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?

Problem type – Classification

- **Data Collection** – Customer financial history, credit scores, and loan repayment.
- **Data P reprocessing** – Remove outliers, normalize transaction amounts, and encode categorical features.
- **Split Dataset** – Into training and testing sets.
- **Choose Algorithm** – Classification Model - Logistic Regression, Random Forest
- **Train the Model** – Fit the model on the training dataset.
- **Evaluate Performance** – Metrics accuracy, precision, recall, AUC-ROC and F1-score.
- **Make Predictions –** Use the model to predict loan default for new applicants.

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?

Problem type – Regression

- **Data Collection** – records with sales data, seasonal trends, and product demand.
- **Data P reprocessing** – Handle missing values, encode categorical variables.
- **Split Dataset** – Into training and testing sets.
- **Choose Algorithm** – Regression model - Linear Regression or Decision Tree Regression.
- **Train the Model** – Fit the model on the training dataset.
- **Evaluate Performance** – Metrics RMSE (Root Mean Square Error) and R^2 score.
- **Make Predictions –** Use the model to predict upcoming sales periods.

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?

Problem type – Classification

- **Data Collection** – Records sensor readings, production details, and defect labels.

- **Data P reprocessing** – Remove outliers, normalize transaction amounts, and encode categorical features.
- **Split Dataset** – Into training and testing sets.
- **Choose Algorithm** – Classification Model - Logistic Regression, Random Forest, SVM
- **Train the Model** – Fit the model on the training dataset.
- **Evaluate Performance** – Metrics accuracy, precision, recall, AUC-ROC and F1-score.
- **Make Predictions –** Use the model to predict defective products in real time.

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?

Problem type – Classification

- **Data Collection** – patient records with symptoms and diagnoses.
- **Data P reprocessing** – Remove outliers, normalize transaction amounts, and encode categorical features.
- **Split Dataset** – Into training and testing sets.
- **Choose Algorithm** – Classification Model - Logistic Regression, Random Forest, SVM
- **Train the Model** – Fit the model on the training dataset.
- **Evaluate Performance** – Metrics accuracy, precision, recall, AUC-ROC and F1-score.
- **Make Predictions –** Use the model to predict disease category based on patient 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?

Problem type – Classification

- **Data Collection** – records labeled as real and fake reviews.
- **Data P reprocessing** – Remove outliers, normalize transaction amounts, and Convert email text to numerical format .
- **Create features -** transaction frequency, average spending, and unusual behavior detection.
- **Split Dataset** – Into training and testing sets.

- **Choose Algorithm** – Naive Bayes, Support Vector Machines, or Neural Networks.
- **Train the Model** – Fit the model on the training dataset.
- **Evaluate Performance** – Metrics accuracy, precision, recall, AUC-ROC and F1-score.
- **Make Classification –** Use the model to classify fake reviews in real-time.

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?

Problem type – Regression

- **Data Collection** – historical data on stock prices, trading volumes, and economic indicators.
- **Data P reprocessing** – Handle missing values, encode categorical variables.
- **Split Dataset** – Into training and testing sets.
- **Choose Algorithm** – Regression model - Linear Regression or Decision Tree Regression.
- **Train the Model** – Fit the model on the training dataset.
- **Evaluate Performance** – Metrics RMSE (Root Mean Square Error) and R^2 score.
- **Make Predictions –** Use the model to predict future stock price movements.

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?

Problem type – Classification

- **Data Collection** – records labeled as account details, activity logs, and engagement patterns.
- **Data P reprocessing** – Remove outliers, normalize transaction amounts, and Convert email text to numerical format .
- **Create features -** transaction frequency, average spending, and unusual behavior detection.
- **Split Dataset** – Into training and testing sets.
- **Choose Algorithm** – Naive Bayes, Support Vector Machines, or Neural Networks etc.
- **Train the Model** – Fit the model on the training dataset.
- **Evaluate Performance** – Metrics accuracy, precision, recall, AUC-ROC score. and F1-score.
- **Make Classification –** Use the model to classify flag fake accounts.

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?

Problem type – Clustering

- **Data Collection** – Uuser click behavior, browsing history, and demographic information.
- **Data P reprocessing** – Normalize data like amount spent
- **Split Dataset** – Into training and testing sets.
- **Choose Algorithm** – K-Means, DBSCAN, or Hierarchical Clustering.
- **Train the Model** – Fit the model on the training dataset.
- **Evaluate Performance** – Apply clustering algorithm to identify user categories ("Action Lovers," "Drama Fans").
- **Make Predictions –** Use Clusters to Suggest "Tech Enthusiasts," "Fashion Lovers" based on preferences.

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?

Problem type – Classification

- **Data Collection** – records like satellite images labeled with land types.
- **Data P reprocessing** – Remove outliers, normalize transaction amounts, and Convert email text to numerical format .
- **Create features -** transaction frequency, average spending, and unusual behavior detection.
- **Split Dataset** – Into training and testing sets.
- **Choose Algorithm** – Naive Bayes, Support Vector Machines, or Neural Networks etc.
- **Train the Model** – Fit the model on the training dataset.
- **Evaluate Performance** – Metrics accuracy, precision, recall, AUC-ROC score. and F1-score.
- **Make Classification –** Use the model to classify new satellite images into land cover types.

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?

Problem type – Classification

- **Data Collection** – records like user engagement data, subscription history, and interaction logs.
- **Data P reprocessing** – Remove outliers, normalize transaction amounts, and Convert email text to numerical format .
- **Create features -** transaction frequency, average spending, and unusual behavior detection.
- **Split Dataset** – Into training and testing sets.
- **Choose Algorithm** – Naive Bayes, Support Vector Machines, or Neural Networks etc.
- **Train the Model** – Fit the model on the training dataset.
- **Evaluate Performance** – Metrics accuracy, precision, recall, AUC-ROC score. and F1-score.
- **Make Classification –** Use the model to classify customers likely to churn and apply retention strategies.