Final Project Report
Portuguese Bank Marketing Campaign

1. Introduction & Business Context

Banks invest significant resources in direct marketing campaigns to promote financial products such as term deposits. However, contacting every customer is costly and inefficient. The objective of this project is to help the bank's marketing team identify customers who are more likely to subscribe to a term deposit, enabling targeted and data-driven decision-making.

2. Data Understanding

The dataset is based on direct phone call marketing campaigns conducted by a Portuguese banking institution between 2008 and 2010. It includes customer demographics, financial background, campaign history, and macroeconomic indicators. The target variable indicates whether the customer subscribed to a term deposit.

3. Data Cleaning & Preprocessing

Missing values were handled carefully to avoid information loss. Categorical variables were encoded using one-hot encoding. The duration variable, although highly predictive, was excluded from realistic production models to avoid data leakage. These steps ensured the dataset was reliable and model-ready.

4. Exploratory Data Analysis (EDA)

EDA revealed that factors such as age, job type, previous campaign outcome, and contact method strongly influenced subscription behavior. Customers contacted via cellular phones and those with successful past campaigns showed higher conversion rates.

5. Model Building Approach

Multiple models were trained, including Logistic Regression, Random Forest, and XGBoost. Given the class imbalance, models were evaluated using ROC-AUC, precision, recall, and F1-score rather than accuracy alone.

6. Model Comparison & Selection

Logistic Regression served as a baseline model. Random Forest improved performance by capturing non-linear patterns. XGBoost delivered the best overall results and was selected as the final production model due to its superior ROC-AUC and recall performance.

7. Key Factors Influencing Subscription

Important predictors included previous campaign outcome, number of contacts, contact type, economic indicators such as Euribor rate, and customer demographics. These insights aligned well with real-world banking intuition.

8. Business Recommendations

The bank should prioritize customers with successful past campaign outcomes, focus on cellular-based contact methods, optimize contact frequency, and leverage favorable economic conditions to maximize subscription rates.

9. Challenges Faced

Key challenges included class imbalance, risk of data leakage, and high-cardinality categorical variables. These were addressed using appropriate evaluation metrics, careful feature selection, and industry-standard preprocessing techniques.

10. Conclusion

This project demonstrates how data-driven techniques can significantly improve marketing efficiency in the banking sector. The final model provides both strong predictive performance and actionable insights for business decision-making.