Scenario: 1) A bank wants to predict whether a loan applicant will default based on credit score, income, and past loan history.

Problem Type: Classification

Steps:

1. Data Collection

2. Data preprocess whether there is missing or null values & converting any categorical data to nominal data.

3. Splitting the input / independent variable as credit score, income and past loan history.

4. Assign the train and test set and fitting the train model to the loan default data.

5. Using the algorithms like Logistic Regression, Decision Tree or Random Forest.

6. Finding the Confusion Matrix and Classification report to find the best model.

7. Evaluating the performance using AUC – ROC curve, F1- Score, Recall and Precision.

8. Predicting the loan applicant default for new customers.

Scenario: 2) A retail store wants to predict the demand for different products to optimize inventory levels.

Problem Type: Regression

Steps:

1. Data Collection for the independent and target variable.

2. Data Preprocess for null, missing values and then converting any categorical values to a nominal data.

3. Splitting Input and Output. Assigning Train and Test set.

4. Using algorithms like Linear, SVM, Decision Tree and Random Forest to find the best R_score value

5. Evaluating the performance using R_score and RMSE to find the best model.

6. Predicting the demand of products for the inventory levels.

Scenario: 3) A factory wants to detect whether a manufactured product is defective based on sensor readings and quality control data.

Problem Type: Classification

Steps:

1. Gathering the data information like sensor readings and quality control data.

2. Data Preprocessing the null/missing values and converting categorical data to nominal if needed.

3. Splitting the train and test set.

4. Using the algorithms like Logistic Regression, Decision Tree or Random Forest.

5. Finding the Confusion Matrix and Classification report to find the best model.

6. Evaluating the performance using AUC – ROC curve, F1- Score, Recall and Precision.

7. Predicting the manufactured product is defective or not.

Scenario:4) A healthcare provider wants to analyze patient symptoms and classify them into different disease categories.

Problem Type: Classification

Steps:

1. Data Collection for patient symptoms of different disease categories.

2. Preprocessing the null/missing values and converting categorical data to nominal if needed.

3. Splitting the train and test set.

4. Using the algorithms like Logistic Regression, Decision Tree or Random Forest.

5. Finding the Confusion Matrix and Classification report to find the best model.

6. Evaluating the performance using AUC – ROC curve, F1- Score, Recall and Precision.

7. Predict the patient's disease based on the symptoms.

Scenario: 5) An e-commerce company wants to identify and remove fake reviews posted by bots or fraudsters.
Problem Type: Classification

Steps:

1. Collection of Data from E-commerce company.

2. Handling null or missing values.

3. Converting the text into Vectors (numbers) for model performance.

4. Feature engineering for recursive feature elimination.

5. Splitting the training set.

6. Using the algorithms like Naïve bayes, Random Forest or any other.

7. Evaluate the best model performance using clf report, recall, precision & Confusion matrix.

8. Predict the fake reviewers from bots or fraudsters.

Scenario: 6) A financial firm wants to predict stock price movements based on historical price data and market indicators.

Problem Type: Regression

Steps:

1. Data Collection of historical price movements and market indicators.

2. Data Preprocess for null, missing values and then converting any categorical values to a nominal data.

3. Splitting Input and Output. Assigning Train and Test set.

4. Using algorithms like Linear, SVM, Decision Tree and Random Forest to find the best R_score value

5. Evaluating the performance using R_score and RMSE to find the best model.

6. Predict the movements of stock price.

Scenario: 7) A social media platform wants to detect fake user accounts based on user activity and profile data.

Problem Type: Classification

Steps:

1. Data Collection of user activity and profile data of user accounts.

2. Handling null or missing values.

3. Converting the text into Vectors (numbers) for model performance.

4. Feature engineering for recursive feature elimination.

5. Splitting the training set.

6. Using the algorithms like Naïve bayes, Random Forest or any other.

7. Evaluate the best model performance using clf report, recall, precision & Confusion matrix.

7. Predict the fake user accounts based on their activity.

Scenario: 8) A marketing agency wants to segment customers into different groups based on their purchasing behavior.

Problem Type: Clustering

Steps:

1. Data Collection of purchasing history and products of marketing agency.

2. Data Preprocess for null, missing values and then converting any categorical values to a nominal data.

3. Feature Engineering to remove repeated data/values and selection the important features using select K method.

4. Using algorithms like K-means, Agglomerative or Fuzzy Clustering.

5. Applying algorithms to train the model.

6. Identify the cluster groups and Predict the groups for customer segmentation.

Scenario: 9) A geospatial research team wants to analyze satellite images to classify different land types (forest, water, urban).

Problem Type: Classification

Steps:

1. Collection of Satellite images.

2. Converting the images into normalize the noises and extract image features for model performance.

3. Feature engineering for recursive feature elimination.

4. Splitting the training set.

5. Using the algorithms like Decision Tree, Random Forest or CNN.

6. Evaluate the best model performance using clf report, accuracy, recall, precision & Confusion matrix.

7. Classify the images into different land types.

Scenario: 10) A streaming service wants to predict which users are likely to cancel their subscriptions.

Problem Type: Classification

Steps:

1. Data Collection of user subscription history.

2. Preprocessing the null/missing values and converting categorical data to nominal if needed.

3. Splitting the train and test set.

4. Using the algorithms like Logistic Regression, Decision Tree or Random Forest.

5. Finding the Confusion Matrix and Classification report to find the best model.

6. Evaluating the performance using AUC – ROC curve, F1- Score, Recall and Precision.

7. Predict the likely users cancelling their subscriptions.