ML-Scenario based

🏠 1. Predicting House Prices

Scenario: A real estate company wants to estimate house prices based on square footage, number of bedrooms, and location.
Problem Type: Regression
Step-by-Step Logic:

1. **Collect Data** – Gather historical records including square footage, bedroom count, and location.

Preprocess Data –

   - Handle missing values
   - Encode categorical variables (location → One-Hot)
   - Detect outliers using IQR or Z-score
   - Feature scaling (optional depending on model)

2. **Split Dataset** – Divide the data into training and testing sets.
3. **Select Algorithm** – Choose regression models such as Linear Regression or Decision Tree Regression.
4. **Train the Model** – Fit the model using the training data.
5. **Evaluate Performance** – Use RMSE and R^2 score to assess accuracy.
6. **Make Predictions** – Predict house prices for new listings.

💳 2. Identifying Fraudulent Transactions

Scenario: A bank aims to detect fraudulent transactions using customer behavior and transaction history.
Problem Type: Classification
Step-by-Step Logic:

1. **Collect Data** – Compile labeled transaction records (fraudulent vs. non-fraudulent).
2. **Preprocess Data** – Remove outliers, normalize transaction amounts, and encode categorical features.
3. **Feature Engineering** – Create features like transaction frequency, average spending, and anomaly indicators.

4. **Split Dataset** – Separate data into training and testing sets.
5. **Select Algorithm** – Use classifiers like Logistic Regression, Random Forest, or Neural Networks.
6. **Train the Model** – Fit the model on labeled data.
7. **Evaluate Performance** – Use accuracy, precision, recall, and F1-score.
8. **Deploy Model** – Enable real-time fraud detection.

🛒 3. Grouping Customers Based on Spending Habits

Scenario: A supermarket wants to segment customers by shopping behavior.
Problem Type: Clustering
Step-by-Step Logic:

1. **Collect Data** – Gather purchase history, spending amounts, and frequency.
2. **Preprocess Data** – Normalize features to reduce bias.
3. **Select Clustering Algorithm** – Use K-Means, DBSCAN, or Hierarchical Clustering.
4. **Determine Optimal Clusters** – Apply the Elbow Method to find the ideal number of clusters.
5. **Train Model** – Cluster customers based on spending patterns.
6. **Analyze Clusters** – Identify segments like high, medium, and low spenders.
7. **Apply Insights** – Tailor marketing strategies for each segment.

👩💼 4. Predicting Employee Salaries

Scenario: A company wants to estimate salaries based on experience, job title, and education.
Problem Type: Regression
Step-by-Step Logic:

1. **Collect Data** – Compile employee records with relevant features and salary.
2. **Preprocess Data** – Handle missing values and encode categorical variables.
3. **Split Dataset** – Divide data into training and testing sets.
4. **Select Algorithm** – Use Linear Regression or Random Forest Regression.
5. **Train the Model** – Fit the model on training data.

6. **Evaluate Model** – Use MAE and R^2 score to measure performance.
7. **Make Predictions** – Estimate salaries for new employees.

📧 5. Detecting Spam Emails

Scenario: An email provider wants to classify emails as spam or not spam.
Problem Type: Classification
Step-by-Step Logic:

1. **Collect Data** – Use labeled datasets of spam and non-spam emails.
2. **Preprocess Data** – Convert text to numerical format using TF-IDF or embeddings.
3. **Split Dataset** – Separate into training and testing sets.
4. **Select Algorithm** – Use Naive Bayes, SVM, or Neural Networks.
5. **Train the Model** – Fit the model on labeled email data.
6. **Evaluate Model** – Use precision, recall, and F1-score.
7. **Deploy Model** – Automatically classify incoming emails.

💬 6. Customer Reviews Sentiment Analysis

Scenario: A company wants to determine sentiment (positive/negative) from customer reviews.
Problem Type: Classification
Step-by-Step Logic:

1. **Collect Data** – Gather labeled reviews.
2. **Preprocess Text** – Remove stopwords, punctuation, and tokenize.
3. **Convert Text to Features** – Use TF-IDF or Word2Vec.
4. **Split Dataset** – Perform train-test split.
5. **Select Algorithm** – Use Logistic Regression, Naive Bayes, or BERT.
6. **Train Model** – Fit the model on training data.
7. **Evaluate Model** – Use accuracy and F1-score.
8. **Make Predictions** – Classify new reviews.

🚗 7. Predicting Car Insurance Claims

Scenario: An insurer wants to predict if a customer will file a claim next year.
Problem Type: Classification
Step-by-Step Logic:

1. **Collect Data** – Include claim history, driving behavior, and demographics.
2. **Preprocess Data** – Handle missing values and encode categorical features.
3. **Split Dataset** – Divide into training and testing sets.
4. **Select Algorithm** – Use Logistic Regression, Decision Tree, or Neural Networks.
5. **Train the Model** – Fit the model on historical data.
6. **Evaluate Model** – Use AUC-ROC and precision-recall metrics.
7. **Deploy Model** – Predict claim likelihood for new customers.

🎬 8. Recommending Movies Based on Viewing History

Scenario: A streaming platform wants to group users and recommend content.
Problem Type: Clustering
Step-by-Step Logic:

1. **Collect Data** – Include genres watched, ratings, and preferences.
2. **Preprocess Data** – Encode categorical genres numerically.
3. **Select Clustering Algorithm** – Use K-Means or Hierarchical Clustering.
4. **Determine Optimal Clusters** – Apply the Elbow Method.
5. **Train Model** – Group users based on preferences.
6. **Analyze Clusters** – Label segments (e.g., "Action Fans").
7. **Recommend Content** – Suggest movies tailored to each group.

🏥 9. Predicting Patient Recovery Time

Scenario: A hospital wants to estimate recovery duration post-surgery.
Problem Type: Regression
Step-by-Step Logic:

1. **Collect Data** – Include age, medical history, and lifestyle.
2. **Preprocess Data** – Normalize features and handle missing values.
3. **Select Algorithm** – Use Random Forest or Linear Regression.

4. **Train Model** – Fit the model on training data.
5. **Evaluate Model** – Use RMSE to assess accuracy.
6. **Make Predictions** – Estimate recovery time for new patients.

🎓 10. Predicting Student Exam Scores

Scenario: A university wants to predict exam scores using study habits and attendance.
Problem Type: Regression
Step-by-Step Logic:

1. **Collect Data** – Include study hours, attendance, and past scores.
2. **Preprocess Data** – Handle missing values and standardize features.
3. **Split Dataset** – Divide into training and testing sets.
4. **Select Algorithm** – Use Linear Regression or Support Vector Regression.
5. **Train Model** – Fit the model on training data.
6. **Evaluate Performance** – Use RMSE and R^2 score.
7. **Make Predictions** – Estimate scores for new students.
