

Prediction of Customer Time Deposit Subscriptions Based on Bootstrap Uncertainty Quantification

Tianrui Zhang

*College of Business, City University of Hong Kong, Hong Kong, China
zhang882-c@my.cityu.edu.hk*

Abstract. Traditional retail bank telemarketing for time deposit subscriptions has a low conversion rate, suffers from data leakage, and fails to account for individual uncertainty in point predictions. To address these issues, this paper constructs a risk-aware analytical system that integrates high-performance ensemble learning with non-parametric uncertainty quantification. First, a strict leakage-prevention protocol is introduced by excluding post-dialing variables, such as call duration, and chronological data splits are used to avoid look-ahead bias. Second, to address macroeconomic non-stationarity over time and space, the absolute account balance is converted into a quintile-based relative wealth ranking. Third, four machine learning models are tested, and XGBoost performs best, with an AUC-ROC of 0.9100, an F1-score of 0.8504, and a total accuracy of 85.49%. Additionally, a parallelized Bootstrap engine is employed to obtain individual predictive interval widths and, from the raw output probabilities, to construct an operational dual-tier decision system. The framework sets up micro-level risk alarms for uncertain predictions and macro-level manual overrides. Based on the above experiments, the system will reduce operating costs in the low-uncertainty marketing area by approximately 80%. Finally, conformal prediction is proposed as a future low-latency path for real-time deployment.

Keywords: Machine Learning, Time Deposit Prediction, Uncertainty Quantification, Bootstrap Resampling, Risk-Aware Decision Engine

1. Introduction

Time deposits are regarded as the primary channel for commercial banks to obtain low-cost, stable funds [1]. However, traditional mass telemarketing not only has a quite low conversion rate but also dramatically harms the customer experience due to frequent harassing calls. So there is an urgent need for retail banks to target potential customer segments using predictive models to increase return on marketing investment.

Although traditional machine learning models such as logistic regression and random forests have been widely applied to predicting customers' willingness to subscribe, two core limitations persist in actual business deployment.

First, point predictions hide uncertainty. Existing models only produce a fixed probability, such as a 75% purchase probability, which corresponds to a deterministic point prediction. According to [2], traditional black-box machine learning models tend to exhibit severe overconfidence when

financial texts or behavioral features fluctuate drastically. Without adequate risk awareness, blindly relying on such posterior probabilities with low confidence could lead to substantial decision bias in banks' automated marketing systems [2]. If banks blindly rely on these low-confidence posterior probabilities, significant risks would arise for automated marketing decisions. Especially during periods of market volatility, automatic decisions without uncertainty calibration would bring huge waste of marketing budgets and potential operational risks. Therefore, the method to overcome the blind spots of point prediction for individual prediction has been the most difficult problem that needs to be solved.

Second, absolute values violate comparability [1]. Many studies directly serve the "Absolute Balance" of customer accounts into models as a continuous variable [1]. Nevertheless, absolute monetary values are entirely incomparable across time and groups because of the differences in residents' wealth driven by inflation and gaps like affluent urban outlets and rural branches [3]. At the same time, absolute balance would lose its effect in a spatial perspective due to the big gap between wealthy urban branches and rural sub-branches [4]. Direct adoption of this indicator hinders the model from capturing nonlinear threshold effects [4].

To make up for the gap in the previous methodology, this research is based on a complete UCL (University of California, Irvine) bank marketing database and establishes an honest and regular analytical framework, which could quantify and predict risks. The core contributions of this article are reflected mainly in the following three dimensions.

First, for data leakage control, a feature removal protocol is strictly implemented before model training, deleting the confounding duration (call length) feature in traditional studies, which essentially avoids Look-ahead Bias to make sure the core model can operate independently as a realistic prediction tool before dialing [5].

Then, for financial engineering, absolute balances would be innovatively transformed into quintile relative rankings (balance relative) in this research, and the problem of incomparability across time and space could be solved.

In terms of decision optimization, Bootstrap is introduced to quantify the individual uncertainty, and a dynamic marketing strategy combined with automated AI and manual intervention would be mentioned in this study.

