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. Findings and Insights

* The model revealed several key insights:  
  Clients who had a successful outcome in a previous marketing campaign (`poutcome\_success`) were significantly more likely to subscribe.
* Economic indicators like **`emp.var.rate`, `euribor3m`, and `nr.employed`** had strong predictive power, indicating macroeconomic conditions influenced client decisions.
* Clients contacted via **`cellular`** were more responsive compared to those contacted via telephone.
* Subscription likelihood was higher in certain months (e.g., December), suggesting time-of-year trends.
* Clients with fewer campaign contacts were more likely to subscribe, hinting at over-contact fatigue.

**These findings suggest that future marketing campaigns should:**

* Prioritize contacts with a history of success
* Optimize timing to align with favorable macroeconomic periods
* Use cellular contacts more than landlines
* Limit the number of repeated contacts to avoid diminishing returns

# 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: Jupyter notebook with full analysis, EDA, modeling, and export.
* streamlit\_app.py: Streamlit app to collect input and display predictions.
* term\_deposit\_model.pkl: Trained Random Forest model saved with **joblib**.
* Term\_Deposit\_Report.docx: Project report with findings and approach.
* requirements.txt – Dependencies list for setup
* bank-additional-full.csv – Dataset used for training and evaluation
* screenshot.png – Visual preview of the deployed Streamlit app
* README.md: This setup and usage guide