**Name:** Rugved Dani **Roll No:** 281057

**Batch:** A3

# Problem Statement:

Assignment 4

Apply appropriate ML algorithm on a dataset. Create confusion matrix based on the data and find

1. Accuracy
2. Precision
3. Recall
4. F-1 score

# Objectives:

* 1. To apply a supervised machine learning algorithm to predict customer response.
  2. To analyze the dataset and preprocess the data for better model performance.
  3. To evaluate model performance using a confusion matrix.
  4. To compute key classification metrics (Accuracy, Precision, Recall, F1-score).

# Resources used:

1. Software used: Visual Studio Code
2. Libraries used: Pandas, Matplotlib, Seaborn, SKLearn

# Theory:

Classification is a supervised learning technique where the model learns to map input features to predefined labels. The goal is to train a model that can accurately classify new

data points into one of the given categories. In this assignment, we focus on binary classification (Customer will respond: Yes or No).

Confusion Matrix:

A confusion matrix is a performance measurement tool for classification models. It consists of four components:

* + True Positives (TP): Correctly predicted positive cases.
  + True Negatives (TN): Correctly predicted negative cases.
  + False Positives (FP): Incorrectly predicted positive cases (Type I Error).
  + False Negatives (FN): Incorrectly predicted negative cases (Type II Error).

Evaluation Metrics:

* + Accuracy: Measures the overall correctness of the model.
  + Precision: Measures how many predicted positive cases were actually positive.
  + Recall: Measures how many actual positive cases were correctly predicted.
  + F1-Score: Harmonic mean of precision and recall, balancing both metrics.

# Methodology:

1. Data Preprocessing
   * Load the dataset using Pandas.
   * Handle missing values (imputation or removal).
   * Encode categorical variables (e.g., gender) using one-hot encoding.
   * Normalize numerical features using MinMaxScaler or StandardScaler.
   * Split the dataset into training and testing sets (e.g., 75% training, 25% testing).
2. Choosing the ML Algorithm

Since the problem is a binary classification task, suitable algorithms include:

* + Logistic Regression
  + Decision Tree Classifier
  + Random Forest Classifier
  + Support Vector Machine (SVM)
  + K-Nearest Neighbors (KNN)
  + Neural Networks (optional for advanced modeling)

1. Model Training & Prediction
   * Train the selected ML model on the training dataset.
   * Predict customer responses on the test dataset.
2. Confusion Matrix & Performance Metrics Calculation
   * Compute the confusion matrix (True Positives, True Negatives, False Positives, False Negatives).
   * Derive the following metrics from the confusion matrix:
     1. Accuracy
     2. Precision
     3. Recall (Sensitivity)
     4. F1-Score

# Conclusion:

* + The chosen ML model was able to predict responses with reasonable accuracy.
  + Based on the evaluation metrics, the model’s performance can be assessed for further improvements.
  + Feature engineering and hyperparameter tuning could further enhance the model's effectiveness.