# TERM DEPOSIT SUBSCRIPTION PREDICTION REPORT

## Introduction

This report presents the findings of a project aimed at predicting whether a bank client will subscribe to a term deposit or not. This is part of a broader effort to improve targeted marketing strategies in the financial services.

## Objectives

The main objective is to build a predictive model that determines the likelihood of a client subscribing to a term deposit, based on historical marketing data.
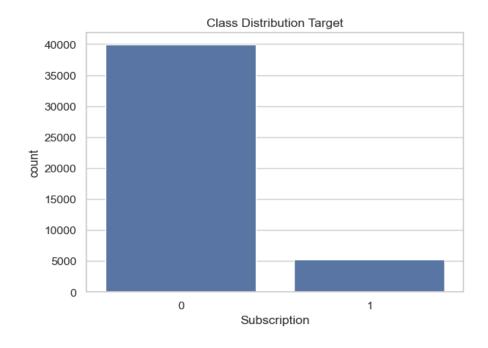
## Methodology

The project followed a structured process consisting of the following stages:

- Exploratory Data Analysis (EDA)
- Feature Engineering & Data Preprocessing
- Model Training and Evaluation

### 1. Exploratory Data Analysis (EDA)

Dataset Overview:  The dataset consists of *45,211 entries* and *17 columns*, with no missing values. The data types are a mix of integers and objects (categorical features).

Class Distribution:  The dataset is *imbalanced*, with a significantly higher number of 'no' than 'yes' . This imbalance was addressed using upsampling .

# TERM DEPOSIT SUBSCRIPTION PREDICTION REPORT

## 2. Feature Engineering

Target Encoding: The target variable, originally named y and renamed to Subscription, indicates whether a client subscribed to a term deposit ('yes' or 'no'). It was encoded into a binary format (1 for 'yes', 0 for 'no') for modelling.

Feature Identification: Categorical and numerical features were identified and separated for appropriate preprocessing before model training.

Numerical Features: The numerical columns are 'age', 'balance', 'day', 'duration', 'campaign', 'pdays', and 'previous'.

Categorical Features: The categorical columns include 'job', 'marital', 'education', 'default', 'housing', 'loan', 'contact', 'month', and 'poutcome'.

## 3. Model Training And Evaluation

An *XGBoost (Extreme Gradient Boosting)* model was used to predict term deposit subscriptions. XGBoost is a powerful and efficient algorithm known for its high performance in classification tasks.

### Model Performance Evaluation

The model's performance was evaluated using the following metrics:

**Accuracy:** The model achieved an accuracy of *87.24%.*

**Confusion Matrix:**

- True Positives (Subscribed): 849
- True Negatives (Not Subscribed): 7040
- False Positives: 945
- False Negatives: 209

**Classification Report:**

- Precision (Subscribed): 47% - the model is correct 47% of the time when it predicts a subscription
- Recall (Subscribed): 80% - the model correctly identifies 80% of all actual subscribers
- F1-Score (Subscribed): 60% - a weighted average of precision and recall

The high recall for subscriptions (80%) is a positive indicator, suggesting the model is effective at identifying potential customers. However, the lower precision (47%) means that a significant number of clients predicted to subscribe will not, which has implications for marketing campaign efficiency.

# TERM DEPOSIT SUBSCRIPTION PREDICTION REPORT

## Findings and Insights

The model demonstrates a good ability to identify clients who are likely to subscribe to a term deposit, as shown by the high recall. The overall accuracy is also high. The main challenge is the precision for the 'yes' class, which could lead to marketing efforts being spent on clients who are not likely to convert.

## Actionable Recommendations for the Marketing Team

Focus marketing efforts on clients with longer call durations and positive past outcomes. Consider real-time scoring of leads during campaigns for optimal resource use. Further investigate client demographics linked to high subscription likelihood.