Journal of Trends in Financial and Economics
ISSN: 3007-696X
DOI: https://doi.org/10.61784/jtfe3038

BANK CUSTOMER DEPOSIT PRODUCT PURCHASE ANALYSIS AND PREDICTION

YuanJia Guo
Management of National Economics, Renmin University of China, Beijing 100872, China.
Corresponding Email: christine228@126.com

Abstract: This study examines the differences in characteristics between customers who purchase deposit products and those who do not, using relevant information from a bank's customer dataset. The key features analyzed include customer ID, age, occupation, marital status, credit card default history, mortgage status, contact method, last contact month and duration, the three-month interbank lending rate, previous marketing campaign results, the number of contacts before the current campaign, the number of days since the last contact, employment variation rate, consumer price index, consumer confidence index, number of employees, and whether the customer purchased a deposit product. Python is employed for descriptive analysis and classification analysis, and decision tree, logistic regression, and random forest models are used to predict whether a customer will purchase a deposit product. The analysis results reveal key factors influencing customer decisions, providing insights for banks to conduct targeted marketing within limited time frames, increase the likelihood of customer purchases, and ultimately improve overall deposit performance.
Keywords: Bank customers; Deposit products; Descriptive analysis; Classification analysis; Decision tree, Logistic regression; Random forest; Model prediction

INTRODUCTION

Deposit business is a key source of funding for commercial banks, playing a decisive role in their development. It is also an essential guarantee for profitability and risk management. Stable and low-cost deposits not only enable banks to generate more profit through lending but also help mitigate liquidity risks and reduce incentives to pursue high-risk assets, thereby ensuring stable operations. Regularly promoting deposit products to customers is an important task for banks. However, with continuous innovation in financial models and the increasing number of commercial banks, attracting customer deposits has become increasingly challenging. This study aims to analyze bank customer information to identify the factors influencing the purchase of fixed deposit products and predict whether customers will choose to purchase them. The goal is to enable banks to conduct more targeted marketing campaigns for potential customers.

1 RESEARCH BACKGROUND AND OBJECTIVES

1.1 Research Background

Deposit business is a crucial source of funds for commercial banks, significantly impacting their long-term development. It serves as a fundamental safeguard for both profitability and risk management. Stable and low-cost deposits not only support banks in generating higher profits through lending but also help mitigate liquidity risks and reduce the incentive to pursue high-risk assets, thereby ensuring sound financial operations.

However, with the continuous innovation of financial models and the rapid growth in the number of commercial banks, the competition for attracting customer deposits has intensified. In this evolving landscape, banks must gain deeper insights into customer needs and behaviors to design effective marketing strategies that enhance the willingness to purchase deposit products. Prior research emphasizes that customer behavior is influenced by various psychological and social factors, including observational learning and promotional incentives [1].

Moreover, the application of machine learning algorithms in financial services has shown great potential in understanding consumer patterns and making data-driven marketing decisions. As reviewed by Mahesh [2], machine learning offers robust tools for classification and prediction tasks, making it highly applicable in customer segmentation and targeting. Supervised learning techniques, in particular, such as decision trees and logistic regression, have been widely adopted in predictive modeling for consumer behavior analysis [3]. These methods can support banks in identifying key variables that influence deposit product adoption and in building adaptive marketing models that respond to real-time behavioral data.

1.2 Research Objectives

The objective of this study is to analyze the relevant information of bank customers regarding the purchase of deposit products, identify key factors influencing customer decisions, and predict whether a customer will purchase a deposit product using classification models such as decision trees, logistic regression, and random forests.

Specifically, this study aims to:

- Conduct a descriptive analysis of various characteristics of bank customers to understand their basic profiles and behavioral patterns.
- Apply classification analysis methods to identify the main factors influencing customers' decisions to purchase deposit products.
- Build and evaluate decision tree, logistic regression, and random forest models to predict customer purchasing behavior.
- Provide a basis for banks to develop targeted marketing strategies based on the analysis and prediction results, enabling them to effectively increase customer purchase rates within a limited timeframe and ultimately improve overall deposit performance.