2. Data description and preprocessing

2.1. Dataset baseline and strict cleaning protocol

The study is experientially based on the UCI Bank Marketing dataset, which includes 45,211 real historical observation samples from telemarketing campaigns [1]. The target binary variable exhibits severe class imbalance; only 11.7% (5,289 entries) represent successful fixed-deposit subscription accounts ($Y=1$) among all samples. To eliminate model bias induced by such imbalance, a balanced class-weight penalty matrix is introduced during the training phase of all candidate classifiers [6].

During the baseline cleaning phase, this research employs a strict leakage-prevention protocol. The duration feature in the original database was removed entirely because this variable is a typical post-hoc variable, and a serious Look-ahead bias would be introduced if it remained, leading to an apparent effectiveness before actual outbound calling. Simultaneously, the repeated macroeconomic indicators nr. employed are also deleted to eliminate multicollinearity. Additionally, to simulate a real-world industrial production environment, random splitting is rejected, which would break temporal independence, while a strict Chronological Split was implemented: The raw dataset is sorted chronologically by marketing date: the first 80% of the time-series records (May 2008 – May

2010) form the training and validation set, while the latter 20% of the data (June 2010 – November 2010) is reserved as the test set. This configuration ensures that model validation remains fully independent of unseen future market conditions [5].

2.2. Quantile stabilization for non-stationarity

Directly feeding the raw continuous numerical feature balance, like the absolute account balance, into the model would lead to serious Non-stationarity. Over many years, absolute monetary values are highly sensitive to inflation, business-cycle fluctuations, and interest-rate adjustments. The nominal balance of € 50,000 in 2008 represents a significantly different customer wealth tier [3, 4]. To remove the nominal balance, this study would transform raw absolute balances into relative financial ranking variables based on local quintiles, denoted as B_{relative} . The mathematical transformation formula below maps the absolute balance x_i , within a rolling time window to its corresponding sample quantile bin:

$$B_{\text{relative}}(x_i) = k \quad \text{if} \quad Q_{k-1}(X) \leq x_i < Q_k(X) \quad (1)$$

where $B_{\text{relative}}(x_i)$ denotes the transformed relative financial ranking of customer x_i , which is their position in the distribution. The term x_i represents the absolute monetary account balance at the start of observation. Start equation k indicates the specific quintile interval index, with $k \in \{1, 2, 3, 4, 5\}$. Cap Q_k and Q_{k-1} denote the threshold values for the start equation k and $k-1$ quintiles of the financial distribution, respectively. As demonstrated in equation (1), the absolute account balance is mapped into its corresponding quintile interval. Finally, Cap X represents the local account balance distribution within the selected moving time window.

3. Methodology

3.1. Baseline classifier selection

To establish a stable experimental baseline, this study evaluates four distinct classification algorithms after preprocessing the data matrix.

Initially, a logistic regression model is used as a linear baseline, trained with maximum likelihood estimation and protected against overfitting via L2 regularization. Next, a decision tree serves as a non-linear baseline, which recursively partitions data based on information gain and Gini impurity. To further improve performance, a random forest consisting of 100 decision trees is employed, reducing variance through bootstrap sampling and random feature selection [7]. Finally, XGBoost, a popular boosting technique, is applied to optimize a regularized loss function by iteratively correcting errors and utilizing a second-order Taylor expansion of the loss [8].

3.2. Parallelized bootstrap engine

The traditional sort system could only generate deterministic point estimates. To quantify predictive risk without rigid Gaussian distribution assumptions, a non-parametric Bootstrap resampling framework is combined. Let D denote the original training set with sample size N . The engine runs $B=100$ independent iterations. In each iteration indexed by b , a bootstrap sample D_b^* of identical size N is constructed via uniform sampling with replacement from D . Although the accuracy of point prediction of XGBoost is the highest, its sequential boosting mechanism results in extremely

slow computation during large-scale iterative resampling, which creates severe computational latency bottlenecks for practical deployment in banking systems [9]. To increase the feasibility of deployment in financial institutions and satisfy the demand of a low-latency concurrent environment, a parallelized hybrid architecture is incorporated: arranging random forest, which natively supports parallel computing, serves as the Proxy Engine, dedicated to running 100 Bootstrap simulation loops. This configuration leverages multi-core CPUs to process resampling in parallel, circumventing the sequential speed bottleneck inherent to boosting models. For any newly input customer feature vector x_{new} , the engine outputs an empirical distribution consisting of 100 subscription probabilities:

$$\hat{P}_b = f_b^*(x_{\text{new}}), \quad b \in \{1, 2, \dots, 100\} \quad (2)$$

Where $\hat{P}_b$ denotes the probability of being a subscriber predicted by the b -th bootstrap estimator, f_b^* denotes the classification model trained on the b -th resampled dataset, x_{new} is the multidimensional feature vector of the newly targeted customer, and b is the bootstrap iteration index ranging from 1 to 100.

