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:**

As we need to predict the house prices based on square footage, number of bedrooms and location it falls under ***Regression*** problem.

**Steps to solve the problem:**

* **Data Collection**: Past house sales details including square footage, number of bedrooms, locations and their respective house price should be collected.
* **Data Pre-processing**: Missing values should be handled. If any standardization required that should be handled using StandardScaler.If there are outliers, it should be replaced with mean, median or mode.
* **Split the dataset**: split the dataset into test and train data
* **Algorithm Selection:** Mulitple Linear regression, SVM regression, Decision Tree or Random forest tree
* **Train the model:** Fit the model using past input and output data(square footage, number of bedrooms, locations and house price)
* **Model Performance**: Evaluate the model performance using metrics like r\_score or rmse
* **Select Best Model**: Based on the metrics select the best model to deploy them
* **Make Predictions**: Predict house price by passing new house details with squared footage, number of bedrooms and location

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:**

As we need to predict the customer is fraud or not based on the customer spending behavior and transaction history, it is a ***Classification*** problem.

**Steps to solve the problem:**

* **Data Collection**: Past customer details like amount spent on different purposes, amount available on account and respective fraudulent status of them should be collected
* **Data Pre-processing**: Missing values should be handled. If any standardization required that should be handled using StandardScaler. Categorical values should be converted to numerical values using encoding technique.
* **Split the dataset**: split the dataset into test and train data
* **Algorithm Selection:** DecisiontreeClassifier, RandomForest, logistic regression or XGBoost Classifier
* **Train the model:** Fit the model using labelled fraudulent data
* **Model Performance**: Evaluate the model performance using metrics like AUC\_ROC, F1\_score,Recall and precision
* **Select Best Model**: Based on the metrics select the best model to deploy them
* **Make Predictions**: Predict the customer is fraud or not by passing customer spending behavior and transaction history

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:**

As we need to segment customers based on their shopping patterns, it is a ***Clustering type***.

**Steps to solve the problem:**

* **Data Collection**: Past customer purchase data like, total purchased amount and frequency of purchase should be collected
* **Data Pre-processing**: Missing values should be handled. If any standardization required that should be handled using StandardScaler.
* **Algorithm Selection:** Kmeans, Meanshift or HDBSCAN
* **Train the model:** Fit the model using total purchased amount and frequency of purchase
* **Model Performance**: Evaluate the model performance using metrics like Silhouette score
* **Select Best Model**: Based on the metrics select the best model to deploy them
* **Make Predictions**: Segment the customer based on the future purchasing data of customer. Check whether the clustering or segmentation of customer is predicted correctly based on the customer purchase data
* **Promotions sent**: Based on the customer shopping patterns(frequency of purchase) personalized promotion details are sent to customers

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:**

As we need to predict the employee’s salary based on their years of experience, job title and education level, it falls under ***Regression*** problem.

**Steps to solve the problem:**

* **Data Collection**: Past employee data including years of experience, job title and education level should be collected.
* **Data Pre-processing**: Missing values should be handled. If any standardization required that should be handled using StandardScaler. If there are outliers, it should be replaced with mean, median or mode.
* **Split the dataset**: Split the dataset into test and train data
* **Algorithm Selection:** Mulitple Linear regression, SVM regression, Decision Tree or Random forest tree
* **Train the model:** Fit the model using past input and output data(years of experience, job title, education level and salary)
* **Model Performance**: Evaluate the model performance using metrics like r\_score or rmse
* **Select Best Model**: Based on the metrics select the best model to deploy them
* **Make Predictions**: Predict employee salary by passing employee years of experience, job title and education level

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:**

As we need to classify whether the incoming email is spam or not based on their content and sender details, this falls under a ***Classification*** Type

**Steps to solve the problem:**

* **Data Collection**: Past Sender details, content of the mail and respective spam status(Yes or no) should be collected
* **Data Pre-processing**: Convert email text to numerical format using word embeddings
* **Algorithm Selection:** Support vector Classifier, Naïve bayes or Neural Networks
* **Train the model:** Fit the model using labelled spam data
* **Model Performance**: Evaluate the model performance using metrics like AUC\_ROC, F1\_score,Recall and precision
* **Select Best Model**: Based on the metrics select the best model to deploy them
* **Make Predictions**: Predict the incoming mail is spam or not by passing sender details and mail content

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:**

As we need to classify whether the sentiment is positive or negative based on the customer reviews on the products, this falls under a ***Classification*** Type

