ML Set 3 Question

1. A power company wants to predict the electricity demand in different regions based on past usage, temperature, and time of year. What type of machine learning approach would be suitable? Outline the steps to build the model.

   Logic:
   1. **Collect Data**: Past electricity usage, temperature records, time of year (season, month, day, hour), regional information.
   2. **Domain Selection**: Supervised Learning: **Regression**.
   3. **Preprocess Data**: Handle missing values, normalize/scale numeric data, encode categorical features (e.g., region, season).
   4. **Split Input/Output**: X: features (usage history, weather, time); y: electricity demand (continuous value).
   5. **Split Dataset**: Training (70%) and Testing (30%) (time-based split if sequential).
   6. **Choose Algorithm**: Linear Regression, Random Forest Regressor, XGBoost Regressor, Neural Networks.
   7. **Train Model**: Fit the model on training data.
   8. **Evaluate Model**: Metrics: RMSE, MAE, R^2.
   9. **Make Predictions**: Forecast electricity demand for future periods in each region.

2. An e-commerce platform wants to identify and filter out fake product reviews based on user behavior, review patterns, and sentiment analysis. What type of machine learning approach would be suitable? Outline the steps to build the model.

   Logic:
   1. **Collect Data**: Review text, ratings, user behavior (frequency of reviews, account age), review patterns, sentiment scores.
   2. **Domain Selection**: Supervised Learning: **Classification**.
   3. **Preprocess Data** –
      a. Clean review text (remove stopwords, punctuation).
      b. Convert text into numbers (TF-IDF, Word2Vec, BERT embeddings).
      c. Handle missing values, normalize numeric features.
   4. **Split Input/Output**: X: review text + behavior features; 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 training dataset.
   8. **Evaluate Model**: Metrics: Accuracy, Precision, Recall, F1-score, ROC-AUC.
   9. **Make Predictions**: Identify and filter out fake reviews automatically

3. A city traffic department wants to estimate traffic congestion levels based on GPS data, road construction reports, and weather conditions. What type of machine learning approach would be suitable? Outline the steps to build the model.

   Logic:
   1. **Collect Data**: GPS data (vehicle speeds, locations), road construction reports, weather conditions, historical congestion levels.
   2. **Domain Selection**: Supervised Learning: **Regression** (predict numeric congestion level).
   3. **Preprocess Data**: Handle missing values, normalize numeric features, encode categorical features (weather, road type), aggregate GPS data into meaningful features (average speed per road segment).

4. **Split Input/Output**: X: GPS + weather + road features; y: congestion level (numeric).
5. **Split Dataset**: Training (70%) and Testing (30%) (time-based split if sequential).
6. **Choose Algorithm**: Linear Regression, Random Forest Regressor, XGBoost Regressor, Neural Networks.
7. **Train Model**: Fit model on training dataset.
8. **Evaluate Model**: Metrics: RMSE, MAE, R^2.
9. **Make Predictions**: Estimate numeric congestion levels to optimize traffic management in real time.

4. A bank wants to determine whether a loan applicant should be approved or rejected based on income, credit history, and previous loan repayment behavior. What type of machine learning approach would be suitable? Outline the steps to build the model.

Logic:

1. **Collect Data**: Applicant income, credit history, past loan repayment records, other financial information.
2. **Domain Selection**: Supervised Learning: **Classification** (Approve = 1 / Reject = 0).
3. **Preprocess Data**: Handle missing values, encode categorical variables, normalize numeric features.
4. **Split Input/Output**: X: applicant financial features; y: label (approved/rejected).
5. **Split Dataset**: Training (70%) and Testing (30%).
6. **Choose Algorithm**: Logistic Regression, Decision Tree, Random Forest, XGBoost, SVM.
7. **Train Model**: Fit model on training dataset.
8. **Evaluate Model**: Metrics: Accuracy, Precision, Recall, F1-score, ROC-AUC.
9. **Make Predictions**: Predict approval or rejection for new applicants

5. A factory wants to automatically detect defective products using images from a quality control camera. What type of machine learning approach would be suitable? Outline the steps to build the model.

