Term Deposit Subscription Prediction Report

# 1. Executive Summary

This project involved building a predictive model to determine whether a client would subscribe to a term deposit, using features derived from direct marketing campaigns conducted by a banking institution. The solution includes data preprocessing, feature engineering, model training using a Random Forest Classifier, and deployment via an interactive Streamlit application.

# 2. Dataset Overview

The primary dataset used is `bank-additional-full.csv`, containing 41,188 records and 20 features (excluding the target). It is part of a larger collection of related datasets that track client responses to marketing campaigns between May 2008 and November 2010. The objective is to predict the binary outcome variable `y` (subscription: yes/no) based on demographic, financial, and contact attributes.

# 3. Exploratory Data Analysis

- Class imbalance observed in the target variable `y`: ~89% 'no', ~11% 'yes'.  
- The `duration` column was excluded due to data leakage risk.  
- Categorical features such as `job`, `education`, and `contact` showed clear variation with respect to the target.

- Economic features such as `euribor3m`, `emp.var.rate`, and `nr.employed` were strongly correlated.

# 4. Data Preprocessing

- One-hot encoding applied to categorical features.  
- Target variable `y` was mapped to binary values.  
- `duration` column was removed.  
- SMOTE was applied to balance the training set.  
- Dataset was split into training (80%) and testing (20%) sets.

# 5. Model Building

A `RandomForestClassifier` was trained on the balanced dataset. The model was evaluated using classification metrics such as accuracy, precision, recall, F1 score, and ROC AUC.

- Accuracy: ~88%  
- ROC AUC Score: High, indicating good model discrimination  
- Evaluation metrics were balanced after SMOTE application

# 6. Feature Importance

Top features identified by the model:  
- poutcome\_success  
- emp.var.rate  
- euribor3m  
- month\_dec  
- contact\_cellular

# 7. Deployment

The trained model was serialized using `joblib` and integrated into a Streamlit web application. The app provides a user-friendly interface for predicting term deposit subscription based on user input. It displays both the prediction (yes/no) and the associated probability.

# 8. Files Included in GitHub Repository

- `term\_deposit\_modeling.ipynb`: Full notebook with extended EDA, preprocessing, modeling, and export  
- `streamlit\_app.py`: Streamlit app code for live prediction  
- `term\_deposit\_model.pkl`: Trained model file  
- `README.md`: Setup and deployment instructions  
- `Term\_Deposit\_Report.docx`: This project report