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.

ANS:

**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**

**a. problem type:** **Regression**

**b. Step-by-step logic:**

* **Collect Data** – Gather historical data with features like square footage, number of bedrooms, and location.
* **Preprocess Data** – Handle missing values, encode categorical variables.
* **Split Dataset** – Divide the dataset into training and testing sets.
* **Choose Algorithm** – Use a regression model like Linear Regression or Decision Tree Regression.
* **Train the Model** – Fit the model on the training dataset.
* **Evaluate Performance** – Use metrics like Root Mean Square Error & 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.

**a. problem type:** **Classification**

**b. Step-by-step logic:**

* **Collect Data** – Gather transaction records labeled as fraudulent or non-fraudulent.
* **Preprocess Data** – Remove outliers, normalize transaction amounts, and encode categorical features.
* **Feature Engineering** – Create features like transaction frequency, average spending, and unusual behavior detection.

**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**

* **Split Dataset** – Divide data into training and testing sets.
* **Choose Algorithm** – Use classification models like Logistic Regression, Random Forest, or Neural Networks.
* **Train the Model** – Fit the model using labeled transaction data.
* **Evaluate Performance** – Use metrics like accuracy, precision, recall, and F1-score.
* **Deploy Model** – Implement real-time fraud detection.

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.

**a. problem type:** **Clustering**

**b. Step-by-step logic:**

* **Collect Data** – Gather customer purchase history, amount spent, and frequency of purchases.
* **Preprocess Data** – Normalize data (e.g., scale spending amounts to avoid bias).
* **Choose Clustering Algorithm** – Use K-Means, DBSCAN, or Hierarchical Clustering.
* **Determine Optimal Clusters** – Use the Elbow Method to find the best number of clusters.
* **Train Model** – Apply 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.

**a. problem type:** **Regression**

**b. Step-by-step logic:**

* **Collect Data** – Gather employee records with years of experience, education, and salary.
* **Preprocess Data** – Handle missing values and encode categorical variables (e.g., job title).
* **Split Dataset** – Separate data into training and testing sets.
* **Choose Algorithm** – Use Linear Regression or Random Forest Regression.
* **Train the Model** – Fit the model on training data.
* **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.

**a. problem type:** **Classification**

**b. Step-by-step logic:**

* **Collect Data** – Use datasets of spam and non-spam emails.
* **Preprocess Data** – Convert email text to numerical format using TF-IDF or word embeddings.
* **Split Dataset** – Divide data into training and testing sets.
* **Choose Algorithm** – Use 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**

* **Train the Model** – Fit the model using labeled email data.
* **Evaluate Model** – Measure accuracy using Precision, Recall, and F1-score.
* **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.

**Identify the problem type:** **Classification**

**b. Step-by-step logic:**

* **Collect Data** – Gather labeled customer reviews (positive/negative).
* **Preprocess Text Data** – Remove stopwords, punctuation, and tokenize words.
* **Convert Text into Features** – Use TF-IDF or Word2Vec to convert text into numerical format.
* **Split Dataset** – Train-test split.
* **Choose Algorithm** – Use Logistic Regression, Naive Bayes, or Transformers (BERT).
* **Train Model** – Fit the model on the training dataset.
* **Evaluate Model** – Use accuracy and F1-score to assess model performance.
* **Make Predictions** – 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.

**a.problem type:** **Classification**

**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**

**b. Step-by-step logic:**

* **Collect Data** – Gather 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, or Neural Networks.
* **Train the Model** – Fit the model using past claims data.
* **Evaluate Model** – Use Precision-Recall, AUC-ROC 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.

**a. problem type:** **Clustering**

**b. Step-by-step logic:**

* **Collect Data** – Gather user movie preferences, genres watched, and ratings.
* **Preprocess Data** – Convert categorical movie genres into numerical format.
* **Choose Clustering Algorithm** – Use K-Means or Hierarchical Clustering.
* **Determine Optimal Clusters** – Use the Elbow Method.
* **Train Model** – Apply clustering algorithm to group users.
* **Analyze Clusters** – Identify user categories (e.g., "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.

**a. problem type:** **Regression**

**b. Step-by-step logic:**

* **Collect Data** – Gather historical recovery data with features like patient age, medical history, and lifestyle habits.
* **Preprocess Data** – Normalize medical features and handle missing values.
* **Choose Regression Algorithm** – Use Random Forest Regression or Linear Regression.
* **Train Model** – Fit the model on training data.
* **Evaluate Model** – Use RMSE (Root Mean Square Error) 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.

**a. problem type:** **Regression**

**b. Step-by-step logic:**

1. **Collect Data** – Gather historical student records with study hours, attendance, and exam scores.
2. **Preprocess Data** – Handle missing values and standardize numerical features.
3. **Split Dataset** – Divide data into training and testing sets.
4. **Choose Algorithm** – Use Linear Regression or Support Vector Regression.
5. **Train the Model** – Fit the model on training data.

**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**

1. **Evaluate Performance** – Use metrics like RMSE and R² score.
2. **Make Predictions** – Estimate exam scores for new students based on input features.