Predictive Modeling of Bank Marketing Campaign Responses Using Machine Learning

Komal Kothawade[*], Mayuri Babar, Deepali Akolkar, Neha Chothe

Department of Statistics, Dr. D. Y. Patil Arts, Commerce and Science College, Pimpri, Pune-18, Maharashtra, India

Abstract: This study aims to develop a predictive model to assess client responses to bank marketing campaigns. Using an open-source dataset derived from a Portuguese bank's marketing efforts and hosted on Kaggle, we apply various classification algorithms including Logistic Regression, Random Forest, and LightGBM. The study involves thorough preprocessing, feature engineering, and model evaluation using ROC-AUC and F1 metrics. The best performing model achieved an ROC-AUC of approximately 0.80 using LightGBM, with SHAP analysis revealing the most influential factors.

Keywords: Bank marketing, customer response prediction, machine learning, SHAP, ROC-AUC

I. Introduction

In today's digital banking environment, understanding customer behavior is essential for designing effective marketing campaigns. Predicting whether a client will subscribe to a financial product, such as a term deposit, allows banks to allocate their outreach efforts more efficiently. With the rise of big data and computational tools, machine learning has become a powerful approach for analyzing customer information and campaign performance. These models can uncover hidden patterns in demographic and behavioral data that traditional techniques may miss. This study applies several machine learning methods to real-world banking data to evaluate their predictive accuracy and interpretability. Emphasis is also placed on identifying the most influential factors affecting customer decisions.

Dataset Description

We utilized a dataset sourced from a Portuguese financial institution containing more than 40,000 client entries, covering client demographics (age, job, marital status), financial indicators (balance, loan), and campaign-related features (number of contacts, outcome of previous campaigns). The target variable is binary, indicating subscription to a term deposit.

## II. Methodology

A. Preprocessing

The dataset underwent several preprocessing steps to ensure it was ready for model training. Categorical variables were converted into numerical format using label encoding and one-hot encoding where appropriate. Missing values were addressed using imputation techniques to maintain data integrity. Numerical features were standardized to bring them onto a similar scale, improving algorithm performance. Outliers were checked and handled to reduce noise in the model. These steps helped in creating a clean and structured dataset, which is crucial for building reliable machine learning models. The preprocessed data served as the foundation for feature selection and model evaluation.

B. Feature Engineering

Feature selection techniques included both Recursive Feature Elimination and correlation-based analysis.Key features included 'duration', 'euribor3m' (The average **interest rate** that **Eurozone banks** are willing to lend money to each other **for a period of 3 months**), and 'employment variation rate'.

C. Model Selection and Evaluation

We implemented Logistic Regression, Random Forest, and LightGBM. Model assessments incorporated metrics such as ROC-AUC, precision, recall, and the F1-score. Cross-validation ensured the robustness of the models.

Feature Importance (SHAP Values)
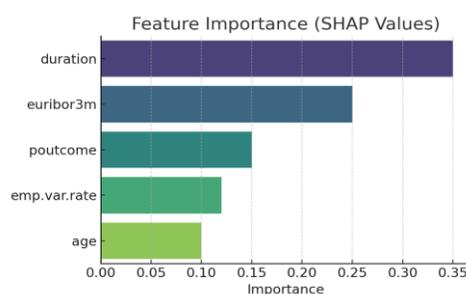
III. Results
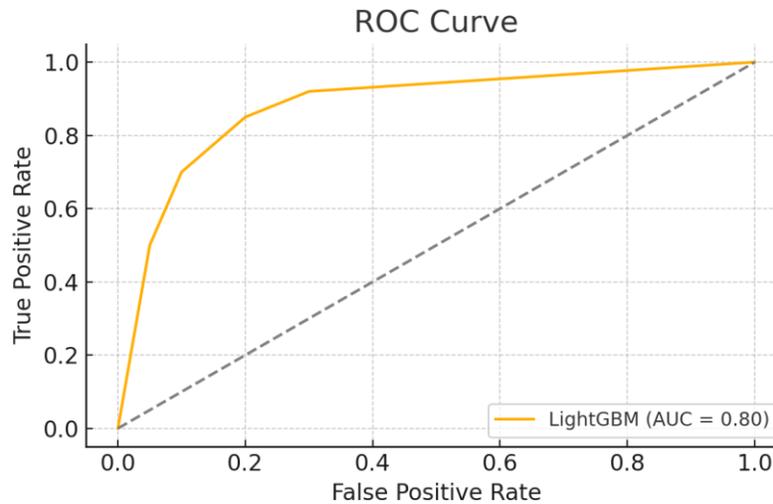
SHAP analysis revealed that **'duration'** (the last contact duration), **'euribor3m'** (3-month Eurozone interest rate), and **'poutcome'** (outcome of previous marketing campaigns) were the most influential features driving model predictions. This interpretability not only validates the model's internal logic but also provides actionable insights for marketers—highlighting that successful engagements and current economic conditions heavily influence customer decisions.

The ROC curve evaluation of the LightGBM model indicates reliable classification capability, achieving an AUC of 0.80. This result suggests that the model effectively separates clients who are likely to subscribe from those who are not. Its high true positive rate, particularly at lower false positive thresholds, makes it a practical tool for targeted marketing—helping banks focus their efforts on genuinely interested clients while avoiding excessive outreach.

IV. Discussion

The results highlight the advantage of using advanced ensemble methods, particularly LightGBM, for marketing response prediction. Compared to traditional models like Logistic Regression, LightGBM captured non-linear relationships more effectively, leading to better performance metrics. The use of SHAP values provided a transparent way to understand feature contributions, making the model insights more accessible to business stakeholders. Variables such as call duration, prior campaign outcome, and economic indicators were found to significantly influence client behavior. These insights can help refine marketing strategies and improve campaign efficiency. The balance between predictive power and interpretability remains key in deploying such models in real-world banking environments.

V. Conclusion

The study showed that machine learning methods are effective in forecasting client responses to bank marketing efforts. Of all the models assessed, LightGBM delivered the highest performance, with an AUC score of 0.80, reflecting strong predictive accuracy. SHAP analysis added interpretability by highlighting key features like call duration and economic indicators. The findings suggest that predictive modeling can significantly enhance targeted marketing strategies in the banking sector. Additionally, the findings advocate for the use of explainable AI tools to enhance transparency and build trust. Future work may explore temporal patterns and neural network-based methods to improve prediction accuracy further.

VI. Acknowledgment

References

1. Kaggle Dataset: https://www.kaggle.com/datasets/kukuroo3/bank-marketing-response-predict Lundberg, S.M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. NIPS.
2. Breiman, L. (2001). Random Forests. Machine Learning Journal.