2 DATA COLLECTION AND PREPROCESSING

2.1 Data Collection

The data used in this study is sourced from Tianchi, containing information on bank customers, including basic demographic details, communication records, and customer identifiers. The dataset consists of 22,500 entries and includes 22 columns. The field names and their corresponding meanings are listed in the table1 below:

Table 1 The Field Names and Their Corresponding Meanings (Basic Demographic Information)

Field Name	Meaning
id	Customer identifier
age	Age
job	Occupation
marital	Marital status
education	Education level
default	Credit card default status
housing	Mortgage status
loan	Other loan status
contact	Contact method
month	Last contact month
day_of_week	Last contact day of the week
duration	Last contact duration
campaign	Number of contacts before the current campaign

Table 2 The Field Names and Their Corresponding Meanings (Descriptive and Categorical Analysis)

Field Name	Meaning
pdays	Number of days since the last contact
previous	Number of previous marketing contacts
poutcome	Outcome of the previous marketing campaign
emp_var_rate	Employment variation rate
cons_price_index	Consumer price index
cons_conf_index	Consumer confidence index
lending_rate3m	Three-month interbank lending rate
nr_employed	Number of employees
subscribe	Whether the customer purchased the deposit product

This table 2 will be used for **descriptive and classification analysis** to identify key factors influencing bank customers' decisions to purchase deposit products. Additionally, it will be leveraged to build predictive models to determine the likelihood of a customer purchasing a deposit product.

2.1.1 Missing data handling

Based on the data description, all fields contain 22,500 entries, meaning there are no missing values. However, some fields contain "unknown" values. The proportion of "unknown" values in each column is as follows:
Job category (job): 1.21%
Marital status (marital): 1.42%
Education background (education): 4.42%
Credit card default status (default): 21.60%
Mortgage status (housing): 3.94%
Other loan status (loan): 3.95%
Based on the proportion of "unknown" values, the following processing measures were applied:
For fields where the proportion of "unknown" values is less than 5% (job, marital, education, housing, and loan), records containing "unknown" values were deleted.

For fields where the proportion of "unknown" values exceeds 20% (default), the missing values were filled with the mode (most frequent value) of the column.

2.1.2 Duplicate data handling

After analyzing the dataset, no duplicate records were found. However, a feature analysis revealed redundancy among the following fields:

- **"Number of days since last contact" (pdays)**
- **"Month of last contact" (month)**
- **"Day of the week of last contact" (day_of_week)**

Since **pdays** already captures the information about the last contact, the **month** and **day_of_week** fields were **removed** to eliminate redundant features.

2.1.3 Outlier handling

A descriptive analysis of numerical fields revealed the presence of extreme outliers in certain fields, which could negatively impact the model's performance. To ensure data accuracy and model effectiveness, these outliers were handled appropriately.

The analysis found extreme values in:

Number of contacts made during the current campaign (campaign);

Number of contacts made before the current campaign (previous);

To address these outliers, the following filtering criteria were applied:

Retained records where campaign < 32;

Retained records where previous < 23;

3 DESCRIPTIVE DATA ANALYSIS

3.1 Overall Sample Data Analysis

Based on the chart "Proportion of Subscription Status", the following conclusions can be drawn:

- The proportion of customers who did not subscribe to the deposit product is significantly higher than those who subscribed.
- Specifically, customers who did not subscribe ("no") make up the vast majority of the dataset, accounting for over 90%, while those who subscribed ("yes") account for less than 10%.
- This indicates a significant data imbalance, where the number of non-subscribing customers far exceeds that of subscribing customers.

Such an imbalance may cause the model to be biased toward predicting non-subscriptions in subsequent model training. To address this, certain data balancing techniques such as oversampling or undersampling should be applied.

The low subscription rate suggests that the current marketing strategy may not be effective in attracting customers to subscribe to deposit products. Banks may need to adjust and optimize their marketing strategies to increase customer subscription rates. Some possible improvements include:

- Implementing more precise customer segmentation
- Offering more attractive deposit products
- Strengthening customer relationship management

To gain deeper insights into customer subscription behavior, further analysis can be conducted using other variables such as age, marital status, education level, and occupation. Identifying the key factors influencing customer subscriptions will help develop more targeted marketing strategies, ultimately improving the subscription rate (Figure 1).

