1. A real estate company wants to develop a system that predicts house prices based on square footage, number of bedrooms, and location.

**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**

**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 – Collect the data for house price prediction. Store the data either in data frame, variable.
* Pre-process Data – check the data for missing values and update. Perform encoding on non-numerical data. One hot encoding, dataset.dummies
* Split Dataset – training set and test set. Xtrain,ytrain., xtest,ytest
  + Input – square footage, number of bedrooms, and location.
  + Output - house prices
* Create and Choose Algorithm – MLR , SVM Regression
* Train the Model – Execute the algorithm with the parameters with the training data.
* Test the Model – Pass the test data to the model and get the results
* Evaluate Performance – Use metrics like RMSE (Root Mean Square Error) and R² score. Validate the various regression algorithms and check the score. Choose the best algorithm.
* Save to the file
* Deploy the file
* Make Predictions – For the best algorithm with all the best parameters, provide the new input(get the user input) and get the desired output. Predict the house price.

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**

* 1. Collect Data – Collect the data for fraud data transactions, customer data base. Store the data either in data frame, variable.
  2. Pre-process Data – check the data for missing values and update. Perform encoding on non-numerical data. One hot encoding, dataset.dummies
  3. Split Dataset – training set and test set.
  4. Input – customer spending behaviour and transaction history.
     1. Output – Fraud Transaction (Yes/No)
  5. Create and Choose Algorithm – Logistic regression , Random forest, Decision Tree (classfier)
  6. Train the Model – Execute the algorithm with the parameters with the training data.
  7. Test the Model – Pass the test data to the model and get the results
  8. Evaluate Performance – Confusion Matrix. Validate the various classification using confusion matrix and check the accuracy, precision recall, f1 score Choose the best algorithm.
  9. Save to the file
  10. Deploy the file
  11. Make Predictions – For the best algorithm with all the best parameters, provide the new input(get the user input) and get the desired output. Predict the transaction is fraud or not

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**

* 1. Collect Data – Collect the data , super market customer database. Store the data either in data frame, variable.
  2. Pre-process Data – check the data for missing values and update. Perform encoding on non-numerical data. One hot encoding, dataset.dummies
  3. Customer data base and details with the shopping history, amount spend
  4. Create and Choose Algorithm – K-means , Affinity Propagation , DBSCAN, BIRCH , OPTICS
  5. Plot the clusters, seaborn
  6. Evaluate Performance – get the score with silhouette\_score
  7. Choose the best algorithm. The score should be near to 1
  8. Save to the file
  9. Deploy the file
  10. Make Predictions – For the best algorithm with all the best parameters, provide the new input(get the user input) and get the desired output. Predict the customer and send the personalised 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**

👉 **b.** Step-by-step logic:

1. Collect Data – Collect the data for employee salary prediction. Store the data either in data frame, variable.
2. Pre-process Data – check the data for missing values and update. Perform encoding on non-numerical data. One hot encoding, dataset.dummies
3. Split Dataset – training set and test set. Xtrain,ytrain., xtest,ytest
   1. Input and output
4. Create and Choose Algorithm – SVM Regression , Decision Tree (Regression), Random Forest (Regression)
5. Train the Model – Execute the algorithm with the parameters with the training data.
6. Test the Model – Pass the test data to the model and get the results
7. Evaluate Performance – Use metrics like RMSE (Root Mean Square Error) and R² score. Validate the various regression algorithms and check the score. Choose the best algorithm.
8. Save to the file
9. Deploy the file
10. Make Predictions – For the best algorithm with all the best parameters, provide the new input(get the user input) and get the desired output. Predict the employee salary.

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**

* 1. Collect Data – Collect the data for email details,userr data base. Store the data either in data frame, variable.
  2. Pre-process Data – check the data for missing values and update. Perform encoding on non-numerical data. One hot encoding, dataset.dummies
  3. Ordinal data or Numerical data
  4. Split Dataset – training set and test set.
  5. Input – user email content and the sender details.
     1. Output –Spam email (yes /no)
  6. Create and Choose Algorithm – Logistic regression , Random forest, Decision Tree (classfier)
  7. Train the Model – Execute the algorithm with the parameters with the training data.
  8. Test the Model – Pass the test data to the model and get the results
  9. Evaluate Performance – Confusion Matrix. Validate the various classification using confusion matrix and check the accuracy, precision recall, f1 score Choose the best algorithm.
  10. Save to the file
  11. Deploy the file
  12. Make Predictions – For the best algorithm with all the best parameters, provide the new input(get the user input) and get the desired output. Predict the Spam email (yes /no)

