

ML Scenario Based Set 2 Questions

1. 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?

Logic:

1. **Collect Data:** Applicant info: credit score, income, past loan history; Label = Default (Yes/No).
 2. **Domain Selection:** Supervised Learning: **Classification**.
 3. **Preprocess Data:** Handle missing values, encode categorical variables, normalize numeric values.
 4. **Split Input/Output:** X: credit score, income, loan history; y: default (1/0).
 5. **Split Dataset:** Training (70%) and Testing (30%).
 6. **Choose Algorithm:** Logistic Regression, Decision Tree, Random Forest.
 7. **Train Model:** Fit model on training data.
 8. **Evaluate Model:** Use metrics: Accuracy, Precision, Recall, F1:score, ROC:AUC.
 9. **Make Predictions:** Predict default risk for new loan 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?

Logic:

1. **Collect Data:** Product sales history, stock levels, demand (units sold).
 2. **Domain Selection:** Supervised Learning: **Regression**.
 3. **Preprocess Data :** Handle missing values, aggregate sales per product, encode categories, scale numeric features.
 4. **Split Input/Output :** X: product features; y: demand (units sold).
 5. **Split Dataset :** Training (70%) and Testing (30%).
 6. **Choose Algorithm :** Linear Regression, Decision Tree Regressor
 7. **Train Model :** Fit model on training data.
 8. **Evaluate Model :** Metrics: RMSE, R^2 .
 9. **Make Predictions :** Predict demand for upcoming periods to optimize 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?

Logic:

1. **Collect Data :** Sensor readings, production line data, quality control labels (defective / not defective).
2. **Domain Selection :** Supervised Learning: **Classification**.

3. **Preprocess Data** : Clean data, handle missing values, normalize sensor readings, encode categorical features.
4. **Split Input/Output** : X: sensor & quality data; y: label (defective = 1 / not defective = 0).
5. **Split Dataset** : Training (70%) and Testing (30%).
6. **Choose Algorithm** : Logistic Regression, Decision Tree, Random Forest, SVM.
7. **Train Model** : Fit model using training data.
8. **Evaluate Model** : Metrics: Accuracy, Precision, Recall, F1:score, ROC:AUC.

4. **Make Predictions** : Detect defective products in real time.

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?

logic:

1. **Collect Data** : Patient symptoms, medical records, lab results, disease diagnosis (labels).
2. **Domain Selection** : Supervised Learning: **Classification**.
3. **Preprocess Data** : Handle missing values, encode categorical symptoms, normalize numeric values.
4. **Split Input/Output** : X: patient symptoms & features; y: disease category (label).
5. **Split Dataset** : Training (70%) and Testing (30%).
6. **Choose Algorithm** : Logistic Regression, Decision Tree, Random Forest, Naive Bayes
7. **Train Model** : Fit model on training dataset.
8. **Evaluate Model** : Metrics: Accuracy, Precision, Recall, F1:score, ROC:AUC
9. **Make Predictions** : Classify new patients into disease categories.

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?

Logic

1. **Collect Data** : Customer reviews, review text, ratings, user activity logs, reviewer profile details.
2. **Domain Selection** : Supervised Learning: **Classification**.
3. **Preprocess Data** : Clean text (remove stopwords, punctuation), convert text into numbers (TF:IDF / Word2Vec), handle missing values.
4. **Split Input/Output** : X: review features (text, metadata); y: label (fake = 1 / genuine = 0).
5. **Split Dataset** : Training (70%) and Testing (30%).
6. **Choose Algorithm** : Logistic Regression, Naive Bayes, Random Forest, XGBoost, Deep Learning (LSTM/BERT).
7. **Train Model** : Fit model on labeled training data.
8. **Evaluate Model** : Metrics: Accuracy, Precision, Recall, F1:score, ROC:AUC.
9. **Make Predictions** : Detect and remove fake reviews automatically.

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?

Logic:

1. **Collect Data** : Historical stock prices, trading volume, technical indicators (moving averages), Past stock price
 2. **Domain Selection** : Supervised Learning: **Regression**.
 3. **Preprocess Data** : Handle missing values, normalize/scale numeric data
 4. **Split Input/Output** : X: historical features & indicators; y: future stock price (continuous value).
 5. **Split Dataset** : Training (past data) and Testing (recent data, time:based split).
 6. **Choose Algorithm** : Linear Regression, Random Forest Regressor, XGBoost Regressor.
 7. **Train Model** : Fit model on training data.
 8. **Evaluate Model** : Metrics: RMSE, R^2 .
 9. **Make Predictions** : 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?

Logic:

1. **Collect Data** : Historical data: bio, User activity logs (Connections, posts, likes, comments), fake = 1 / genuine = 0
 2. **Domain Selection** : Supervised Learning: **Classification**.
 3. **Preprocess Data** : Handle missing values, encode categorical features, normalize numeric values, extract activity:based features.
 4. **Split Input/Output** : X: user activity & profile features; y: label (fake = 1 / genuine = 0).
 5. **Split Dataset** : Training (70%) and Testing (30%).
 6. **Choose Algorithm** : Logistic Regression, Decision Tree, Random Forest, SVM, XGBoost, Neural Networks.
 7. **Train Model** : Fit model using training dataset.
 8. **Evaluate Model** : Metrics: Accuracy, Precision, Recall, F1:score, ROC:AUC.
 9. **Make Predictions** : Detect and flag fake accounts automatically.
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?

Logic:

1. **Collect Data** : Customer purchase history, Salary, spending amount
2. **Domain Selection** : Unsupervised Learning: **Clustering**.
3. **Preprocess Data** : Handle missing values, normalize/spending scaling, encode categorical features if needed.
4. **Split Input**: Salary, spending amount

5. **Select Clustering Algorithm** : K:Means, Hierarchical, DBSCAN.
6. **Find Optimal Clusters** : Use Elbow Method or Silhouette Score.
7. **Train Model** : Apply chosen clustering algorithm to form groups.
8. **Analyze Clusters** : Identify groups (e.g., high spenders, budget shoppers, seasonal buyers).

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?

Logic:

1. **Collect Data** : historical data: Satellite images data with labeled regions (forest, water, urban).
2. **Domain Selection** : Supervised Learning: **Classification**
3. **Preprocess Data** : Find image data match with historical data in numeric.
4. **Split Input/Output** : X: satellite image data features, y: land type label (forest / water / urban).
5. **Split Dataset** : Training (70%) and Testing (30%).
6. **Choose Algorithm** : Logistic Regression, Decision Tree, Random Forest, SVM, XGBoost, Neural Networks.
7. **Train Model** : Fit model on training images.
8. **Evaluate Model** : Metrics: Accuracy, Precision, Recall, F1:score, Confusion Matrix.
9. **Make Predictions** : Classify land type for new satellite images.

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?

Logic:

1. **Collect Data** : User watch history, subscription details, login frequency, complaints, payment history.
2. **Domain Selection** : Supervised Learning: **Classification**.
3. **Preprocess Data** : Handle missing values, encode categorical variables, normalize numeric features.
4. **Split Input/Output** : X: user behavior & profile features; y: label (cancel = 1 / active = 0).
5. **Split Dataset** : Training (70%) and Testing (30%).
6. **Choose Algorithm** : Logistic Regression, Decision Tree, Random Forest, XGBoost.
7. **Train Model** : Fit model using training dataset.
8. **Evaluate Model** : Metrics: Accuracy, Precision, Recall, F1:score, ROC:AUC.
9. **Make Predictions** : Predict which users are likely to cancel and take retention actions