Python Programming

Machine Learning Assignment

Bank Term Deposit Subscription Prediction

Domain: Banking, Marketing

Problem Statement:

A Portuguese bank conducted marketing campaigns to promote term deposit subscriptions. The goal is to predict whether a client will subscribe (yes or no) to a term deposit based on their profile and campaign interaction details.

Dataset Overview:

You will use the **Bank Marketing Dataset** from UCI repository.
Target Column: y

(yes = client subscribed to term deposit, no = did not subscribe)

Feature	Description
age	age of the client
job	job type (admin., technician, etc.)
marital	marital status
education	education level
default	has credit in default?
balance	average yearly account balance
housing	has housing loan?
loan	has personal loan?
contact	contact communication type
day	last contact day of the month
month	last contact month of year
duration	last contact duration
campaign	number of contacts during campaign
previous	number of contacts before this campaign
poutcome	outcome of previous campaign

Assignment Tasks:

1. **Load and Explore the Dataset**

   - Handle missing or unknown values (e.g., `unknown` in categorical features).

   - Display basic stats and visualize class distribution.

2. **Preprocess the Data**

   - Convert categorical variables using **Label Encoding** or **One-Hot Encoding**.

   - Scale numeric features (e.g., using **StandardScaler**).

3. **Split the Data**

   - Use 80% data for training and 20% for testing.

   - Apply `train_test_split()`.

4. **Train Classification Models**

   - Train the following models:

     - **Logistic Regression**

     - **K-Nearest Neighbors**

     - **Random Forest Classifier**

5. **Evaluate the Models**

   - Compare using:

     - Accuracy

     - Confusion Matrix

     - Classification Report

     - ROC-AUC score

6. **Visualize Results**

   - Plot confusion matrix and ROC curves.