict stock market trends by processing and analyzing vast amounts of structured and unstructured data from various sources. This data can include historical stock prices, news articles, social media sentiment, economic indicators, and more. Predictive analytics models can use this data to identify patterns, correlations, and trends that human analysts might miss. By analyzing this data, machine learning models can make predictions about future stock prices and market trends.

3. **The Twitter data satisfies characteristics of big data:** Twitter data satisfies the characteristics of big data because it is typically large in volume, generated at a high velocity, and comes in diverse formats. Twitter generates a massive number of tweets in real-time, and this data is unstructured, making it challenging to process and analyze. Furthermore, the data can vary in terms of its reliability, leading to veracity issues. Big data technologies and analytics are often required to manage and gain insights from Twitter data effectively.

4. **Big Data in eCommerce and advantages:** Big Data in eCommerce involves analyzing large volumes of data related to customer behavior, sales, and product performance. Advantages include personalizing customer experiences, optimizing inventory management, improving marketing strategies, and enhancing fraud detection.

5. **Features of Confusion Matrix:**

   - **True Positive (TP):** The number of correct positive predictions.

   - **True Negative (TN):** The number of correct negative predictions.

   - **False Positive (FP):** The number of incorrect positive predictions (Type I error).

   - **False Negative (FN):** The number of incorrect negative predictions (Type II error).

6. **Pros and cons of K-means clustering algorithm:**

   - Pros:

     - Simple and easy to implement.

- Efficient for large datasets.

- Suitable for a wide range of applications.

- Provides hard clustering, where each data point belongs to one cluster.

- Cons:

  - Sensitive to the initial choice of centroids.

  - Assumes clusters are spherical and equally sized, which may not be the case in some datasets.

  - May converge to a local minimum.

  - Not suitable for non-linearly separable data.

7. **Unsupervised learning method with an example:** An unsupervised learning method is clustering, which involves grouping similar data points together. For example, in customer segmentation, you can use clustering to group customers with similar purchasing behavior. If you have an eCommerce dataset, you can apply clustering to identify segments like "frequent shoppers," "occasional shoppers," and "bargain hunters."