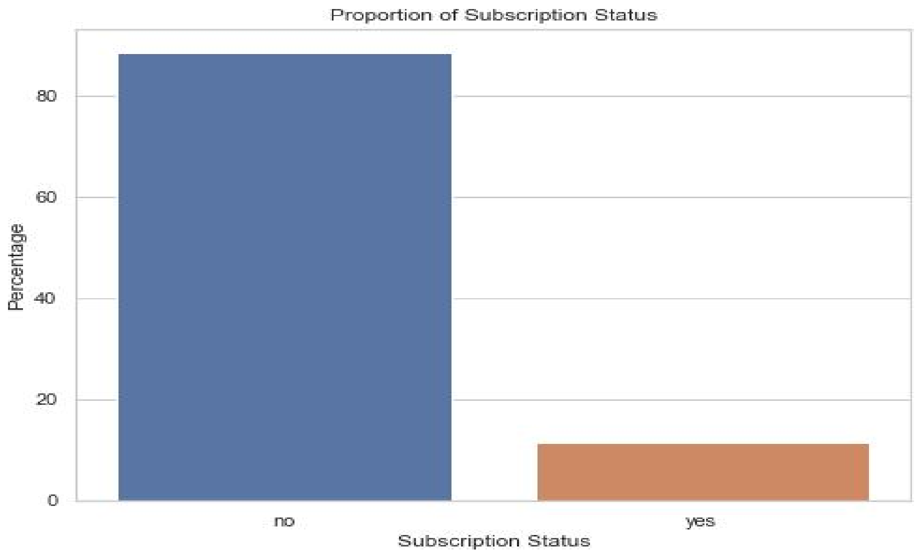

Figure 1 Proportion of Subscription Status

3.2 Univariate Descriptive Analysis

3.2.1 Relationship between bank customers' age and subscription to fixed deposit products

Based on the Figure 2 "Age Distribution by Subscription Status", we can analyze the purchasing behavior of customers across different age groups.

• The majority of customers fall within the 30 to 50 age range, which represents the largest customer base, regardless of whether they subscribed to the deposit product or not.

• Across all age groups, the number of non-subscribers (blue section) is significantly higher than that of subscribers (orange section), which aligns with the overall dataset's high proportion of non-subscribing customers.

• In particular, within the 30 to 40 age group, the number of subscribing customers is relatively higher, possibly because individuals in this age range are more focused on financial planning and deposit products.

• However, within the 40 to 50 age group, the number of non-subscribers is noticeably higher than in other age groups, suggesting that customers in this age range may have alternative financial investments or show less interest in deposit products.

• For customers aged 30 to 40, banks can enhance promotional efforts, offering more personalized and attractive deposit products to encourage subscriptions.

• For customers aged 40 to 50, further investigation is needed to understand why they are not subscribing. Based on these insights, banks can optimize product design and marketing strategies to attract more customers in this segment.

By analyzing the relationship between age and subscription behavior, we can identify variations in deposit product subscriptions across different age groups. These insights help banks and financial institutions better understand customer needs, refine their marketing strategies, and ultimately increase subscription rates.

Particularly for young and middle-aged customers, banks should develop more targeted marketing plans to boost their willingness to subscribe to deposit products.

