1. A bank wants to predict whether a loan applicant will default based on credit score, income, and past loan history. What type of ML problem is this, and what steps would you take to solve it?

   Type of ML Problem: It is a ***Supervised Learning*** problem, Specifically, a ***Classification*** problem. Because the goal is to ***predict Yes or No*** (whether a person will default on a loan).

   Steps to Solve:

   - **Collect Data**
     Get data like credit score, income, and loan history along with whether the person defaulted or not.
   - **Prepare the Data**
     - Fill or remove missing values
     - Convert text data to numbers (like past loan history)
     - Split data into training and testing sets
   - **Choose a Model**
     Use a classification model like:
     - Logistic Regression
     - Decision Tree
     - Random Forest
   - **Train the Model**
     Use the training data to let the model learn.
   - **Test the Model**
     Check how well the model works using the testing data.
   - **Make Predictions**
     Use the model to predict if new loan applicants will default.

---

2. A retail store wants to predict the demand for different products to optimize inventory levels. What type of ML problem is this, and what steps would you take to solve it?

   Type of ML Problem: It is a ***Supervised Learning*** problem. Specifically, a ***Regression*** problem. Because the goal is to ***predict a number*** (how many products will be sold).

   Steps to Solve:

   - **Collect Data**
     Sales history, Product info, Dates, holidays, promotions.
   - **Prepare the Data**
     - Clean missing values
     - Convert dates
     - Split into training and testing sets
   - **Choose a Model**
     - Linear Regression
     - Decision Tree
     - Random Forest

- **Train the Model**
  Teach the model using past data
- **Test the Model**
  Check how well it predicts on new data
- **Predict Demand**
  Use it to forecast future sales and manage inventory

---

3. A factory wants to detect whether a manufactured product is defective based on sensor readings and quality control data. What type of ML problem is this, and what steps would you take to solve it?

   Type of ML Problem: It is a ***Supervised Learning*** problem, Specifically, a ***Classification*** problem. Because the goal is to ***predict Yes or No*** (Is the product defective?)
   Steps to Solve:
   - **Collect Data** – Gather sensor measurements, inspection reports, and defect labels (defective or not).
   - **Preprocess Data** – Handle missing values, normalize sensor data, and encode categorical features.
   - **Split Dataset** – Divide the data into training and testing sets.
   - **Choose Algorithm** – Use Logistic Regression, Decision Trees, Random Forest, or SVM.
   - **Train Model** – Fit the model using labeled data showing defective and non-defective products.
   - **Evaluate Performance** – Use Accuracy, Precision, Recall, F1-score, and ROC-AUC to assess model performance.
   - **Make Predictions** – Predict if new products are defective based on sensor readings.

---

4. A healthcare provider wants to analyze patient symptoms and classify them into different disease categories. What type of ML problem is this, and what steps would you take to solve it?

   Type of ML Problem: It is a ***Classification*** problem. Because we want to ***predict categories*** (disease types).
   Steps to Solve:
   - **Collect Data:** Patient symptoms, Medical history, Disease diagnosis (labels).
   - **Prepare the Data:**
     - Clean missing or incorrect data
     - Convert symptoms to numerical format
     - Split into training and testing sets
   - **Choose a Model:** Logistic Regression, Decision Tree, Random Forest, Naive Bayes.
   - **Train the Model:** Teach the model using labeled patient data.

- **Test the Model:** Check how well the model classifies diseases.
- **Make Predictions:** Predict disease category for new patients based on symptoms.

---

5. An e-commerce company wants to identify and remove fake reviews posted by bots or fraudsters. What type of ML problem is this, and what steps would you take to solve it?

   Type of ML Problem: ***Classification*** Problem. We want to classify whether a review is real or fake.
   Steps to Solve:
   - **Collect Data**
     Get review text, ratings, user info (like time, IP, etc.).
   - **Clean the Data**
     Remove symbols, stopwords, duplicates, etc.
   - **Create Features**
     Length of review. How often the user posts. Time between reviews. Sentiment of the text. Same text used in many reviews?.
   - **Train a Model (Use models like):**
     Logistic Regression, Random Forest, SVM, Or simple NLP models
   - **Test the Model**
     Check how well it works using accuracy, precision, recall, etc.
   - **Use the Model**
     Use it to flag or remove fake reviews automatically.
   - **Keep Improving**
     Update with new data to catch smarter fraud.

---

6. A financial firm wants to predict stock price movements based on historical price data and market indicators. What type of ML problem is this, and what steps would you take to solve it?

