1. A real estate company wants to develop a system that predicts house prices based on square footage, number of bedrooms, and location.  
    **Q:** Identify the problem type and outline the step-by-step logic to solve it.

**Problem Statement:** House Price Prediction

**Domain:** Machine Learning

**Prediction/Output:** Regression (numerical data)

**Step-by-step logic:**

* **Collect Data** – Gather a dataset with features like square footage, number of bedrooms, and location.
* **Preprocess Data** – Perform the One Hot Encoding technique used to convert categorical data into numerical data(e.g., location).
* **Split Dataset** – Divide the dataset into training and testing sets.
* **Choose Algorithm** – Use a regression model like Multiple Linear Regression, Support Vector Machine, Random Forest, or Decision Tree Regression.
* **Train the Model** – Fit the model on the training dataset.
* **Evaluate Performance** – Use metrics like R² score.

**Make Predictions** – Use the model to predict house prices for new data.

1. A bank wants to build a model to detect fraudulent transactions by analyzing customer spending behavior and transaction history.  
    **Q:** Identify the problem type and outline the step-by-step logic to solve it.

**Problem Statement:** Predicting Fraud Transactions

**Domain:** Machine Learning

**Prediction/Output:** Classification (categorical data)

**Step-by-step logic:**

* **Collect Data** – Collect a dataset with features like Customer details and transaction history
* **Preprocess Data** – Perform the One Hot Encoding technique, which is used to convert categorical data into numerical data and bring the data to a specific range using standardization.
* **Split Dataset** – Divide the dataset into training and testing sets.
* **Choose Algorithm** – Use a classification algorithm like KNN, Naive Bayes, Support Vector Machine, Random Forest, or Decision Tree Regression.
* **Train the Model** – Fit the model on the training dataset.
* **Evaluate Performance** – Use metrics like the classification report and find the accuracy, precision, recall, and F1 score.

**Make Predictions** – Use the model to predict whether a transaction is fraudulent or not

1. A supermarket wants to segment its customers based on their shopping patterns to provide personalized promotions.  
    **Q:** Identify the problem type and outline the step-by-step logic to solve it.

**Problem Statement:** Customer Segmentation Based on Shopping Patterns

**Domain:** Machine Learning

**Prediction/Output:** Clustering

**Step-by-step logic:**

* **Collect Data** – Collect a dataset with features like Customer details, purchase amount, frequency of purchase, and month of purchase.
* **Preprocess Data** – Perform the standardization preprocessing technique, which helps to bring the data to a specific range.
* **Choose Algorithm** – Use an algorithm like K Means, Affinity Propagation, DBSCAN, OPTICS, Hierarchical Clustering, Mean Shift, or BIRCH, etc
* **Determine Optimal Clusters –** Use the Elbow Method or Silhouette Method to find the best number of clusters.
* **Train the Model** – Apply a clustering algorithm to group customers.
* **Analyze Clusters** – Interpret results to identify high-spending, medium-spending, and low-spending customer groups.

**Use Clusters for Marketing** – Target each segment with personalized promotions.

1. A company wants to estimate an employee’s salary based on their years of experience, job title, and education level.  
    **Q:** Identify the problem type and outline the step-by-step logic to solve it.

**Problem Statement:** Employee Salary Prediction

**Domain:** Machine Learning

**Prediction/Output:** Regression (numerical data)

**Step-by-step logic:**

**Collect Data** – Gather a dataset with features like the number of experience, designation, education, etc

**Preprocess Data** – Perform the One Hot Encoding technique used to convert categorical data into numerical data(e.g., designation, education).

**Split Dataset** – Divide the dataset into training and testing sets.

**Choose Algorithm** – Use a regression model like Multiple Linear Regression, Support Vector Machine, Random Forest, or Decision Tree Regression.

**Train the Model** – Fit the model on the training dataset.

**Evaluate Performance** – Use metrics like R² score.

**Make Predictions** – Use the model to predict salary for a new employee.

1. An email provider wants to automatically classify incoming emails as spam or not spam based on their content and sender details.  
    **Q:** Identify the problem type and outline the step-by-step logic to solve it.

**Problem Statement:** Email Spam Detection

**Domain:** Machine Learning

**Prediction/Output:** Classification (categorical data)

**Step-by-step logic:**

**Collect Data** – Collect a dataset with features like Email.

**Preprocess Data** – Perform TF - IDF (For ML) or Word Embedding (NLP Domain)(Deep Learning)

**Split Dataset** – Divide the dataset into training and testing sets.

**Choose Algorithm** – Use a classification algorithm like KNN, Naive Bayes, Support Vector Machine, Random Forest, or Decision Tree Regression.

**Train the Model** – Fit the model on the training dataset.

**Evaluate Performance** – Use metrics like the classification report and find the accuracy, precision, recall, and F1 score.