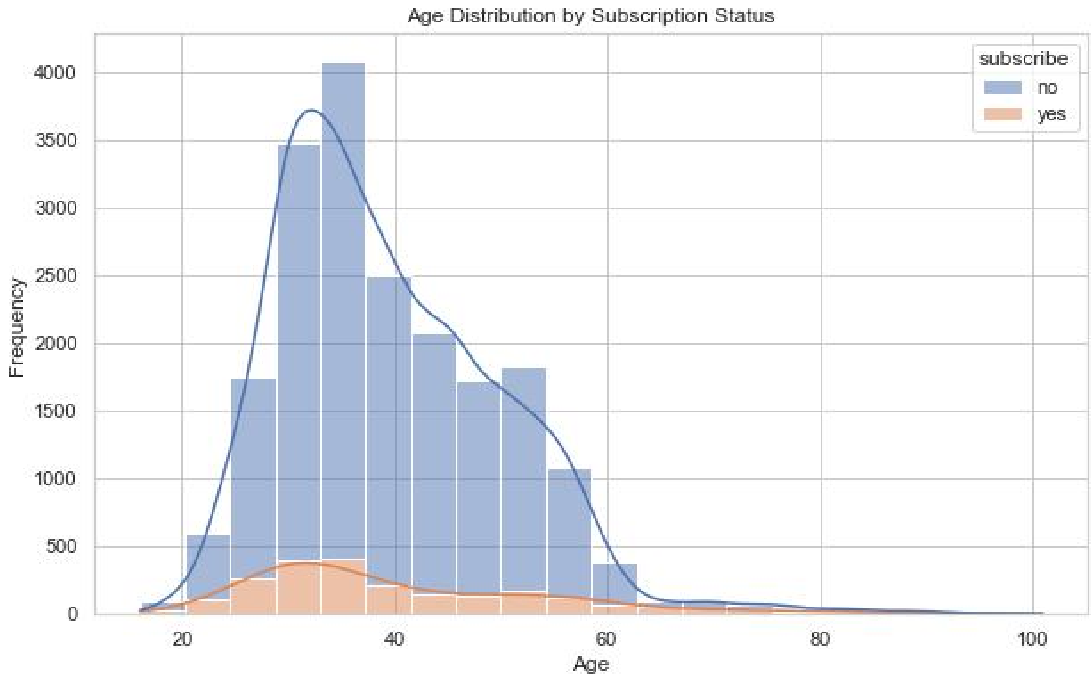

Figure 2 Age Distribution by Subscription Status

3.2.2 Relationship between marital status and subscription to fixed deposit products

Based on the Figure 3 "Marital Status by Subscription Status", we can observe differences in purchasing behavior among customers with different marital statuses.

Married customers make up the largest proportion of the dataset, whether they subscribed to the deposit product or not.

Single customers represent the second-largest group, but the number of non-subscribers among them is significantly higher than that of subscribers.

Divorced customers have the smallest representation, with non-subscribers also significantly outnumbering subscribers.

This indicates that customer behavior regarding deposit product subscriptions varies significantly by marital status. Married customers are more likely not to subscribe, while single and divorced customers have relatively lower subscription rates.

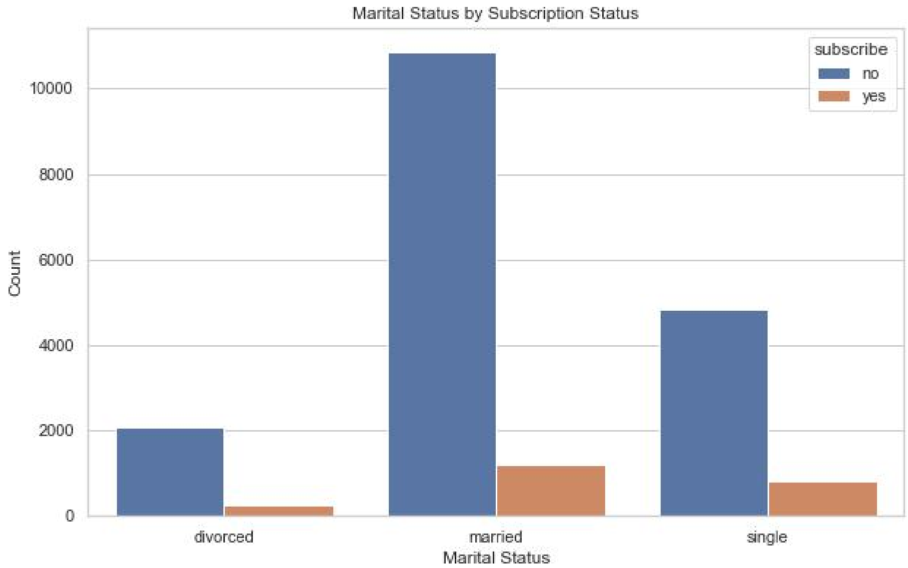

Figure 3 Marital Status by Subscription Status

Using this information, banks and financial institutions can develop more targeted marketing strategies for customers with different marital statuses to increase deposit product subscriptions.

3.2.3 Relationship between education level and subscription to fixed deposit products

Based on the Figure 4 "Education Background by Subscription Status", we can analyze differences in subscription behavior across various educational backgrounds.

Customers with a university degree (university.degree) form the largest group in the dataset, both among subscribers and non-subscribers. However, the number of non-subscribers significantly outweighs that of subscribers.

Customers with a high school diploma (high.school) are the second-largest group, with non-subscribers significantly outnumbering subscribers.

