**1. What is Machine Learning?**

Machine Learning (ML) is a subset of Artificial Intelligence (AI) that enables computers to learn from data and make decisions or predictions without being explicitly programmed. It identifies patterns in data and improves performance over time with experience.

**2. Types of Machine Learning:**

* **Supervised Learning**The algorithm learns from labeled data (input-output pairs).  
  *Example: Predicting house prices, Email spam detection.*
* **Unsupervised Learning**The algorithm finds patterns or groups in unlabeled data.  
  *Example: Customer segmentation, Anomaly detection.*
* **Reinforcement Learning**An agent learns by interacting with an environment, receiving rewards or penalties.  
  *Example: Game-playing AI, Robotics.*

**3. Difference between AI and ML:**

|  |  |  |
| --- | --- | --- |
| **Aspect** | **Artificial Intelligence (AI)** | **Machine Learning (ML)** |
| **Scope** | Broader field of making machines smart | Subset of AI that focuses on learning from data |
| **Goal** | Mimic human intelligence | Learn from data to make predictions/decisions |
| **Example** | Self-driving cars | Predicting stock prices |

**4. What is the difference between supervised and unsupervised learning?**

|  |  |
| --- | --- |
| **Supervised Learning** | **Unsupervised Learning** |
| Trained on labeled data | Trained on unlabeled data |
| Output is known (e.g., a class or value) | Output is unknown or inferred |
| Tasks: Classification, Regression | Tasks: Clustering, Association |

**5. What is a model in ML?**

An ML model is a mathematical representation trained on data to recognize patterns and make predictions or decisions.

*Example: A linear regression model predicts house prices based on size and location.*

**6. What are features and labels?**

* Features are the input variables (independent variables).  
  *Example: Age, Salary, Education.*
* Label is the output or target variable (dependent variable).  
  *Example: Will the person buy a product (Yes/No)?*

**7. What is the difference between classification and regression?**

|  |  |
| --- | --- |
| **Classification** | **Regression** |
| Predicts categories or classes | Predicts continuous values |
| Example: Email is spam or not | Example: Predicting house price |

**8. What is a training set and a test set?**

* Training Set: Data used to train the model.
* Test Set: Data used to evaluate the model’s performance on unseen data.

**9. What is cross-validation? Why is it used?**

Cross-validation is a technique to assess a model’s performance by splitting data into multiple training and validation sets.  
The most common is k-fold cross-validation, where data is split into *k* parts.

Purpose:

* To ensure the model generalizes well to unseen data.
* To reduce overfitting and get a more reliable performance estimate.

**10. How do you evaluate an ML model’s performance?**

Depends on the problem type:

* For Classification:
  + Accuracy, Precision, Recall, F1-Score
  + Confusion Matrix
  + ROC-AUC score
* For Regression:
  + Mean Absolute Error (MAE)
  + Mean Squared Error (MSE) / RMSE
  + R² Score (Coefficient of Determination)

**✅ 2. Algorithms & Models**

**11. What is Linear Regression?**

Linear Regression is a supervised learning algorithm used for predicting continuous values.  
It models the relationship between input feature(s) and output by fitting a straight line (y = mx + c).

Example: Predicting house price based on size.

**12. What is Logistic Regression?**

Logistic Regression is a classification algorithm (despite the name "regression").  
It is used when the output is categorical (e.g., Yes/No, 0/1).  
It uses the sigmoid function to output probabilities between 0 and 1.

Example: Spam email detection.

**13. What is K-Nearest Neighbors (KNN)?**

K-Nearest Neighbors (KNN) is a supervised machine learning algorithm. It's used for both classification and regression tasks,

KNN is a lazy learning algorithm that classifies a data point based on how its neighbors are classified.

* It looks at the K closest data points in the training set.
* The majority label among the K is assigned to the new point.

Example: Classifying whether a person will buy a product based on age and salary.

