1. What is Machine Learning?

Machine Learning (ML) is a branch of Artificial Intelligence (AI) that allows systems to automatically learn and improve from experience without being explicitly programmed.
It focuses on building algorithms that can recognize patterns, make decisions, and predict outcomes based on data.

2. Give one real-life example of Machine Learning.

A common example is **spam email detection**.
Email providers (like Gmail) use ML algorithms to learn from data (spam vs. non-spam emails) and automatically filter out unwanted emails into the spam folder.

---

3. What is the difference between Artificial Intelligence (AI) and Machine Learning (ML)?

- **AI**: The broader concept of creating machines that can mimic human intelligence, perform reasoning, problem-solving, and decision-making.
- **ML**: A subset of AI that focuses specifically on training systems to learn from data and improve performance over time.

Example:

- AI: A self-driving car making decisions about navigation.
- ML: The car's algorithm learning how to recognize stop signs from images.

---

4. What are the types of Machine Learning?

The main types are:

1. **Supervised Learning** – Trained on labelled data.
2. **Unsupervised Learning** – Trained on unlabelled data.
3. **Reinforcement Learning** – Learns through trial and error with feedback (rewards/punishments).
4. **Semi-Supervised Learning** – A mix of labelled and unlabelled data.
5. **Online Learning** – Learns continuously from new data streams.

---

5. What is supervised learning? Give an example.

Supervised learning is when a model is trained on **labelled data** (input + output provided). The algorithm learns the mapping between input (features) and output (labels).

Example: Predicting house prices using features like size, location, and number of rooms.

- Input (features): Size = 1200 sq. Ft., Location = City Centre, Rooms = 3.
- Output (label): Price = ₹50 Lakhs.

---

6. What is unsupervised learning? Give an example.

Unsupervised learning is when the model is trained on **unlabelled data**, meaning only input data is given, and the algorithm tries to find hidden patterns or groupings.

Example: Customer segmentation in marketing.
The algorithm groups customers into clusters based on buying behaviour without knowing in advance which customer belongs to which group.

---

7. What is reinforcement learning? Give an example.

Reinforcement learning is when an agent learns by interacting with its environment and receiving feedback in the form of **rewards** or **penalties**.

Example: A robot learning to walk.

- If it walks correctly → gets a reward.
- If it falls → gets a penalty.

Over time, it learns the best strategy to maximize rewards.

---

8. What is the difference between training data and test data?

- **Training Data**: Used to train the ML model (helps it learn patterns and relationships).
- **Test Data**: Used to evaluate the model's performance on new, unseen data (checks if it generalizes well).

Think of training data like study material and test data like the exam.

---

9. What are features in a dataset?

Features are the **input variables** that describe each data point.
They represent measurable properties used to make predictions.

Example: For predicting student exam scores:
Features could be study hours, sleep hours, and attendance.

---

10. What are labels (targets) in a dataset?

Labels are the **output variables** that the model is trying to predict.

Example: In predicting exam scores:

- Input (features): Study hours, sleep hours.
- Output (label): Actual exam score.

---

11. What is the difference between classification and regression?

- **Classification**: Predicts a **category/class**.
  Example: Whether an email is Spam or Not Spam (Yes/No).
- **Regression**: Predicts a **continuous numerical value**.
  Example: Predicting the price of a house (₹50,000, ₹60,000, etc.).

---

12. What does the term model mean in ML?

A **model** in ML is the final output of the training process. It is the mathematical representation of the real-world problem that can make predictions on new data.

Example: After training on house data, the model can predict prices of unseen houses.

---

13. What is overfitting in ML?

Overfitting happens when a model learns the **training data too well**, including noise and irrelevant patterns.
As a result, it performs well on training data but poorly on test data.

Example: Memorizing past exam questions instead of understanding the concepts.

---

14. What is underfitting in ML?

Underfitting happens when a model is **too simple** and cannot capture the underlying patterns in data. It performs poorly on both training and test data.

Example: Preparing too little for an exam, so you cannot answer even basic questions.

---

15. What is a confusion matrix used for?

A confusion matrix is a table used to evaluate classification models. It shows how many predictions were:

- Correct (True Positives & True Negatives)
- Incorrect (False Positives & False Negatives)

Example: In spam detection, it shows how many spam emails were correctly detected and how many normal emails were wrongly flagged as spam.

---

16. Why do we split data into training and testing sets?

We split data so the model can:

- **Learn** from training data.
- **Be evaluated** on test data to check if it works well on unseen data.

This prevents overfitting and ensures generalization.

---

17. What is a decision tree?

A decision tree is a model that makes decisions by splitting data into branches based on conditions. Each internal node represents a decision (feature), each branch represents an outcome, and each leaf node represents a final prediction.

Example:
Predicting if someone will play tennis:

- Is it sunny? Yes → Play.
- Is it rainy? No → Don't Play.

---

18. What is a linear regression model used for?

Linear regression is used to model the relationship between **input variables (features)** and a **continuous output (label)** by fitting a straight line.

Example: Predicting salary based on years of experience.

---

19. What does accuracy mean in ML?

Accuracy measures the percentage of correctly predicted values out of the total predictions.

Formula:

Accuracy=Correct Predictions Total Predictions×100\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{Total Predictions}} \times 100Accuracy=Total Predictions Correct Predictions ×100

Example: If 80 out of 100 predictions are correct, accuracy = 80%.

---

20. What are some common applications of Machine Learning in daily life?

- Spam detection in emails.
- Recommendation systems (Netflix, Amazon, YouTube).
- Fraud detection in banking.
- Voice assistants (Alexa, Siri).
- Face recognition (phone unlock).
- Self-driving cars.
- Healthcare (disease diagnosis from scans).