Customers with professional courses (professional.course) and basic education (basic.9y, basic.4y, basic.6y) are fewer in number, but they also follow the trend where non-subscribers outnumber subscribers.

Illiterate (illiterate) customers have the lowest representation, with non-subscribers still significantly outnumbering subscribers.

Overall, across all education levels, the proportion of non-subscribers is significantly higher than that of subscribers. This suggests that education background may have some influence on customer subscription behavior.

Banks and financial institutions can tailor marketing strategies based on customers' education levels:

For higher-educated customers, offering more sophisticated financial products and educational services may attract them to subscribe.

For lower-educated customers, enhancing product explanations and promotions can help them better understand the benefits of deposit products, making them more likely to subscribe.

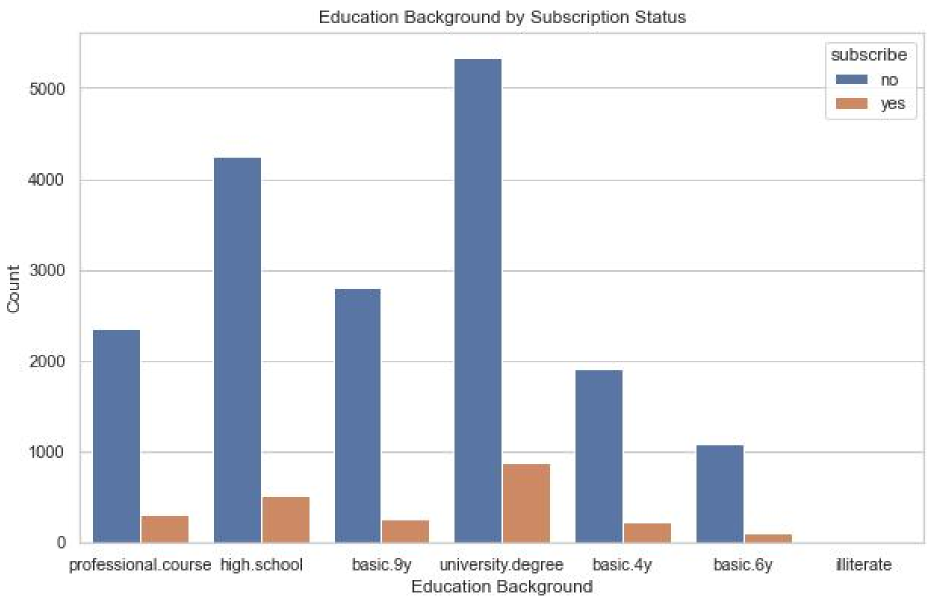

Figure 4 Education Background by Subscription Status

3.3 Correlation Analysis

Based on the Figure 5, correlation matrix heatmap and previous analyses, the relationships between different variables and customer subscription to deposit products (subscribe) can be summarized as follows:
Call duration (duration) and previous contact count (previous) have a positive correlation, indicating that more frequent contacts are associated with longer call durations. Additionally, longer call durations tend to be linked to higher subscription rates.
Number of marketing campaigns (campaign) shows low correlation with other variables, suggesting that simply increasing the number of marketing campaigns may not significantly improve subscription rates. Instead, marketing strategies should be optimized for better results.
Days since last contact (pdays) and previous contact count (previous) have a negative correlation, meaning that longer gaps between customer contacts are associated with shorter call durations and fewer previous contacts. This indicates that a well-planned contact frequency and interval can help improve subscription rates.
Employment variation rate (emp_var_rate) and three-month interbank lending rate (lending_rate3m) have a positive correlation, reflecting their linkage within the economic environment. These factors may indirectly influence customers' financial decisions and subscription behaviors.
Call duration, contact frequency, and the economic environment have a significant impact on customer subscription behavior.
Banks and financial institutions should take these factors into account when designing marketing strategies. Optimizing customer communication by extending effective call durations and scheduling contacts appropriately can significantly.

