Proceedings of ICFTBA 2025 Symposium: Data-Driven Decision Making in Business and Economics
DOI: 10.54254/2754-1169/2025.BL29270

Findings from the Bank Marketing Dataset: Using Machine Learning to Forecast Term Deposit Subscriptions

Shurui Du

College of Art and Science, Boston University, Boston, USA
shuruidu@bu.edu

Abstract. Banks face the challenge of determining which customers are most likely to react effectively to marketing initiatives in a financial climate that is becoming more and more competitive. Using a large dataset of campaign-related, financial, and demographic data, this study applies machine learning techniques to forecast term deposit subscriptions. To compare many classifiers based on various performance metrics, the study uses a methodical procedure that includes data pretreatment, exploratory data analysis, model training, and evaluation. The findings show that, in comparison to traditional models like logistic regression and decision trees, ensemble approaches—in particular, gradient boosting—achieve improved prediction accuracy and generalization. Gradient Boosting is the most effective classifier for reducing class imbalance in subscription prediction since it attains the biggest area under the Receiver Operating Characteristic (ROC) curve while maintaining a fair balance between precision and recall. These findings demonstrate how targeted marketing campaigns in the banking industry could benefit from sophisticated predictive modeling. Machine learning models maximize marketing return on investment, decrease needless interactions, and improve customer happiness by more accurately identifying high-potential clients. The study emphasizes how crucial it is for financial marketing strategy to include data-driven decision-making as a fundamental element.

Keywords: Predictive modeling, machine learning, financial marketing strategy, customer segmentation, precision-recall

1. Introduction

Effective marketing strategies are crucial for attracting and keeping customers in the cutthroat banking sector. Low conversion rates from traditional telemarketing campaigns frequently lead to resource waste and dissatisfied customers. In response to this challenge, recent studies have increasingly used machine learning and predictive analytics techniques to improve campaign targeting and design [1].

There have always been issues with traditional telemarketing, including low conversion rates, high costs, and insufficient customer interaction. This demonstrates how outmoded approaches are ineffective in the modern economy. In order to address these issues and look into more effective marketing strategies, both academia and business have embraced data-driven approaches more and more. The goal is to enhance campaign outcomes by using predictive analytics and customer

DOI: 10.54254/2754-1169/2025.BL29270

segmentation. Banks can now anticipate client reactions and make better use of their resources thanks to machine learning.In this context, the study employs the Bank Marketing Dataset, obtained from telemarketing campaigns conducted by a Portuguese banking institution [2]. This dataset, which combines campaign-specific variables with demographic data, has become a widely used standard for subscription forecasting predictive modeling tasks. The first goal of this study is to assess prediction models using interpretable baselines like logistic regression and advanced ensemble techniques like random forests and gradient boosting. The second goal is to extract practical insights that banks can use to improve the effectiveness of their campaigns and customer responsiveness.

The study advances methodological development and managerial practice by using systematic preprocessing, exploratory analysis, and stringent model evaluation. The findings enhance contemporary research that underscores the importance of ensemble approaches and interpretable machine learning in financial contexts, offering comparative information about algorithmic effectiveness and practical ramifications for data-driven marketing tactics [3,4].

The dataset analyzed in this study originates from a comprehensive telemarketing campaign promoting term deposit subscriptions. There are 45,211 observations in the dataset, spread over 17 categories. These include demographic, financial, and campaign-related characteristics. The dependent variable is binary, indicating whether a client has subscribed to a term deposit. There are both categories and numerical factors in the attributes. Work, marital status, education, type of contact, and outcomes of previous campaigns are all examples of categorical variables. Many of these include "unknown" values that are encoded in a systematic way. Age, balance, duration, and campaign frequency are all examples of numerical variables that have skewed distributions with big tails and outliers, especially in account balance and call duration. A key aspect of the numbers is the considerable class imbalance, as only 11% of clients have opted for a term deposit. This difference means that you need to carefully choose your evaluation measures, since accuracy alone may hide poor performance in finding positive cases. There are three main problems with the dataset: the target variable is very unbalanced, certain categorical fields have "unknown" values, and some numeric features have long-tailed distributions. These traits helped us decide how to preprocess and model the data for the next study.

