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 type – Regression

- **Data Collection** – historical data with features like square footage, number of bedrooms, and location.

- **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 R² score.

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

2. 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 type – Classification

- **Data Collection** – transaction records labeled as fraudulent or non-fraudulent.

- **Data P reprocessing** – Remove outliers, normalize transaction amounts, and encode categorical features.

- **Create features -** transaction frequency, average spending, and unusual behavior detection.

- **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 real-time fraud detection.

3. 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 type – Clustering

- **Data Collection** – Customer purchase history, amount spent, and frequency of purchases.

- **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 group customers and Analyze Clusters

- **Make Predictions –** Use Clusters for Marketing

4. 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 type – Regression

- **Data Collection** – employee records with years of experience, education, and salary.

- **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 salary based on new employee data.

5. 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 type – Classification

- **Data Collection** – records labeled as spam and non-spam emails.

- **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 incoming emails as spam or not spam.

6. 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 type – Classification

- **Data Collection** – records labeled as customer reviews (positive/negative).

- **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 score. and F1-score.

- **Make Classification –** Use the model to classify new customer reviews as positive or negative.

7. 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 type – Classification

- **Data Collection** – records labeled as past claim history, driving behavior, and customer demographics.

- **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 score. and F1-score.

- **Make Classification –** Use the model to classify new customer ikely to file a claim in the next year.

8. 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 type – Clustering

- **Data Collection** – User movie preferences, genres watched, and ratings.

- **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 movies based on preferences.

9. 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 type – Regression

- **Data Collection** – historical data ecovery data with features like patient age, medical history, and lifestyle habits.

- **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 R² score.

- **Make Predictions –** Use the model to predict recovery time for new patients based on medical records.

10. 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 type – Regression

- **Data Collection** – historical data student records with study hours, attendance, and exam scores.

- **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 exam scores for new students based on input features.