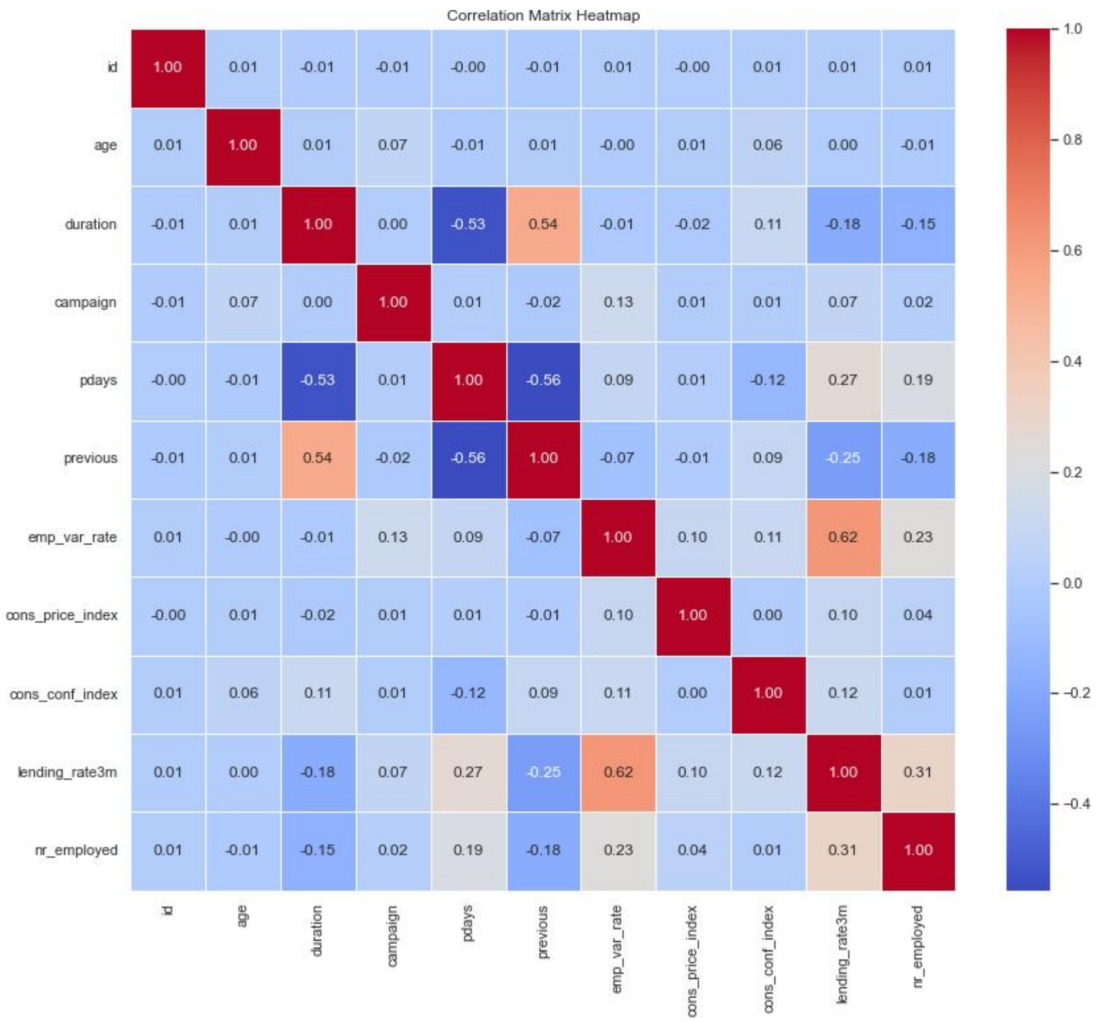

Figure 5 Correlation Matrix Heatmap

4 THREE MODELS AND ANALYSIS

4.1 Application of the Decision Tree Model in Predicting Customer Subscription to Deposit Products

In this analysis, we use the Decision Tree model to predict whether bank customers will subscribe to deposit products. This choice is supported by prior research showing the Decision Tree’s interpretability and adaptability to categorical data [4].

