

1. What do you mean by Machine Learning?

Machine Learning (ML) is a subset of artificial intelligence that enables systems to learn from data and make decisions or predictions without being explicitly programmed.

2. Why is machine learning important for Big Data?

ML helps extract meaningful patterns and insights from vast and complex Big Data, enabling automated decision-making and predictive analytics.

3. What are the different types of machine learning?

The main types are:

- Supervised learning.
- Unsupervised learning.
- Reinforcement learning.
- Semi-supervised learning.

4. What is classification in machine learning?

Classification is a supervised learning task where the goal is to predict a categorical label (e.g., spam or not spam) based on input data.

Section - B

1. What are the challenges of applying machine learning to Big Data?

ans: Challenges include:

- High Computational cost
- Data variety and quality issues

- Data Privacy Concerns

- Model Scalability and deployment difficulties.

The sheer volume of data requires high computational power and distributed processing systems.

The variety of data formats (text, images, audio etc) complicates data integration and analysis.

Velocity - the speed at which new data is generated makes it hard to keep models up to date in real time.

Ensuring data quality and removing noise from large datasets is difficult.

Scalability of machine learning algorithms and ensuring data privacy and security are major concerns.

2. Explain the difference between supervised, unsupervised and Reinforcement learning.

one In supervised learning, the model is trained on labeled data where the input and output are known.

Examples include classification (spam detection) and regression (price prediction).

Unsupervised learning: On the other hand deals with unlabeled data and aims to find hidden patterns or grouping for example, customer segmentation using clustering.

Reinforcement learning is based on agents learning through trial and error by receiving rewards or penalties. It's used in environments like game playing and robotics. Each learning type serves different purposes based on the data and problem.

Section - c

1. Discuss the importance of Data Exploration and Data Preparation in machine learning.

- Data Exploration and Data Preparation are critical initial steps in any machine learning project.

Data Exploration involves understanding the structure, patterns, and distribution of the data using summary statistics and visualization techniques. This helps identify data quality issues such as missing values, outliers and inconsistencies.

Data Preparation includes cleaning, transforming, and encoding the data to make it suitable for machine learning algorithms. Proper exploration and

Preparation ensures the models are trained on high-quality, meaningful data, which directly impacts the accuracy and performance of the final results.

2. Explain Regression in machine learning with an example.

Regression is a type of supervised learning techniques used to predict continuous numerical outcomes based on input features.

One of the most common regression methods is Linear Regression, where the relationship between the dependent and independent variables is modeled with a straight line.

For example: Predicting a house price based on feature like area, number of rooms, and location is a regression problem.

The model learns the coefficient during training uses them to make predictions like forecasting sales, estimating risks, and financial modeling.