2. Data preprocessing

The preprocessing was meant to fix schema integrity, handle category unknowns, fix numeric skewness, and reduce class imbalance. To keep their predictive value, categorical variables with "unknown" entries were kept as separate categories. Binary fields, including loans and housing, were turned into indicator variables. Other nominal variables were one-hot encoded, with categories that didn't happen very often being combined.

Winsorization was used to deal with outliers in balance and duration for numerical variables, and standardized z-scores were used to make sure that features could be compared. Two modeling methods were used to figure out how long the duration would last, which was thought to be prone to information leaks. One method used it to estimate the best possible performance, and the other method left it out to guess how quickly it could be deployed in real time.

Stratified sampling was used to divide the dataset into training and test subsets so that the classes stayed balanced. Stratified k-fold cross-validation helped optimize the model during training. Class imbalance was alleviated by employing class weights and conducting sensitivity analyses that included resampling. To stop data leakage, all changes were only made to training folds. This

DOI: 10.54254/2754-1169/2025.BL29270

preprocessing methodology ensured statistical integrity, preserved essential categorical signals, and established a foundation for thorough and impartial model evaluation.

3. Exploratory Data Analysis (EDA)

Exploratory study revealed substantial differences in demographic, financial, and campaign-related factors. The age distribution was skewed to the right, with most clients being between 30 and 50 years old, with the most clientele being in their late thirties. Clients over 60 had higher subscription rates, but they made up a smaller part of the sample. Occupational and educational variables strongly influenced outcomes: elderly and students exceeded 20 percent subscription rates, whilst blue-collar and service workers stayed below 10 percent. Clients with tertiary education demonstrated a greater likelihood of subscribing, suggesting that financial awareness is a crucial factor.

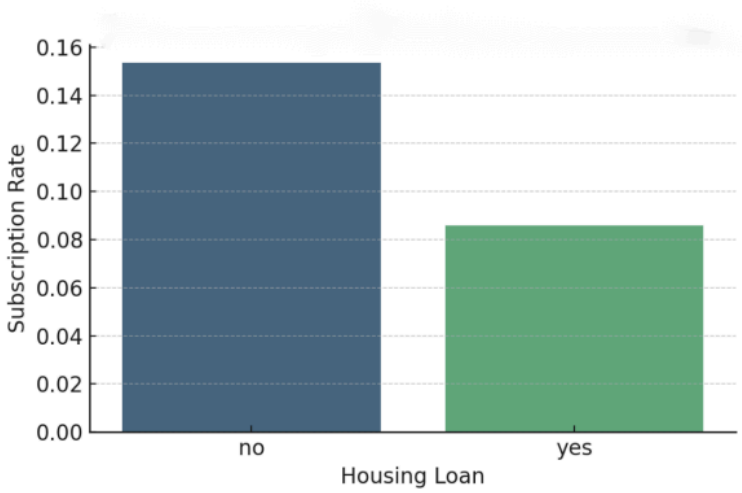

Figure 1. Subscription rate by housing loan

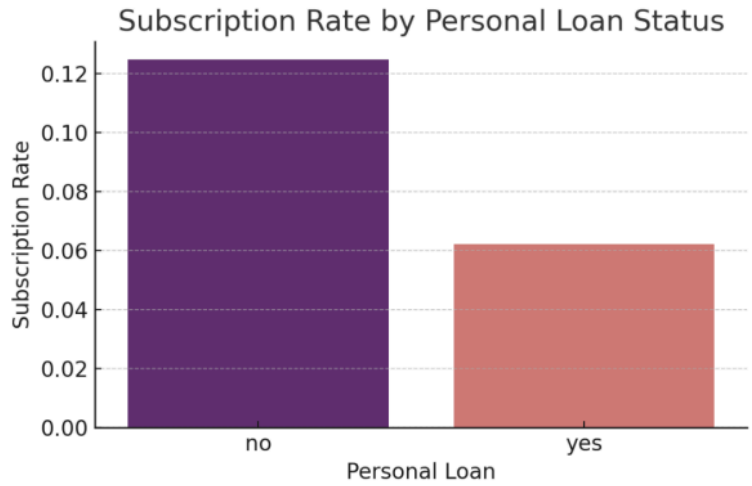

Figure 2. Personal loan status

The financial characteristics highlighted the importance of stability. Higher account balances were linked to higher responsiveness, but many clients had account balances that were close to zero

DOI: 10.54254/2754-1169/2025.BL29270

or negative, which meant they were more likely to go over their limits. Debt obligations exhibited a negative association with subscription rates. Figure 1 shows that clients who didn't have house loans signed up more often than clients who did. Figure 2 also shows that clients who didn't have personal loans had a lot more answers than those who did. These findings underscore the manner in which liabilities constrain investment options and reduce flexibility.