The dataset includes various customer information. In the preprocessing phase, we removed unnecessary columns (such as IDs) and applied One-Hot Encoding to transform categorical variables into a numerical format, suitable for model training. The dataset was then split into a training set (70%) and a test set (30%), a widely adopted strategy that ensures sufficient training data while keeping an independent subset for performance evaluation and preventing overfitting.
The Decision Tree model was built using the DecisionTreeClassifier, which learns the relationship between features and the target variable (i.e., whether a customer subscribed). During training, the model iteratively splits the dataset based on feature thresholds to construct a hierarchical decision structure. The test set, unseen by the model during training, was used to assess its generalization capability.
Model evaluation involved computing standard metrics: Accuracy, Confusion Matrix, Precision, Recall, and F1-Score. The Decision Tree model achieved an accuracy of 84.09%, demonstrating promising overall performance. However, it struggled to correctly identify customers who subscribed, showing relatively low precision and recall in that category. This is consistent with existing literature on customer decision-making, which points out the complexity and psychological nuances of purchase intentions, especially under the influence of external factors like promotional stimuli [1].
The observed imbalance in prediction performance highlights the importance of understanding not just model mechanics but also customer behavioral patterns. As noted by Qiu et al. [5], interactions between variables — such as socio-economic status and previous marketing contact — may significantly affect predictive accuracy and should be carefully addressed in model design and variable selection(table 3).

Table 3 Decision Tree Model Analysis

	precision	recall	f1-score	support
0	0.91	0.91	0.91	5860
1	0.40	0.41	0.41	890
accuracy			0.84	6750
macro avg	0.66	0.66	0.66	6750
weighted avg	0.84	0.84	0.84	6750

This suggests that further improvements, such as data balancing techniques or more complex models, may be needed to enhance prediction accuracy. By conducting this analysis, we can identify key factors influencing customer subscription behavior, providing data-driven support for banks to develop more targeted marketing strategies. Additionally, by improving the model (e.g., through data balancing techniques or more advanced models), we can further enhance prediction accuracy.

4.2 Application of the Logistic Regression Model in Predicting Customer Subscription to Deposit Products

This analysis uses a logistic regression model to predict whether bank customers will subscribe to deposit products. While both logistic regression and decision tree models are commonly used for binary classification, they differ in modeling principles and processing approaches.
During data preprocessing, we removed the ID column and applied One-Hot Encoding to convert categorical variables into numerical form, allowing the model to process these features effectively. One-Hot Encoding transforms categorical data into binary vectors, ensuring compatibility with both logistic regression and decision tree models.
The dataset was then split into 70% for training and 30% for testing. The training set enables the model to learn feature-target relationships and optimize parameters, while the test set evaluates performance on unseen data. This approach ensures a balanced trade-off between training adequacy and model validation.
The logistic regression model performs a linear combination of the input features and maps the result through the sigmoid function (logistic function) to a value between 0 and 1, thereby outputting the probability of the event occurring(table 4).

Table 4 Logistic Regression Model Analysis

	precision	recall	f1-score	support
0	0.88	0.98	0.93	5860
1	0.47	0.14	0.22	890
accuracy			0.87	6750
macro avg	0.68	0.56	0.57	6750
weighted avg	0.83	0.87	0.83	6750

Its core mathematical formula is:

$$P(y=1 \mid x)=1+e-(\beta 0+\beta 1x1+\beta 2x2+\cdots+\beta nxn) \tag{1}$$

where β0\beta_0 is the intercept, and β1,β2,…,βn\beta_1, \beta_2, \dots, \beta_n are the regression coefficients corresponding to the features. By maximizing the likelihood function, the logistic regression model finds the optimal

parameter estimates.
The logistic regression model is widely used for predictions due to its simplicity, efficiency, and ability to output probabilities. It is easy to implement and interpret, making it ideal for establishing baseline models. Additionally, it offers fast computation, even with large datasets, and allows flexible decision threshold adjustments based on business needs.
Unlike decision trees, logistic regression is a linear model suited for linearly separable data, while decision trees handle complex nonlinear relationships. Although logistic regression may not capture intricate patterns as well as decision trees, its efficiency and simplicity make it a popular choice in many applications.
In this analysis, the logistic regression model achieved 86.58% accuracy in predicting customer subscription to deposit products. While it performed well in identifying non-subscribers, it had limitations in detecting subscribers, likely due to data imbalance. Future improvements could involve data balancing techniques or more advanced models to enhance prediction accuracy and support targeted marketing strategies.