**Make Predictions** – Use the model to detect whether an email is spam or not spam

1. A business wants to analyze customer reviews of its products and determine whether the sentiment is positive or negative.  
    **Q:** Identify the problem type and outline the step-by-step logic to solve it.

**Problem Statement:** Sentiment Analysis of Customer Reviews

**Domain:** Machine Learning

**Prediction/Output:** Classification (categorical data)

**Step-by-step logic:**

**Collect Data** – Collect features like customer reviews with both positive and negative.

**Preprocess Data** – Perform TF - IDF (For ML) or Word Embedding (NLP Domain)(Deep Learning)

**Split Dataset** – Divide the dataset into training and testing sets.

**Choose Algorithm** – Use a classification algorithm like KNN, Naive Bayes, Support Vector Machine, Random Forest, or Decision Tree Regression.

**Train the Model** – Fit the model on the training dataset.

**Evaluate Performance** – Use metrics like the classification report and find the accuracy, precision, recall, and F1 score.

**Make Predictions** – Use the model to analyse whether a review is positive or negative

1. An insurance company wants to predict whether a customer is likely to file a claim in the next year based on their driving history and demographics.  
    **Q:** Identify the problem type and outline the step-by-step logic to solve it.

**Problem Statement:** Predicting Car Insurance Claims

**Domain:** Machine Learning

**Prediction/Output:** Classification (categorical data)

**Step-by-step logic:**

**Collect Data** – Collect past claim history, driving behavior, and customer demographics.

**Preprocess Data** – Handle missing values and encode categorical features.

**Split Dataset** – Divide data into training and testing sets.

**Choose Algorithm** – Use Logistic Regression, Decision Tree, SVM, Random Forest, KNN, and Naive Bayes.

**Train the Model** – Fit the model using past claims data.

**Evaluate Model** – Use metrics like the classification report and find the accuracy, precision, recall, and F1 score.

**Make Predictions** – Use the model to analyse whether a customer claims insurance in the next year

1. A streaming platform wants to recommend movies to users by grouping them based on their viewing preferences and watch history.  
    **Q:** Identify the problem type and outline the step-by-step logic to solve it.

**Problem Statement:** Recommending Movies Based on Watch History

**Domain:** Machine Learning

**Prediction/Output:** Clustering

**Step-by-step logic:**

**Collect Data** – Collect a dataset with features like recent watch history, interests, etc

**Preprocess Data** – One Hot Encoding to convert history into numerical data

**Choose Algorithm** – Use an algorithm like K Means, Affinity Propagation, DBSCAN, OPTICS, Hierarchical Clustering, Mean Shift, or BIRCH, etc

**Determine Optimal Clusters –** Use the Elbow Method or Silhouette Method to find the best number of clusters.

**Train the Model** – Apply a clustering algorithm to group customers.

**Analyze Clusters** – Interpret results to identify thriller movies, comedy movies, and commercial movie groups.

**Use Clusters for Recommending Movies** – Target each segment with specific movies.

1. A hospital wants to predict the recovery time of patients after surgery based on their age, medical history, and lifestyle habits.  
    **Q:** Identify the problem type and outline the step-by-step logic to solve it.

**Problem Statement:** Predict Patients' Recovery Time

**Domain:** Machine Learning

**Prediction/Output:** Regression (numerical data)

**Step-by-step logic:**

**Collect Data** – Gather a dataset with features like patients' age, medical history, lifestyle habits, etc

**Preprocess Data** – Perform the One Hot Encoding technique used to convert categorical data into numerical data(e.g., medical history, lifestyle habits) and do Standardization

**Split Dataset** – Divide the dataset into training and testing sets.

**Choose Algorithm** – Use a regression model like Multiple Linear Regression, Support Vector Machine, Random Forest, or Decision Tree Regression.

**Train the Model** – Fit the model on the training dataset.

**Evaluate Performance** – Use metrics like R² score.

**Make Predictions** – Use the model to predict recovery time for the new patient.

1. A university wants to predict a student’s final exam score based on study hours, attendance, and past academic performance.  
    **Q:** Identify the problem type and outline the step-by-step logic to solve it.

**Problem Statement:** Student Exam Score Prediction

**Domain:** Machine Learning

**Prediction/Output:** Regression (numerical data)

**Step-by-step logic:**

**Collect Data** – Gather a dataset with features like the study hours, attendance, past academic score, etc

**Preprocess Data** – Perform the One Hot Encoding technique used to convert categorical data into numerical data, and Standardization

**Split Dataset** – Divide the dataset into training and testing sets.

**Choose Algorithm** – Use a regression model like Multiple Linear Regression, Support Vector Machine, Random Forest, or Decision Tree Regression.

**Train the Model** – Fit the model on the training dataset.

**Evaluate Performance** – Use metrics like R² score.

**Make Predictions** – Use the model to predict the exam score for the students.