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**: Machine Learning 🡪 Supervised Learning 🡪 Regression

**Step-by-step logic**:

1. **Data collection**: Collect historical data containing features like square footage, number of bedrooms, location and house prices.
2. **Data Preprocessing**: Encode the categorical data to numerical data (eg: location) and handle any missing values and outliers.
3. **Input/Output split**: Categorize the input and output variables.
4. **Training & test set split**: Split the dataset into training and test sets.
5. **Model creation**: Using training test, create models using various regression models such as linear regression, support vector machine, decision tree, random forest.
6. **Evaluation**: Using test set, evaluate the model performance using R-score metrics.
7. **Make prediction**: Predict the house prices for the given inputs using the best saved model.
8. 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**: Machine Learning 🡪 Supervised Learning 🡪 Classification

**Step-by-step logic**:

1. **Data collection**: Collect transaction details labelled as fraudulent and non-fraudulent with customer details like their transaction frequency, average spending amount.
2. **Data Preprocessing**: Handle any outliers and categorical data encoding and normalize the data.
3. **Input/Output split**: Split the dataset into input and output variables.
4. **Training & test set split**: Divide the dataset as train and test sets.
5. **Model creation**: Using various classification models like Logistic regression, Navies Bayes, KNN, Support Vector Machine, Decision Tree, Random Forest, create model.
6. **Evaluation**: Evaluate the model performances using confusion matrix, F1- score, accuracy and classification report.
7. **Make prediction**: For the model with best accuracy, classify the transactions as fraud and not fraud for real time transactions.
8. 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**: Machine Learning 🡪 Unsupervised Learning (Clustering)

**Step-by-step logic**:

1. **Data collection**: Customers data like amount spent on shopping, period of the month at which shopping is done (like start, mid or end of the month), frequency of shopping, product purchased are collected.
2. **Data Preprocessing**: Data is preprocessed for outliers, missing values and encoding categorical data and normalization to avoid bias in spending.
3. **Model creation**: Use K-Means algorithm, Hierarchical algorithms, DBSCAN, to create models. Use the Elbow method to find the optimal number of clusters.
4. **Analyze clusters**: Interpret the results to identify high spending, medium spending and low spending customer groups.
5. **Use clusters for marketing**: Target each segment with personalized promotions.
6. 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**: Machine Learning 🡪 Supervised Learning 🡪 Regression

**Step-by-step logic**:

1. **Data collection**: Employee’s details like years of experience, job title, education level and salary are collected.
2. **Data Preprocessing**: The collected data are preprocessed for any missing values, outliers and categorical data.
3. **Input/Output split**: Input and output variables are separated.
4. **Training & test set split**: Split the training set and test set for both input and output variables.
5. **Model creation**: Create models using various regression models such as linear regression, support vector machine, decision tree, random forest.
6. **Evaluation**: Evaluate the model’s performances using R2 value.
7. **Make prediction**: Using the best model, find the employee’s salary for the given new input.
8. 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**: NLP 🡪 Supervised Learning 🡪 Classification

**Step-by-step logic**:

1. **Data collection**: Gather spam and non-spam emails dataset.
2. **Data Preprocessing**: Remove stopwords, punctuation and tokenize words.
3. **Convert text to number**: Use TF-IDF or Word2Vector to convert categorical data into numerical format.
4. **Training & test set split**: Split data into training and test sets.
5. **Model creation**: Use classification algorithms like Logistic regression, Navies Bayes, KNN to create models.
6. **Evaluation**: Use precision, accuracy, F1-score to evaluate the models.
7. **Make prediction**: Automatically classify mails as spam and non- spam.
8. 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**: NLP 🡪 Supervised Learning 🡪 Classification

**Step-by-step logic**:

1. **Data collection**: Gather labelled (positive / negative) customer reviews.
2. **Data Preprocessing**: Remove stopwords, punctuation and tokenize words.
3. **Convert text to number**: Use TF-IDF or Word2Vector to convert categorical data into numerical format.
4. **Training & test set split**: Split data into training and test sets.
5. **Model creation**: Use classification algorithms like Logistic regression, Navies Bayes, KNN to create models.
6. **Evaluation**: Use precision, accuracy, F1-score to evaluate the models.
7. **Make prediction**: For new customer reviews, make use of the best model to categorize into positive or negative.
8. 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**: Machine Learning 🡪 Supervised Learning 🡪 Classification

**Step-by-step logic**:

1. **Data collection**: Customers’ driving behavior, past claim history and demographics data are collected.
2. **Data Preprocessing**: Handle any outliers and categorical data encoding.
3. **Input/Output split**: Split the dataset into input and output variables.
4. **Training & test set split**: Divide the dataset as train and test sets.
5. **Model creation**: Using various classification algorithms like logistic regression, Navies Bayes, KNN, Support Vector Machine, Decision Tree, Random Forest, create model.
6. **Evaluation**: Evaluate the model performances using confusion matrix, F1- score, accuracy and classification report.
7. **Make prediction**: Predict the claims likelihood for new customers.
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**: Machine Learning 🡪 Unsupervised Learning (Clustering)

**Step-by-step logic**:

1. **Data collection**: Gather customer data like their movie preferences, watch history and ratings.
2. **Data Preprocessing**: Data is preprocessed for outliers, missing values and encoding categorical data.
3. **Model creation**: Using K-Means algorithm, Hierarchical algorithms, DBSCAN, models are created. Use the Elbow method to estimate the optimal number of clusters.
4. **Analyze clusters**: Group people based on clustering algorithms and recommend movies.
5. 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**: Machine Learning 🡪 Supervised Learning 🡪 Regression

**Step-by-step logic**:

1. **Data collection**: Recovery data along with patient’s age, medical history, and lifestyle habits are collected.
2. **Data Preprocessing**: The collected data are preprocessed for any missing values, outliers and categorical data.
3. **Input/Output split**: Input and output variables are categorized.
4. **Training & test set split**: Split the data into training set and test set.
5. **Model creation**: Create models using various regression models such as linear regression, support vector machine, decision tree, random forest.
6. **Evaluation**: Evaluate the model’s performances using R2 value.
7. **Make prediction**: Using the best model, find the recovery time based on the given medical reports.
8. 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**: Machine Learning 🡪 Supervised Learning 🡪 Regression

**Step-by-step logic**:

1. **Data collection**: Gather students’ data like their study hours, attendance, past academic performance and scores.
2. **Data Preprocessing**: Preprocess the data for any missing values, outliers and categorical data and normalize the numerical values.
3. **Input/Output split**: Input and output variables are categorized.
4. **Training & test set split**: Split the dataset into training set and test set.
5. **Model creation**: Create models using various regression models such as linear regression, support vector machine, decision tree, random forest.
6. **Evaluation**: Evaluate the model’s performances using R2 value.
7. **Make prediction**: Estimate the students’ marks based on the given input.