

### **Scenario\_SET1\_ Answer**

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

**Problem type:** Numerical values in the label so it is **Regression**.

**Step 1 :** Data Cleaning – identifying and replacing nulls.

**Step 2:** Data Preprocessing – Categorical values into numerical by one hot encoding.

**Step 3: Input Split** – Square footage, no.of bedrooms and location.

**Output Split** – House Prices

**Step 4:** Splitting Train and Test Set

**Step 5:** Using algorithms like linear Regression, Decision Tree or SVM

**Step 6:** Finding the RMSE, R\_score value and saving the best model.

**Step 7:** Predict the house prices based on our requirements from the created model.

2. A bank wants to build a model to detect fraudulent transactions by analyzing customer spending behavior and transaction history.

**Problem type:** Fraud/Not- Fraud classification in the label so it is **Classification**.

**Step 1 :** Data Cleaning – identifying and replacing nulls.

**Step 2:** Data Preprocessing – Categorical values into numerical by one hot encoding or label encoding.

**Step 3:** Input Split and Output Split

**Step 4:** Splitting Train and Test Set

**Step 5:** Using algorithms like Decision Tree, Logistic Regression or SVM.

**Step 6:** Determine classification matrix to find precision, recall, F1 measure.

**Step 7:** Predict the fraud transactions from the created model.

3. A supermarket wants to segment its customers based on their shopping patterns to provide personalized promotions.

**Problem type:** Segmenting customers in the label so it is **Clustering**.

**Step 1:** Data collection from customer history and shopping patterns.

**Step 2:** Data Preprocessing – Removing / dropping unnecessary data.

**Step 3:** Input Split and Output Split

**Step 4:** Using algorithms like K-means, DBScan, OPTICS.

**Step 5:** Determine clusters and outliers.

**Step 6:** Plot the graph and identify the number of clusters.

**Step 7:** Predict the customer segmentation from the created model.

4. A company wants to estimate an employee's salary based on their years of experience, job title, and education level.

**Problem type:** Numerical values in the label so it is **Regression**.

**Step 1 :** Data Cleaning – identifying and replacing nulls.

**Step 2:** Data Preprocessing – Categorical values into numerical by one hot encoding.

**Step 3: Input Split** – years of experience, job title, and education level.

**Output Split** – salary

**Step 4:** Splitting Train and Test Set

**Step 5:** Using algorithms like linear Regression, Decision Tree or SVM

**Step 6:** Finding the RMSE, R\_score value and saving the best model.

**Step 7:** Predict the salary based on our requirements from the created model.

5. An email provider wants to automatically classify incoming emails as spam or not spam based on their content and sender details.

**Problem type:** Spam/Not-spam in the target label so it is **Classification**.

**Step 1:** Data Cleaning – identifying and replacing nulls.

**Step 2:** Data Preprocessing

**Step 3:** Input Split and Output Split

**Step 4:** Splitting Train and Test Set

**Step 5:** Using algorithms like Decision Tree, Logistic Regression or SVM.

**Step 6:** Determine classification matrix to find precision, recall, F1 measure.

**Step 7:** Classifying the spam messages from the created model.

6. A business wants to analyze customer reviews of its products and determine whether the sentiment is positive or negative.

**Problem type:** Analysing the customer reviews in the target label so it is **Classification**.

**Step 1:** Data Cleaning – identifying and replacing nulls.

**Step 2:** Data Preprocessing and converting text to numerical features.

**Step 3:** Splitting Train and Test Set

**Step 4:** Using algorithms like Decision Tree, Logistic Regression or SVM.

**Step 5:** Determine classification matrix to find precision, recall, F1 measure.

**Step 6:** Classifying the positive or negative reviews from the created model.

7. 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.

**Problem type:** Analysing the customer to file insurance in the target label so it is **Classification**.

**Step 1:** Data Collection – collecting customers history and details.

**Step 1:** Data Cleaning – identifying and replacing nulls.

**Step 2:** Data Preprocessing and converting categorical columns to numerical columns using one hot encoding.

**Step 3:** Splitting Train and Test Set

**Step 4:** Using algorithms like Decision Tree, Logistic Regression or SVM.

**Step 5:** Determine classification matrix to find precision, recall, F1 measure and plot AUC-ROC curve for evaluation.

**Step 6:** Predict the claim from the created model.

8. A streaming platform wants to recommend movies to users by grouping them based on their viewing preferences and watch history.

**Problem type:** Grouping movie preferences in the label so it is **Clustering**.

**Step 1:** Data collection from users watch history and viewing patterns.

**Step 2:** Data Preprocessing – Removing / dropping unnecessary data.

**Step 3:** Input Split and Output Split

**Step 4:** Using algorithms like K-means, DBScan, OPTICS.

**Step 5:** Determine clusters and outliers.

**Step 6:** Plot the graph to identify the number of cluster groups.

**Step 7:** Predict the user recommendations from the created model.

9. A hospital wants to predict the recovery time of patients after surgery based on their age, medical history, and lifestyle habits.

**Problem type:** Numerical values in the target label so it is **Regression**.

**Step 1:** Data Collection – Patients age, medical history and lifestyle habits.

**Step 2:** Data Cleaning – identifying and replacing nulls.

**Step 3:** Data Preprocessing – Categorical values into numerical by one hot encoding.

**Step 4: Input Split** – age, medical history and lifestyle habits.

**Output Split** – Recovery time after surgery

**Step 5:** Splitting Train and Test Set

**Step 6:** Using algorithms like Multilinear Regression, Decision Tree and SVM

**Step 7:** Finding the RMSE, R\_score value and saving the best model.

**Step 8:** Predict the recovery time of patients based on our requirements from the created model.

10. A university wants to predict a student's final exam score based on study hours, attendance, and past academic performance.

**Problem type:** Numerical values (score) in the target label so it is **Regression**.

**Step 1:** Data Collection – study hours, attendance, and past academic performance.

**Step 2:** Data Cleaning – identifying and replacing nulls.

**Step 3:** Data Preprocessing – Categorical values into numerical by one hot encoding.

**Step 4: Input Split** – study hours, attendance and past academic performance. **Output Split** – Exam score

**Step 5:** Splitting Train and Test Set

**Step 6:** Using algorithms like linear Regression, Decision Tree or SVM

**Step 7:** Finding the RMSE, R\_score value and saving the best model.

**Step 8:** Predict the exam score of students based on our requirements from the created model.