Using the empirical distribution of $\{\hat{P}_b\}_{b=1}^{100}$, the uncertainty interval width $W(x_{\text{new}})$ is then computed as follows:

$$W(x_{\text{new}}) = \text{Percentile}(\hat{P}, 97.5) - \text{Percentile}(\hat{P}, 2.5) \quad (3)$$

Based on the dynamic configuration in equation (3), the system automatically issues a micro-level Risk Alert. Here, $W(x_{\text{new}})$ represents the confidence interval size for the new observation x_{new} . The value $\hat{y}_{\text{upper}}(x_{\text{new}})$ denotes the upper bound of the predicted probability distribution (the 97.5th percentile), while $\hat{y}_{\text{lower}}(x_{\text{new}})$ indicates the lower bound (the 2.5th percentile) of this distribution for the customer observation.

4. Results and discussion

4.1. Model performance and feature importance evaluation

All candidate models are comprehensively evaluated on an independent chronological test set, and the key performance metrics are presented in Table 1. The results of the experiment show that the XGBoost classifier depicts the strongest discriminative ability, achieving the highest AUC-ROC (0.9100) and balanced F1-score (0.8504) [7, 8]. The competitive F1-scores of all ensemble models confirm that the class weight scheme introduced during the data preparation phase successfully addresses the inherent minority class imbalance issue within the dataset [6].

Table 1. Model performance summary matrix

Model	Accuracy	Precision	Recall	F1-score	AUC-ROC
Logistic regression	0.7973	0.7979	0.7662	0.7817	0.8500
Decision tree	—	—	—	—	0.7600
Random forest	0.8489	0.8520	0.8645	0.8443	0.8800

Table 1. (continued)

XGBoost	0.8549	0.8314	0.8702	0.8504	0.9100
---------	--------	--------	--------	--------	--------

An analysis of feature importance in the champion XGBoost model shows that the main predictor for classification has a relative importance score around 0.15. It is followed by the self-developed, very low relative balance feature (balance_relative_Very Low) and historical contact days (pdays). This empirical evidence supports the hypothesis that time-normalized relative financial rankings provide better and more consistent generalization performance than raw absolute monetary values [3].

4.2. Risk-aware decision framework and business value

Banks can transform the interval width $W(x_{\text{new}})$ output from the parallelised Bootstrap engine into a two-tier risk-aware operating model for translating marketing decisions.

An alert for risk will be triggered at the individual level. When the gap between two marginal customers is very small (e.g., 0.04), it indicates strong model agreement and automatically triggers an outbound call. Conversely, if Customer B has a large interval width (e.g., 0.42), it indicates considerable statistical noise, so they are excluded from the outbound list and immediately flagged as a high-risk case. Thus, some marketing resistance will be avoided [2].

At the macro level, the system has a time-based Manual Override function. Interval width is at its maximum in May and November, corresponding to substantial changes in the economy at those times, and is generally over 0.30. Therefore, the AI outbound calling script will be suspended during periods of high volatility. At the same time, the leads will be assigned to senior, experienced private bankers for top-level support. This approach is more economical to operate and will be relatively inexpensive. For campaigns in low-uncertainty areas, the AI's predictions are reliable, and calls can be fully automated to reduce outbound labor costs by as much as 80% in these cases. Therefore, resources and skilled personnel can be reasonably reallocated to high-value groups with higher uncertainty to optimize resource allocation and risk control.

4.3. Research limitations and future outlook

Although the framework in this study has good discriminatory ability and commercial prospects, it also has some shortcomings that will be examined in future work.