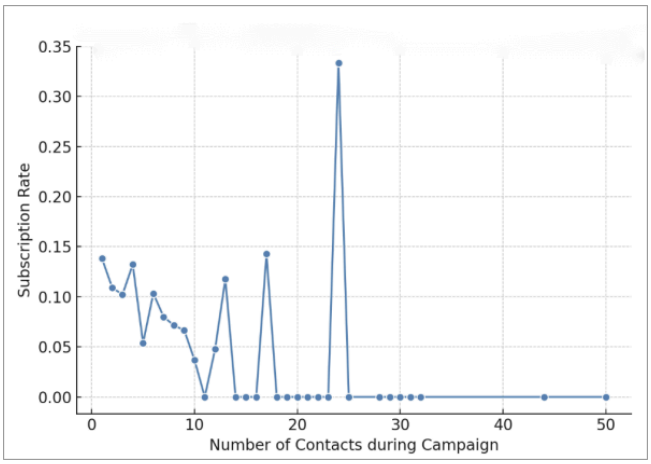

Figure 3. Subscription rate versus number of campaign contacts

The way the campaign worked had a big effect on the results. Subscription rates went up when people were called often at first, but they went down after they were asked too many times, showing that the returns were getting smaller. Figure 3 shows this tendency. It shows that while persistence can improve results at first, too much effort after three or four meetings makes things far less successful. Outliers with high contact counts, like the spike at 24 contacts, come from very small sample sizes and shouldn't be seen as important patterns.

Overall, the exploratory research showed that demographic profiles, financial capacities, and campaign strategies all work together to affect subscription outcomes. These results informed later feature engineering and highlighted the need for models that can account for nonlinear interactions and uneven outcomes.

4. Model selection

Model selection was informed by the necessity to equilibrate interpretability, predictive efficacy, and resilience in the context of class imbalance. Four algorithms were recognized, demonstrating escalating levels of complexity and prediction proficiency.

Logistic regression was chosen as the baseline because it is easy to understand and is well-known for being important in financial analytics. To make generalization better, the regularization parameters were changed. To fix the class imbalance, class weights and threshold modifications were made.

Decision tree is to explain how nonlinear linkages and interactions function. They were able to avoid overfitting by carefully controlling the depth, node size, and pruning settings.

This method got better when random forests combined a lot of decision trees using bootstrap sampling and random feature selection. This made the variance smaller and the generalization better [5]. Recent research has confirmed their efficacy in financial forecasting tasks, especially when compared to simpler benchmarks [3].

DOI: 10.54254/2754-1169/2025.BL29270

Gradient boosting was selected as the most advanced method due to its proven ability to characterize complex nonlinearities and repeatedly correct residual mistakes [6]. Recent studies have highlighted the improved precision and clarity of gradient boosting and related methodologies such as CatBoost, particularly in the context of bank marketing prediction difficulties [1]. To avoid overfitting, hyperparameters including learning rate, depth, and subsampling ratios were improved by halting early.

Precision, recall, F1-score, and Receiver Operating Characteristic curve (AUC) were among the measures used to assess each model using stratified k-fold cross-validation. This guaranteed equitable evaluation in the face of inequity and offered a thorough perspective on predictive potential.

5. Model evaluation

Four classifiers, Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting, were trained on the dataset using a 70–30 stratified split with cross-validated hyperparameter optimization in order to fully evaluate the effectiveness of the model. Accuracy, precision, recall, F1-score, and Receiver Operating Characteristic (ROC)-AUC were used to assess performance. This multi-metric approach is essential in imbalanced classification, since recall, F1-score, and AUC provide more reliable metrics of model efficacy, while accuracy may conceal insufficient identification of minority outcomes [7-9].

