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.

**Ramishahope Artificial Intelligence Pvt Ltd**

**36, Old Anandas, SG Arcade, Marudhamalai Main Road, Vadavalli, Coimbatore -641041.**

**+91 6385383227 | [www.hopelearning.net](http://www.hopelearning.net/) | [mdaravind@hopelearning.net](mailto:mdaravind@hopelearning.net) | 33AAMCR3722R1ZU**

**Problem type -**  **Regression**

**Step-by-step logic:**

**Step 1 : Collect Data** – Collect data with features like square footage, number of bedrooms, and location.

**Step 2 :Preprocess Data** – Handle missing values, encode categorical variables

**Step 3: Split Dataset** – Split training and testing sets.

**Step 4: Choose Algorithm** – Use a regression model like Linear Regression or Decision Tree Regression.

**Step 5: Train the Model** – Fit the model on the training dataset.

**Step 6: Evaluate Performance** – Use metrics like RMSE (Root Mean Square Error) and 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 type -**  **Classification**

**Step 1 : Collect Data** – Collect data about spending behaviour and transaction history

**Step 2 :Preprocess Data** – Handle missing values, encode categorical variables

**Step 3: Split Dataset** – Split training and testing sets.

**Step 4: Choose Algorithm** – Use a classification model like Logistic Regression or Decision Tree Regression.

**Step 5: Train the Model** – Fit the model on the training dataset.

**Step 6: Evaluate Performance** – Use cofusion matrix to evaluate the performance.

**Make Predictions** – Use the model to detect fraudulent transactionsfor new data.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

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 type -**  **Clustering**

**Step 1 : Collect Data** – Collect customers purchase history, frequent of purchases

**Step 2 :Preprocess Data** – Apply normalization techniques

**Step 3: Choose Algorithm** – Use clustering algorithms like DBSCAN, K-means or Hierarchical clustering.

**Step 5: Train the Model** – Fit the model to group the customers.

**Step 6: Analyze Clusters** – Analyse clustering results to identify high-spending, medium-spending, and low-spending customer groups.

**Step 7: 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 type -**  **Regression**

**Step 1 : Collect Data** – Collect employees data like years of experience, job title, and education level.

**Step 2 :Preprocess Data** – Handling missing values, Apply stndard scalar techniques

**Step 3: Split Dataset** – Separate data into training and testing sets

**Step 4: Choose Algorithm** – Use Regreesion algorithms like random forest, decision tree

**Step 5: Train the Model** – Fit the model on training data..

**Step 6: Evaluate Model** – Use Mean Absolute Error (MAE) and R² score for accuracy measurement.

**Make Predictions** – Predict salary based on new employee data.  
\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

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 type:** **Classification**

**Step 1: Collect Data** – Use datasets of spam and non-spam emails.

**Step 2: Preprocess Data** – Convert email text to numerical format using TF-IDF or word embeddings.

**Step 3: Split Dataset** – Divide data into training and testing sets.

**Step 4: Choose Algorithm** – Use classification algorithms like Naive Bayes, Support Vector Machines, or Neural Networks.

**Ramishahope Artificial Intelligence Pvt Ltd**

**36, Old Anandas, SG Arcade, Marudhamalai Main Road, Vadavalli, Coimbatore -641041.**

**+91 6385383227 | [www.hopelearning.net](http://www.hopelearning.net/) | [mdaravind@hopelearning.net](mailto:mdaravind@hopelearning.net) | 33AAMCR3722R1ZU**

**Step 5: Train the Model** – Fit the model using labeled email data.

**Step 6: Evaluate Model** – Measure accuracy using Precision, Recall, and F1-score.

**Step 7: Deploy Model** – Automatically classify incoming emails as 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 type:** **Classification**

**Step 1: Collect Data** – Gather labeled customer reviews (positive/negative).

**Step 2: Preprocess Text Data** – Remove stopwords, punctuation, and tokenize words.

**Step 3: Convert Text into Features** – Use TF-IDF or Word2Vec to convert text into numerical format.

**Step 4: Split Dataset** – Train-test split.

**Step 5: Choose Algorithm** – Use Logistic Regression, Naive Bayes

**Step 6: Train Model** – Fit the model on the training dataset.

**Step 7: Evaluate Model** – Measure accuracy using Precision, Recall, and F1-score

**Deploy Model** – Classify new customer reviews as 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 type:** **Classification**

**Step 1: Collect Data** – Gather past claim history, driving behavior, and customer demographics.

**Step 2: Preprocess Data** – Handle missing values and encode categorical features.

**Step 3:Split Dataset** – Divide data into training and testing sets.

**Step 4: Choose Algorithm** – Use Logistic Regression, Decision Tree, or Neural Networks.

**Step 5: Train the Model** – Fit the model using training data.

**Step 6: Evaluate Model** – Use Precision-Recall, f1 score.

**Deploy Model** – Predict claims likelihood for new customers.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

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 type:** **Clustering**

**Step 1: Collect Data** – Gather user movie preferences, genres watched, and ratings.

**Step 2: Preprocess Data** – Convert categorical movie genres into numerical format.

**Step 3: Choose Clustering Algorithm** – Use K-Means or Hierarchical Clustering.

**Step 4: Train Model** – Apply clustering algorithm to group users.

**Step 5: Analyze Clusters** – Identify user categories like Action Lovers,Drama Fans.

**Recommend Content** – Suggest movies based on cluster preferences

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

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 type:** **Regression**

**Step 1: Collect Data** – Collect data with features like patient age, medical history, and lifestyle habits.

**Step 2: Preprocess Data** – Use normaliztion technique and handle missing values.

**Step 3: Choose Regression Algorithm** – Use Random Forest Regression or Linear Regression.

**Step 4: Train Model** – Fit the model on training data.

**Step 5: Evaluate Model** – Use R2 or RMSEmetrics to check accuracy.

**Make Predictions** – Predict recovery time for new patients based on medical records.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

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 type:** **Regression**

**Step 1: Collect Data** – Collect data with features like students study hours, attendance and past academic performance.

**Step 2: Preprocess Data** – Handle missing values and use standardization techniques.

**Step 3: Choose Regression Algorithm** – Use Support vector machine or Linear Regression.

**Step 4: Train Model** – Fit the model on training data.

**Step 5: Evaluate Model** – Use R² or RMSEmetrics to check accuracy.

**Make Predictions** – Predict students final exam score based on input features.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Ramishahope Artificial Intelligence Pvt Ltd**

**36, Old Anandas, SG Arcade, Marudhamalai Main Road, Vadavalli, Coimbatore -641041.**

**+91 6385383227 | [www.hopelearning.net](http://www.hopelearning.net/) | [mdaravind@hopelearning.net](mailto:mdaravind@hopelearning.net) | 33AAMCR3722R1ZU**