**14. What is Decision Tree?**

A Decision Tree is a tree-structured model that splits the data based on conditions (like yes/no questions) at each node.  
Each internal node represents a decision, and each leaf node a final class or value.

Advantages: Easy to interpret, handles both classification and regression.

**15. What is Random Forest?**

Random Forest is an ensemble method that builds multiple decision trees and combines their results.

* For classification: majority vote
* For regression: average of predictions

It improves accuracy and reduces overfitting compared to a single decision tree.

**16. What is Naive Bayes?**

Naive Bayes is a probabilistic classifier based on Bayes’ Theorem, assuming features are independent (naive assumption).  
It’s fast and works well on text classification problems like spam detection.

**17. What is Support Vector Machine (SVM)?**

SVM is a supervised learning algorithm used for classification and regression.  
It finds the optimal hyperplane that best separates the data into classes, maximizing the margin between them.

Can also use kernel tricks to work on non-linear data.

**18. What is K-Means Clustering?**

K-Means is an unsupervised learning algorithm used to divide data into K clusters.

* It assigns each point to the nearest cluster centroid.
* Centroids are updated until they stabilize.

Example: Customer segmentation.

**19. Explain how a Neural Network works.**

A Neural Network mimics how the human brain works:

* Made up of layers: input layer → hidden layer(s) → output layer.
* Each layer has neurons (nodes) connected with weights.
* Neurons perform weighted summation + activation function (like ReLU, Sigmoid).

Training involves:

* Forward propagation to calculate prediction.
* Loss function to find error.
* Backpropagation to adjust weights using gradient descent.

**20. What is a Confusion Matrix? Explain TP, FP, TN, FN.**

A confusion matrix is a performance table used in classification problems. It compares the predicted labels vs actual labels.

|  | Predicted Positive | Predicted Negative |
| --- | --- | --- |
| Actual Positive | TP (True Positive) | FN (False Negative) |
| Actual Negative | FP (False Positive) | TN (True Negative) |

* TP: Correctly predicted positive cases
* TN: Correctly predicted negative cases
* FP: Incorrectly predicted positive (Type I error)
* FN: Incorrectly predicted negative (Type II error)

Used to calculate: Accuracy, Precision, Recall, F1 Score

**✅ 3. Performance Metrics**

1. **What is accuracy, precision, recall, and F1-score?**
2. **When would you use precision over recall?**
3. **What is ROC-AUC curve?**
4. **What is Mean Squared Error (MSE)?**
5. **What is R² (R-squared) value?**

**✅ 4. Model Optimization & Validation**

1. **What is overfitting and underfitting?**
2. **How can you avoid overfitting?**
3. **What is regularization (L1 and L2)?**
4. **What is bias-variance trade-off?**
5. **What is grid search and random search?**
6. **What is hyperparameter tuning?**
7. **What are learning curves?**

**✅ 5. Data Preprocessing**

1. **Why is data cleaning important in ML?**
2. **What is feature engineering?**
3. **What is normalization and standardization?**
4. **What are outliers and how do you handle them?**
5. **What is feature selection and dimensionality reduction?**
6. **What is Principal Component Analysis (PCA)?**

**✅ 6. Deep Learning (If mentioned in resume)**

1. **What is Deep Learning? How is it different from ML?**
2. **What is a perceptron?**
3. **What are activation functions (ReLU, Sigmoid, Tanh)?**
4. **What is a convolutional neural network (CNN)?**
5. **What is backpropagation?**
6. **What is dropout in neural networks?**

**✅ 7. Project-Based & Scenario Questions**

1. **Explain your Dream11 Prediction ML model.**
2. **What algorithm did you use for prediction and why?**
3. **How did you handle missing or noisy data in your project?**
4. **What metrics did you use to evaluate your model’s accuracy?**
5. **How did you deploy your ML model (Flask, API)?**
6. **What would you do if your model's accuracy was low?**