Table 1. Model performance comparison

Model	Accuracy	Precision	Recall	F1-score	AUC
Logistic Regression	0.8917	0.5692	0.2372	0.3348	0.8732
Decision Tree	0.8850	0.5000	0.3462	0.4091	0.8413
Random Forest	0.8924	0.6190	0.1667	0.2626	0.8741
Gradient Boosting	0.8917	0.5517	0.3077	0.3951	0.8882

Table 1 summarizes the findings. Despite all models attaining excellent accuracy (>0.88), the class-sensitive measures indicated significant trade-offs. Logistic Regression yielded balanced albeit moderate performance (F1 ≈ 0.33; AUC ≈ 0.87). The Decision Tree achieved the highest recall (≈0.35) and F1 score (≈0.41), although it exhibited the lowest AUC (≈0.84), suggesting constrained generalization. Random Forest had the highest accuracy (0.8924) and robust precision (0.62), although it exhibited very low recall (0.17), resulting in the lowest F1 score (0.26). Gradient Boosting yielded the most reliable overall outcomes, with a robust AUC of 0.8882 and balanced precision of 0.55 alongside a recall of 0.31, resulting in an F1 score of approximately 0.40.

DOI: 10.54254/2754-1169/2025.BL29270

Figure 4. Visual model preformance comparison

Figure 4 provides a visual comparison of accuracy, F1-score, and AUC. Although accuracy values were closely grouped among models, F1 and AUC provided a more distinct differentiation in performance, affirming that accuracy alone is inadequate in the context of class imbalance [7].

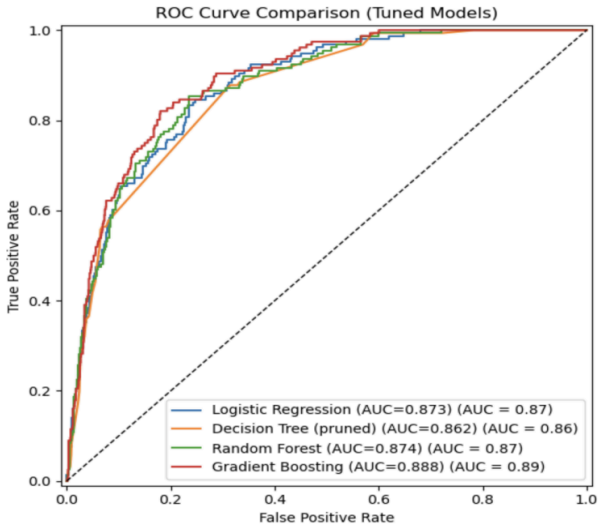

Figure 5. ROC curve comparison

Figure 5 illustrates ROC curves. Gradient Boosting regularly neared the optimal frontier, whereas the Decision Tree fell short. Insignificant discrepancies between tabular AUC values and integrated ROC curves arose from floating-point precision and curve discretization, without influencing findings.

Despite the models demonstrating competitive performance, particularly Gradient Boosting with balanced precision and recall, the overall recall values were still small. This suggests that the classifiers continue to have difficulty identifying a greater share of real positive subscriptions, a prevalent challenge in imbalanced financial datasets. This happened for a few reasons: (1) the inherent class imbalance, with just 11% of consumers subscribing; (2) the inclusion of categorical categories with "unknown" values that weaken predictive signals; and (3) distorted numerical variables like balance and duration that add noise even after preprocessing. In future investigations, these problems could be rectified by utilizing better resampling methods like Synthetic Minority

DOI: 10.54254/2754-1169/2025.BL29270

Over-sampling TEchnique (SMOTE) or Adaptive Synthetic Sampling (ADASYN), cost-sensitive learning, or sophisticated ensemble methods like Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), or Categorical Boosting (CatBoost). Also, banks can utilize tools like SHapley Additive exPlanations (SHAP) or Local Interpretable Model-agnostic Explanation (LIME) to assist them understand how certain aspects affect the outcome. This will help them make better choices. Recall and robustness may be increased by supplementing the dataset with more recent or thorough consumer behavior features.

In comparison to simpler models, ensemble approaches, particularly Gradient Boosting, showed improved precision-recall trade-offs and increased generalization. When assessing imbalanced financial prediction tasks, these results highlight the need to use evaluation metrics other than accuracy [8,9].