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**

* 1. Collect Data – Collect the customer database, reviews, sentiment details Store the data either in data frame, variable.
  2. Pre-process Data – check the data for missing values and update. Perform encoding on non-numerical data. One hot encoding, dataset.dummies
  3. Ordinal data or Numerical data
  4. Split Dataset – training set and test set.
  5. Input – user email content and the sender details.
     1. Output –sentiment (positive /negative)
  6. Create and Choose Algorithm – Logistic regression , Random forest, Decision Tree (classfier)
  7. Train the Model – Execute the algorithm with the parameters with the training data.
  8. Test the Model – Pass the test data to the model and get the results
  9. Evaluate Performance – Confusion Matrix. Validate the various classification using confusion matrix and check the accuracy, precision recall, f1 score Choose the best algorithm.
  10. Save to the file
  11. Deploy the file
  12. Make Predictions – For the best algorithm with all the best parameters, provide the new input(get the user input) and get the desired output. Predict the feedback sentiment analysis positive / 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**

* 1. Collect Data – Collect the customer database, insurance policy , claims details, health details Store the data either in data frame, variable.
  2. Pre-process Data – check the data for missing values and update. Perform encoding on non-numerical data. One hot encoding, dataset.dummies
  3. Ordinal data or Numerical data
  4. Split Dataset – training set and test set.
     1. Input ,Output
  5. Create and Choose Algorithm – Logistic regression , Random forest, Decision Tree (classfier)
  6. Train the Model – Execute the algorithm with the parameters with the training data.
  7. Test the Model – Pass the test data to the model and get the results
  8. Evaluate Performance – Confusion Matrix. Validate the various classification using confusion matrix and check the accuracy, precision recall, f1 score Choose the best algorithm.
  9. Save to the file
  10. Deploy the file
  11. Make Predictions – For the best algorithm with all the best parameters, provide the new input(get the user input) and get the desired output. Predict the customer will file a claim yes /no

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**

* 1. Collect Data – Collect the data , movie customer database. Store the data either in data frame, variable.
  2. Pre-process Data – check the data for missing values and update. Perform encoding on non-numerical data. One hot encoding, dataset.dummies
  3. Customer data base and details with the shopping history, amount spend
  4. Create and Choose Algorithm – K-means , Affinity Propagation , DBSCAN, BIRCH , OPTICS
  5. Plot the clusters, seaborn
  6. Evaluate Performance – get the score with silhouette\_score
  7. Choose the best algorithm. The score should be near to 1
  8. Save to the file
  9. Deploy the file
  10. Make Predictions – For the best algorithm with all the best parameters, provide the new input(get the user input) and get the desired output. Predict and recommend movies to viewers

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**

👉 **b.** Step-by-step logic:

* 1. Collect Data – Collect the data for patience database. Store the data either in data frame, variable.
  2. Pre-process Data – check the data for missing values and update. Perform encoding on non-numerical data. One hot encoding, dataset.dummies
  3. Split Dataset – training set and test set. Xtrain,ytrain., xtest,ytest
     1. Input – square footage, number of bedrooms, and location.
     2. Output - house prices
  4. Create and Choose Algorithm – SVM Regression , Decision Tree (Regression), Random Forest (Regression)
  5. Train the Model – Execute the algorithm with the parameters with the training data.
  6. Test the Model – Pass the test data to the model and get the results
  7. Evaluate Performance – Use metrics like RMSE (Root Mean Square Error) and R² score. Validate the various regression algorithms and check the score. Choose the best algorithm.
  8. Save to the file
  9. Deploy the file
  10. Make Predictions – For the best algorithm with all the best parameters, provide the new input(get the user input) and get the desired output. Predict the recovery time of patience after surgery

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.
   1. Collect Data – Collect the data for students database and predict the exam score. Store the data either in data frame, variable.
   2. Pre-process Data – check the data for missing values and update. Perform encoding on non-numerical data. One hot encoding, dataset.dummies
   3. Split Dataset – training set and test set. Xtrain,ytrain., xtest,ytest
      1. Input(study hours, attendance, and past academic performance.  
         ) and output(exam score)
   4. Create and Choose Algorithm – SVM Regression , Decision Tree (Regression), Random Forest (Regression)
   5. Train the Model – Execute the algorithm with the parameters with the training data.
   6. Test the Model – Pass the test data to the model and get the results
   7. Evaluate Performance – Use metrics like RMSE (Root Mean Square Error) and R² score. Validate the various regression algorithms and check the score. Choose the best algorithm.
   8. Save to the file
   9. Deploy the file
   10. Make Predictions – For the best algorithm with all the best parameters, provide the new input(get the user input) and get the desired output. Predict the student final exam score.