   Type of ML Problem: ***Regression*** Problem. We want to ***predict a number*** — the future stock price. Or ***Classification*** Problem (if predicting up/down) Just predicting if ***price will go up or down*****.**
   Steps to Solve:
   - **Collect Data**
     Get historical stock prices, volume, indicators (like RSI, moving average, etc.).
   - **Clean the Data**
     Fill missing values, remove errors.
   - **Create Features**
     - Previous day's prices.
     - Technical indicators (moving average, MACD, RSI).
     - Market news sentiment (optional).

    - Time features (day, month).
- **Choose the Target**
    - If predicting price: Regression
    - If predicting up/down: Classification
- **Train a Model (Use models like):**
  Linear Regression, Decision Trees / Random Forest, XGBoost, LSTM (for time series data).
- **Test the Model (Check accuracy using metrics like):**
    - MAE, RMSE (for regression)
    - Accuracy, precision (for classification)
- **Predict Future Movements**
    - Use the model to predict tomorrow's price or trend.

---

7. A social media platform wants to detect fake user accounts based on user activity and profile data. What type of ML problem is this, and what steps would you take to solve it?

Type of ML Problem: ***Classification*** Problem. We want to classify accounts as ***real or fake.***

Steps to Solve:

- **Collect Data:** Profile age, Number of posts, Friends/followers, Activity pattern, Bio info, profile picture, etc.
- **Clean the Data:** Remove missing or wrong values.
- **Create Features:** How active the user is, How old the account is, Number of posts per day, Is the bio empty?, Same IP used for many accounts?.
- **Train a Model**
    - Decision Tree
    - Random Forest
    - Logistic Regression
    - Or others
- **Test the Model**
  Measure how good it is (accuracy, precision, etc.).
- **Use the Model**
  Detect and flag fake accounts.
- **Improve Over Time**
  Update the model with new data as fake behavior changes.

---

8. A marketing agency wants to segment customers into different groups based on their purchasing behavior. What type of ML problem is this, and what steps would you take to solve it?

Type of ML Problem: Clustering Problem (Unsupervised Learning). We want to group customers based on their behavior, without any labels.

Steps to Solve:

- **Collect Data**
  customer age, income, total spend, frequency of purchase, etc.
- **Clean the Data**
  Handle missing values, remove outliers, scale the data (normalize).
- **Choose Features**
  How much they spend. How often they buy. What types of products they buy.
- **Apply Clustering Algorithm**
  - K-Means (most common)
  - DBSCAN or Hierarchical Clustering (if data is complex)
- **Decide Number of Groups**
  Use methods like elbow method to find best number of clusters.
- **Analyze the Groups:** One group may be high spenders. Another may buy rarely but regularly.
- **Use the Groups:** Target each group with personalized marketing. Improve product recommendations.

9. A geospatial research team wants to analyze satellite images to classify different land types (forest, water, urban). What type of ML problem is this, and what steps would you take to solve it?

   Type of ML Problem: Image Classification Problem. We want to classify images into land types like forest, water, urban, etc.
   Steps to Solve It:
   - **Collect Data**
     Get labeled satellite images (with land type info).
   - **Preprocess the Images**
     - Resize images
     - Normalize pixel values (e.g., scale from 0 to 1)
     - Augment data (flip, rotate, zoom)
   - **Build a Model:**
     - CNN (Convolutional Neural Networks)
     - Pre-trained models like ResNet, VGG, MobileNet
   - **Train the Model**
     Feed images and labels into the model to learn patterns.
   - **Test the Model**
     Check accuracy and other metrics on new images.
   - **Use the Model**
     Classify new satellite images automatically.
   - **Update When Needed**
     Add new images to keep the model accurate over time.

---

10. A streaming service wants to predict which users are likely to cancel their subscriptions. What type of ML problem is this, and what steps would you take to solve it?

Type of ML Problem: Classification Problem. We want to predict whether a user will cancel (churn) or not.

Steps to Solve It:

- **Collect Data**
  - User activity (watch time, login frequency)
  - Subscription type
  - Payment history
  - Complaints or feedback
- **Clean the Data**
  - Fill missing values
  - Remove duplicates
  - Convert dates, categories, etc.
- **Create Features**
  - Days since last login
  - Total watch hours
  - Number of devices used
  - Payment failures
- **Label the Users**
  - "1" if user canceled (churned)
  - "0" if user stayed
- **Train a Model**
  - Logistic Regression
  - Random Forest
  - XGBoost
- **Test the Model**
  - Accuracy
  - Precision & Recall
  - F1-score
- **Use the Model**
  Predict which users are likely to cancel — and take action (like offering discounts or alerts).