**Steps to solve the problem:**

* **Data Collection**: Past customer reviews and their respective sentiment status i.e. positive or negative should be collected
* **Data Pre-processing(Text)**: Remove stopwords, punctuation and tokenwise words
* **Algorithm Selection:** SupportVector Classifier, Naïve Bayes or Logistic Regression
* **Train the model:** Fit the model using labelled sentiment data
* **Model Performance**: Evaluate the model performance using metrics like AUC\_ROC, F1\_score,Recall and precision
* **Select Best Model**: Based on the metrics select the best model to deploy them
* **Make Predictions**: Predict the customer review is positive or not for a product based on the review content

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:**

As we need to classify whether the customer is likely to claim insurance or not based on their driving history and demographics, this falls under a ***Classification*** *Type*

**Steps to solve the problem:**

* **Data Collection**: Customer Insurance History details like driving history i.e (driving speed, driving experience) and demographics with their customer insurance claim status(Y/N) should be collected
* **Data Pre-processing**: Missing values should be handled. If any standardization required that should be handled using StandardScaler. Categorical values should be converted to numerical values using encoding technique.
* **Split the dataset**: split the dataset into test and train data
* **Algorithm Selection:** DecisiontreeClassifier, RandomForest, logistic regression or XGBoost Classifier
* **Train the model:** Fit the model using labelled Insurance Claim data
* **Model Performance**: Evaluate the model performance using metrics like AUC\_ROC, F1\_score,Recall and precision
* **Select Best Model**: Based on the metrics select the best model to deploy them
* **Make Predictions**: Predict the customer is likely to claim insurance or not by passing Driving details and demographics

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:**

As we need to recommmend movies to users based on their viewing preferences and watch history, it is a ***Clustering type***.

**Steps to solve the problem:**

* **Data Collection**: Customer viewing preferences(genre) and watch history like movies they watched previously should be collected
* **Data Pre-processing**: Missing values should be handled. If any standardization required that should be handled using StandardScaler.
* **Algorithm Selection:** Kmeans, Meanshift, HDBSCAN or Spectral Clustering
* **Train the model:** Fit the model using movie preference and watch history data
* **Model Performance**: Evaluate the model performance using metrics like Silhouette score
* **Select Best Model**: Based on the metrics select the best model to deploy them
* **Make Predictions**: Segment the customers by passing movie preference and watch history. Check if the clustering is done properly as per the customer movie preference and watch history
* **Recommendations sent**: Based on the customer view history and preferences, new movies are recommended to them to watch them next.

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:**

As we need to predict the patient recovery time after surgery based on their age, medical history and lifestyle habits, it falls under ***Regression*** problem.

**Steps to solve the problem:**

* **Data Collection**: Patient history like age, medical history, lifestyle habits and recovery time after surgery should be collected
* **Data Pre-processing**: Missing values should be handled. If any standardization required that should be handled using StandardScaler. If there are outliers, it should be replaced with mean, median or mode.
* **Split the dataset**: split the dataset into test and train data
* **Algorithm Selection:** Mulitple Linear regression, SVM regression, Decision Tree or Random forest tree
* **Train the model:** Fit the model using past input and output data(age, medical history, lifestyle habits and recovery time after surgery)
* **Model Performance**: Evaluate the model performance using metrics like r\_score or rmse
* **Select Best Model**: Based on the metrics select the best model to deploy them
* **Make Predictions**: Predict patient recovery time after surgery by passing new patient details like age, medical history and lifestyle habits

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:**

As we need to predict the student exam scores based on study hours, attendance and past academic performance, it falls under ***Regression*** problem.

**Steps to solve the problem:**

* **Data Collection**: Student history data like study hours, attendance and academic performance(i.e exam scores) should be collected
* **Data Pre-processing**: Missing values should be handled. If any standardization required that should be handled using StandardScaler. If there are outliers, it should be replaced with mean, median or mode.
* **Split the dataset**: split the dataset into test and train data
* **Algorithm Selection:** Mulitple Linear regression, SVM regression, Decision Tree or Random forest tree
* **Train the model:** Fit the model using past input and output data(study hours, attendance, exam scores)
* **Model Performance**: Evaluate the model performance using metrics like r\_score or rmse
* **Select Best Model**: Based on the metrics select the best model to deploy them
* **Make Predictions**: Predict exam scores by passing study hours and attendance