Logic:

1. **Collect Data**: Images of defective and non-defective products.
2. **Domain Selection**: Supervised Learning: **Classification**.
3. **Preprocess Data**: Resize, normalize, and augment images.
4. **Train Model**: Use CNN or pre-trained model on training images.
5. **Evaluate & Predict**: Check accuracy and detect defective products automatically.

6. An agricultural company wants to predict the best crops to grow based on soil composition, rainfall, and past harvest data. What type of machine learning approach would be suitable? Outline the steps to build the model.

Logic:

1. **Collect Data**: Soil composition, rainfall, past harvest data.
2. **Domain Selection**: Supervised Learning: **Classification** (best crop) or **Regression** (expected yield).
3. **Preprocess Data**: Handle missing values, encode categories, scale numbers.
4. **Split Input/Output**: X: soil & weather features; y: crop type or yield.
5. **Split Dataset**: Training (70%) and Testing (30%).
6. **Choose Algorithm**: Decision Tree, Random Forest, XGBoost.
7. **Train Model**: Fit model on training data.
8. **Evaluate Model**: Accuracy (classification) or RMSE/MAE (regression).
9. **Make Predictions**: Recommend best crop or expected yield for a field.

7. A business wants to analyze customer responses to different types of advertisements and promotions to determine the most effective marketing strategy. What type of machine learning approach would be suitable? Outline the steps to build the model.

Logic:

1. **Collect Data**: Customer responses to ads and promotions, demographics, purchase history, engagement metrics.
2. **Domain Selection**: Supervised Learning: **Classification** (response: positive/negative) or **Regression** (response score, purchase amount).
3. **Preprocess Data**: Handle missing values, encode categorical variables, normalize numeric features.
4. **Split Input/Output**: X: ad type + customer features; y: response (binary or numeric).
5. **Split Dataset**: Training (70%) and Testing (30%).
6. **Choose Algorithm**: Classification: Logistic Regression, Decision Tree, Random Forest, XGBoost; Regression: Linear Regression, Random Forest Regressor.
7. **Train Model**: Fit model on training data.
8. **Evaluate Model**: Classification → Accuracy, Precision, Recall, F1-score; Regression → RMSE, MAE, R^2.
9. **Make Predictions**: Identify which ads and promotions are most effective for each customer segment.

8. A software company wants to build a system that can analyze source code and predict whether a particular piece of code is likely to contain a bug. What type of machine learning approach would be suitable? Outline the steps to build the model.

Logic:

1. **Collect Data**: Source code files, code metrics (lines of code, complexity), past bug reports, commit history.
2. **Domain Selection**: Supervised Learning: **Classification** (buggy = 1 / clean = 0).
3. **Preprocess Data**: Clean code text, extract features (e.g., code metrics, TF-IDF or embeddings for code), handle missing values.
4. **Split Input/Output**: X: code features; y: label (buggy or not).
5. **Split Dataset**: Training (70%) and Testing (30%).
6. **Choose Algorithm**: Logistic Regression, Decision Tree, Random Forest, XGBoost, or Deep Learning models for code embeddings.
7. **Train Model**: Fit model on training dataset.
8. **Evaluate Model**: Metrics: Accuracy, Precision, Recall, F1-score, ROC-AUC.
9. **Make Predictions**: Predict whether new code is likely to contain bugs

9. A fitness app wants to recommend personalized workout plans for users based on their exercise history, fitness level, and preferences. What type of machine learning approach would be suitable? Outline the steps to build the model.

Logic:

1. **Collect Data**: User exercise history, fitness level, preferences, goals.
2. **Domain Selection**: Unsupervised Learning: **Clustering** or Recommender System.
3. **Preprocess Data**: Handle missing values, encode categories, scale numbers.
4. **Select Algorithm**: K-Means, Hierarchical Clustering, or Collaborative Filtering.
5. **Train Model**: Group similar users or learn preferences.
6. **Analyze & Apply**: Recommend personalized workout plans for each user.

10. A social media platform wants to detect fake accounts by analyzing user activity, posting patterns, and interactions with other users.Give steps to Achieve.

Logic:

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