First, in terms of calculation efficiency, the 100-layer Bootstrap introduces considerable latency and may cause throughput problems in ultra-large-scale, real-time, and high-concurrency environments. Second, in terms of datasets and features, empirical validation is based on datasets with relatively static demographic and historical reach-related features [1]. Because there are no high-frequency, dynamic transaction data or real-time online channel usage, such as browsing history from mobile banking apps, such data cannot be leveraged to identify an abrupt change in a customer's buying behavior. Finally, in terms of application scenarios, the current decision-making engine is limited to outbound telemarketing of fixed-deposit products and has not been fully integrated into the omnichannel retail banking system.

The two main avenues for future research will be presented below. Conformal Prediction will be selected as a low-latency algorithmic alternative [10]. This method does not assume a particular data distribution; therefore, it provides full-coverage prediction regions and is computationally inexpensive for a real-time system. Dynamically changing behavioral attributes will also be added at the application level. A risk-adaptive decision-making system will also be introduced to span all

links of an omnichannel model for precise deposit marketing, including cross-selling and recommendation workflows across all bank products.

5. Conclusion

This paper presents a system that integrates high-performance classification algorithms with nonparametric uncertainty quantification to address several issues in prior work, including data leakage, macroeconomic non-stationarity, and the lack of risk awareness in traditional point predictions for retail bank time deposit subscription forecasts. It does not require adding variable-duration post hoc for leakage prevention and uses a quintile-based relative financial ranking transformation to improve the model's robustness and generalization across economic cycles while reducing look-ahead bias. Based on the experimental results, the proposed architecture uses XGBoost as the main classifier and a parallel-running random forest surrogate engine to achieve an accuracy of 85.49% and an AUC-ROC of 0.9100, without the latency issues caused by large-scale resampling in multi-core parallel computing.

This paper has successfully realized the concept of abstract mathematical uncertainty, such as the width of a confidence interval, in practice through a mechanism of micro-level risk alerts and macro-level manual overrides. The new decision-making system will be applied to all risks in the precision marketing of retail banking to improve overall operating efficiency. Based on the above empirical data, this system can reduce labor operating expenses for low-risk or uncertain customer segments by as much as 80 percent and improve the efficiency of labor resource utilization.

This study has shown that a feasible level of discrimination accuracy and commercial applicability can be achieved; however, some computational challenges remain. Nested Bootstrap loops have a high computational cost in large-scale, real-time, concurrent systems. Therefore, the following research will focus on combining Conformal Prediction to achieve prediction regions with strong mathematical coverage guarantees at almost no extra computation and without distributional assumptions. This approach can serve as a low-latency alternative to the current model, increase the risk-adaptive control function of the intelligent banking system, and maintain high-frequency, real-time response capabilities.

References

- [1] Moro, S., Cortez, P., & Rita, P. (2014). A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*, 62, 22-31. <https://doi.org/10.1016/j.dss.2014.03.001>
- [2] Subrahmanian, V. S., & Kumar, S. (2023). Calibrated uncertainty quantification in deep and ensemble learning for financial risk. *IEEE Transactions on Knowledge and Data Engineering*, 35(4), 3412-3425.
- [3] Di Francesco, M., & Gupta, R. (2024). Handling non-stationarity in macroeconomic time series features for banking risk models. *Journal of Financial Econometrics*, 22(2), 145-168.
- [4] Standard, A., & Zhou, Y. (2022). Spatial heteroskedasticity and outlet variance in retail banking networks. *Journal of Banking & Finance*, 138, 106-120.
- [5] Kaufman, S., Rosset, S., Perlich, C., & Stitelman, O. (2012). Leakage in data mining: Formulation, detection, and avoidance. *ACM TKDD*, 6(4), 1-21.
- [6] Alshboul, M., & Wang, J. (2023). Cost-sensitive ensemble learning for extreme class imbalance in financial marketing predictive analytics. *Expert Systems with Applications*, 213, 119-132.
- [7] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
- [8] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of ACM SIGKDD* (pp. 785-794).
- [9] Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC Press.
- [10] Angelopoulos, A. N., & Bates, S. (2021). A gentle introduction to conformal prediction. *Foundations and Trends in Machine Learning*, 14(5), 403-482.