**Task 1: Term Deposit Subscription Prediction (Bank Marketing Dataset)**

# Problem Statement and Objective

The goal of this task is to predict whether a client will subscribe to a term deposit based on demographic and campaign-related features from the Bank Marketing Dataset (UCI Repository). By analyzing customer attributes and campaign performance, the model can help financial institutions identify which clients are more likely to subscribe to term deposits.

# Dataset Description and Loading

The dataset contains client-related attributes (e.g., age, job, marital status, education), as well as information about the marketing campaign (e.g., contact type, number of contacts, days since last contact). The target variable is 'y', which indicates whether the client subscribed to a term deposit ('yes' or 'no').

# Data Cleaning and Preprocessing

Steps performed:  
1. Handled missing values (if any).  
2. Converted categorical features into numerical form using Label Encoding / One-Hot Encoding.  
3. Normalized numerical features such as age, balance, and campaign-related variables.  
4. Split dataset into training and testing sets.

# Exploratory Data Analysis (EDA)

Exploratory Data Analysis was performed to understand feature distributions and relationships:  
- Age distribution of clients was right-skewed, with most clients between 30–40 years.  
- Married clients formed the largest group in the dataset.  
- Education level and job type showed correlation with term deposit subscription.  
- Clients contacted via cellular methods had a higher chance of subscription.

# Model Building and Evaluation

Multiple machine learning models were trained, including Logistic Regression and Decision Tree Classifier.  
  
Evaluation metrics used:  
- Accuracy

- Confusion Matrix  
- Precision, Recall, and F1-score  
- ROC Curve and AUC  
  
Results:  
- Logistic Regression performed well with balanced precision and recall.  
- Decision Tree provided higher interpretability but slightly lower generalization performance.

# Visualizations

Several charts were plotted during EDA and model evaluation:  
- Distribution plots of age and balance.  
- Bar charts for categorical features like job and education.  
- ROC Curve showing model performance.  
- Confusion Matrix heatmap to visualize classification results.

# Final Conclusion with Insights

The analysis revealed that term deposit subscriptions depend strongly on client demographics and contact strategies. Logistic Regression was effective in predicting subscriptions, while Decision Tree offered interpretability for business users. Insights indicate that campaigns should focus on middle-aged clients, cellular communication, and targeting specific job/education groups for better success rates.