# Loan Prediction based on Marketing Campaigns

## *Executive Summary*

This project aimed to build a predictive model capable of identifying whether a bank client would subscribe to a term deposit after being contacted during a marketing campaign.

Using the Bank Marketing Dataset from a Portuguese institution, we conducted a full data-driven analysis combining exploratory data analysis (EDA), modelling, and business interpretation.

The key findings are summarized below:

- Model Performance: Among the models tested, Bagging achieved the best results, with an accuracy of 84% and ROC-AUC of 0.84, outperforming both Logistic Regression (A = 0.82) and Decision Tree (A = 0.80).

- Important Predictors:

  ⇒ **Call duration** — longer calls are strongly associated with higher subscription probability.

  ⇒ **Previous successful contact (poutcome)** — clients who had previously subscribed are much more likely to do so again.

  ⇒ **Number of previous interactions (previous)** — consistent follow-ups improve conversion likelihood.

  ⇒ **Account balance** and **contact type** also influence the likelihood of subscription.

- Business Insight: Marketing success is highly influenced by customer engagement rather than demographic attributes. Optimizing contact strategy (longer, better-timed calls with previously interested clients) could significantly improve campaign ROI.

## *Technical Approach*

### 1) Data Understanding and Preparation

The dataset contained 11 162 records representing individual marketing contacts. Each record included demographic, financial, and campaign-related variables.

Data integrity checks ensured there were:

- No missing values or duplicates.
- Logical categorical fields (e.g., job types, education levels).

Categorical variables were one-hot-encoded, and numeric variables were standardized using StandardScaler to prepare them for model input.

### 2) Exploratory Data Analysis (EDA)

EDA focused on understanding variable distributions and their relationships with the target variable (deposit).

- Age: Most clients were between 25–45 years old.
- Balance: Highly skewed, with few clients having large positive balances.
- Duration: Positively correlated with deposit subscriptions — longer calls typically led to higher conversion.
- Campaign and Previous Contacts: Most clients were contacted once, but multiple previous interactions increased success probability.

The target variable (**deposit**) was nearly balanced, which ensured stable model training without resampling.

### 3) Modelling Strategy

Three classification algorithms were trained and compared:

| Model | Type | Accuracy | ROC-AUC | Key Notes |
|---|---|---|---|---|
| **Logistic Regression** | Linear | 0.82 | 0.90 | Baseline, interpretable |
| **Decision Tree** | Non-linear | 0.80 | 0.87 | Captures interactions, interpretable |
| **Bagging** | Ensemble | **0.84** | **0.91** | Best generalization and stability |

Evaluation metrics: Accuracy, Precision, Recall, F1-Score, and ROC-AUC.

Confusion matrices were used to visualize true/false positives and negatives.

### 4) Best Solution – Bagging

Why it performed best:

- It combines multiple trees to reduce variance and improve generalization.

- Handles complex nonlinear relationships between demographic and behavioural variables.
- Less sensitive to overfitting due to bootstrapping and feature randomization.

Quantitative Summary:

⇒ **Accuracy**: 0.836

⇒ **Precision**: 0.809

⇒ **Recall**: 0.861

⇒ **F1-Score**: 0.834

⇒ **ROC-AUC:** 0.915

The Bagging model provided both strong predictive power and robustness, outperforming single-tree and linear approaches.

## *Recommendations for Managers*

- **Prioritize client engagement quality over contact quantity.**

Longer and more meaningful conversations can lead to higher deposit conversions. Agents should be encouraged to spend more time with receptive clients rather than making brief, repetitive calls.

- **Target previously engaged clients.**

Clients who previously subscribed or interacted positively (successful poutcome) are the strongest leads.

Campaigns should segment and prioritize these customers first.

- **Use predictive modelling in campaign planning.**

Integrating the Bagging model into the CRM system can help score clients by likelihood to subscribe, enabling data-driven targeting.

- **Optimize campaign frequency.**

The variable campaign shows diminishing returns for clients contacted too often. Future strategies should focus on quality follow-ups instead of high repetition.

- **Balance resource allocation.**

Since not all demographic groups respond equally, marketing managers should invest more resources in segments with high predicted probabilities rather than uniform outreach.

## *Additional Observations & Future Work*

**Explainability:** decision trees remain valuable for transparency. Even if Bagging perform best, a pruned tree can be extracted for management understanding of "why" a client is predicted to subscribe.

**Feature Engineering Opportunities:** future versions could incorporate engineered variables such as:

- Interaction terms (e.g., duration × previous).
- Binning of continuous variables (e.g., grouping call durations).
- Handling outliers in balance via log transformation.

**Deployment Potential:** with 84% accuracy and 91% ROC-AUC, the Bagging model is sufficiently mature to be deployed as a lead-scoring tool. Integrating it into marketing workflows can directly support campaign decisions.

## *Conclusion*

This analysis demonstrates that machine learning can significantly improve the efficiency of marketing campaigns in the banking sector.
With EDA, interpretable modelling, and ensemble techniques, the project successfully identified the key behavioural and engagement factors influencing client decisions.

Overall, this study confirms that predictive analytics can transform traditional outreach efforts into smarter, more efficient customer engagement strategies with tangible financial benefits.