6. Discussion and implications

The assessment shows that ensemble methods, especially Gradient Boosting, work better for making predictions when classes are not balanced. This has both methodological and practical consequences.

From a methodological standpoint, the findings underscore the necessity of employing multi-metric assessment frameworks that extend beyond mere accuracy, as recall, F1-score, and ROC-AUC provide more dependable information in contexts characterized by imbalanced categorization. For structured financial data, the results further corroborate the comparative benefits of ensemble methods over linear and single-tree models.

From the standpoint of the application, more precise identification of likely subscribers enables more efficient use of resources in marketing campaigns. By cutting down on pointless interactions and increasing targeting accuracy, banks can lower operating expenses while simultaneously raising customer happiness. Furthermore, noting essential variables like account balance, loan status, and earlier campaign outcomes provides useful information for campaign planning and client relationship management.

In conclusion, the study links methodological advancement with managerial decision-making and highlights the need of integrating predictive modeling into financial marketing strategies.

7. Conclusion

This study examined how machine learning may improve telemarketing outcomes for term deposit subscriptions using predictive modeling on the Bank Marketing dataset. To assess logistic regression, decision trees, random forests, and gradient boosting, the study employed exploratory analysis, systematic preprocessing, and model evaluation.

The results demonstrate that all models were very accurate, although ensemble techniques had apparent benefits. Gradient boosting was the most fair when it came to recall, F1-score, and ROC-AUC. This shows that it can handle class imbalance. The results are in line with new studies that illustrate how well ensemble learning works for predicting financial outcomes.

There are two conclusions. The study emphasizes the importance of multi-metric evaluation for equitable assessment in instances with skewed distributions. The findings indicate that enhancing targeting accuracy and minimizing superfluous outreach can enable predictive models to facilitate more efficient and customer-centric marketing strategies.

In conclusion, new ensemble methods, such as Gradient Boosting, improve forecast accuracy and provide valuable insights into what influences how customers respond. These models simulate

DOI: 10.54254/2754-1169/2025.BL29270

nonlinear interactions between demographic, financial, and campaign-related factors to help explain why particular clients are more likely to respond to term deposit offers. With this newfound insight, marketing managers may devise targeted plans that target high-potential prospects while reducing waste. Furthermore, gradient boosting is adaptable and scalable, making it suited for use in real-time decision support systems. This enables banks to continuously improve their marketing strategies as fresh data becomes available. These advantages demonstrate the importance of ensemble learning in two ways: it increases the accuracy of approaches for dealing with unbalanced classification problems and provides essential information for making data-driven judgments in the banking industry.

References

[1] Yu, Q. et al. (2025) Enhancing Bank Term Deposit Predictions: A Machine Learning Approach With CatBoost and SHAP. Applied and Computational Engineering, 120, 171-180.
[2] Moro, S., Cortez, P. & Rita, P. (2014) A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, 62, 22-31.
[3] UCI Machine Learning Repository. (2024) Bank Marketing Data Set. Retrieved from https: //archive.ics.uci.edu/dataset/222/bank+marketing
[4] Peter, M. et al. (2025) Predicting Customer Subscription in Bank Telemarketing Campaigns Using Hybrid Ensemble Models. Journal of Data Science and Artificial Intelligence, 3(1), 15-28.
[5] Tanvir, M. F., Hossain, M. & Asifuzzaman, J. (2024) Bayesian Regression for Predicting Subscription to Bank Term Deposits in Direct Marketing Campaigns. ArXiv Preprint, arXiv: 2410.21539.
[6] Breiman, L. (2001) Random Forests. Machine Learning, 45(1), 5-32.
[7] Friedman, J. H. (2001) Greedy Function Approximation: A Gradient Boosting Machine. Annals of Statistics, 29(5), 1189-1232.
[8] Chen, W., Zhan, Z. & Wang, J. (2024) A Survey on Imbalanced Learning: Latest Research and Future Directions. Artificial Intelligence Review.
[9] López-Pinaya, W. H. et al. (2023) Evaluating Classifier Performance With Highly Imbalanced Big Data. Journal of Big Data, 10, 54.