Bank Marketing Campaign Analysis Report

---

1. Introduction

The objective of this project is to analyze the Bank Marketing Dataset and predict whether a client will subscribe to a term deposit.

The dataset contains client information (job, marital status, education, etc.), economic indicators (interest rates, consumer price index), and details of marketing campaigns.
By applying data analysis and machine learning models, this project aims to:
Identify the key factors that influence client decisions.
Build predictive models to improve campaign targeting.
Provide recommendations for future marketing strategies.

---

2. Data Analysis

➢ 2.1 Dataset Overview

Source: bankmarketing.csv
Rows: ~41,000
Target Variable: y (Yes = client subscribed, No = client did not subscribe)

➢ 2.2 Data Cleaning & Preparation

Handled categorical variables using one-hot encoding.
Checked for missing values and data imbalances.
Split data into train (80%) and test (20%) sets.

➢ 2.3 Exploratory Data Analysis (EDA)

Target distribution: Around 88% “No” and 12% “Yes” → indicates class imbalance.
Important insights from EDA:
Clients with previous successful outcomes are more likely to subscribe.
Longer call duration is positively correlated with success.
Certain months (March, August, December) show higher subscription rates.
Economic indicators (interest rates, CPI) affect decision-making.

---

3. Model Building

Models Used

Logistic Regression
Decision TreeRandom Forest
Support Vector Machine (SVM)

Evaluation Metrics

Accuracy
Confusion Matrix
Classification Report (Precision, Recall, F1-score)
ROC-AUC Score

---

4. Results

➢ 4.1 Feature Importance (Logistic Regression Coefficients)

Top contributing features:

- Month (March, August, December)
- Duration of call
- Past campaign outcome (success)
- Consumer price index (cons.price.idx)
- Interest rate (euribor3m)

Key Insight: Campaign success depends mainly on timing, call quality, and past behavior.

➢ 4.2 Model Comparison

Model	Accuracy	ROC-AUC
Logistic Regression	91.6%	0.94
Decision Tree	89.4%	~0.89
Random Forest	91.8%	0.95
Support Vector Machine (SVM)	91.4%	0.93

Best Model: Random Forest (highest accuracy + ROC-AUC).
Most Interpretable Model: Logistic Regression (easy to explain coefficients).

---

5. Recommendations

➢ 5.1 For Model Deployment

Deploy the Random Forest model for prediction in real campaigns.
Use Logistic Regression for generating explainable insights for managers.

➢ 5.2 For Marketing Strategy

Focus on campaign timing: March, August, and December are high-response months.
Improve call quality: Encourage longer and more engaging customer calls.
Target past positive respondents: They have a much higher chance of subscribing.

➢ 5.3 For Monitoring

Regularly track economic conditions (interest rates, CPI) to adapt strategies accordingly.

---

6. Conclusion

This project demonstrated that machine learning can significantly improve marketing campaign strategies in the banking sector.
Random Forest emerged as the most accurate and reliable model.
Key features such as call duration, campaign month, and past success drive customer subscription.
By adopting these insights, the bank can reduce costs, improve targeting, and maximize campaign effectiveness.