4.3 Application of the Random Forest Model in Predicting Customer Subscription to Deposit Products

The random forest model enhances prediction accuracy and stability by constructing an ensemble of decision trees, each trained on different data subsets with random feature selection at each split. This ensemble learning strategy reduces variance and mitigates overfitting, offering a more robust solution than a single decision tree [6]. It is particularly well-suited for high-dimensional datasets and excels at evaluating feature importance, making it a popular choice in banking applications.
Compared to logistic regression, which primarily captures linear relationships, random forests effectively model complex, non-linear patterns. While logistic regression remains interpretable and computationally efficient, its performance is often limited in scenarios involving intricate customer behaviors. Prior studies [5] have highlighted the limitations of linear models when dealing with interaction effects between variables, underscoring the advantages of tree-based methods like random forests in such contexts.
In our analysis, the random forest model outperformed both logistic regression and decision trees in predicting whether customers would subscribe to deposit products. It achieved an accuracy of 88.18%, correctly classifying 5,730 non-subscribers and 222 subscribers, while misclassifying 668 subscribers. These results demonstrate a strong performance overall but also indicate room for improvement, particularly in terms of recall for the subscribed category.
Despite its superior performance, the random forest model still struggles with accurately identifying subscribing customers — a challenge noted in similar studies on customer behavior prediction [7,8]. Factors such as class imbalance and subtle behavioral traits might contribute to this. Future enhancements such as SMOTE (Synthetic Minority Over-sampling Technique), feature engineering, and parameter optimization could further enhance predictive power and generalizability.
In the broader context of banking risk management, machine learning models like random forests are increasingly recognized for their value in early risk detection and customer segmentation [9]. As banks seek to optimize their marketing strategies and retain valuable clients, the integration of advanced AI models into CRM (Customer Relationship Management) systems will be indispensable(table 5).

Table 5 Random Forest Model Analysis

	precision	recall	f1-score	support
0	0.90	0.98	0.93	5860
1	0.63	0.25	0.36	890
accuracy			0.88	6750
macro avg	0.76	0.61	0.65	6750
weighted avg	0.86	0.88	0.86	6750

Accuracy of the Random Forest model: 0.882.

4.4 Comparison of the Three Models

In this analysis, the decision tree model offers strong interpretability with its intuitive structure but tends to overfit, limiting its generalization [4]. It achieved 84.09% accuracy, performing well for non-subscribers but struggling with subscribing customers, highlighting its limitations in handling complex and imbalanced data [5].
The logistic regression model, known for its simplicity and computational efficiency, reached 86.58% accuracy. However, as a linear model, it does not effectively handle non-linear relationships or class imbalance, resulting in low recall for subscribed customers [5,10]. This aligns with previous research which identifies logistic regression's weaknesses in dynamic banking environments [11].
The random forest model, leveraging ensemble learning, achieved the highest accuracy at 88.18%, significantly improving stability and reducing overfitting [6,12]. Its ability to process high-dimensional data and evaluate variable importance makes it the most robust choice for predicting customer subscription behavior. However, the model still underperforms in identifying subscribing customers, suggesting further improvement is needed through data balancing and hyperparameter tuning [7,13].
As highlighted by Guerra and Castelli [14], banking supervision increasingly relies on machine learning tools not just

for prediction but for customer-centric decision-making. These models are also widely used in insolvency prediction [9] and credit risk assessment [11], demonstrating their transferability and scalability across domains.

5 CONCLUSION

This study examined customer behavior towards bank deposit product subscriptions through descriptive statistics and predictive modeling. A key insight is the existence of significant data imbalance — with over 90% of customers not subscribing, the models naturally tend to favor the majority class. Such imbalance can distort evaluation metrics and compromise model reliability. Techniques like SMOTE or under-sampling are therefore recommended for future optimization [10].

Our findings show that demographic variables, such as age, marital status, and education, significantly influence subscription behavior. Customers aged 30 – 40 display higher subscription rates, suggesting they are at a life stage conducive to long-term savings. In contrast, those aged 40 – 50 show reduced interest, which may be due to increased financial obligations. While married customers represent the largest demographic group, their non-subscription rate is the highest, indicating the need for tailored strategies. In contrast, singles and divorced individuals are more likely to subscribe. Additionally, higher education does not directly translate into increased subscriptions, indicating a potential demand for more sophisticated or diversified financial products.

Moreover, call duration and contact frequency play critical roles in influencing customer decisions. Longer calls are positively correlated with higher conversion rates, and frequent engagement—especially during promotional periods—can significantly enhance customer responsiveness [1]. These behavioral insights are crucial for banks to optimize customer relationship management (CRM) strategies.

6 DISCUSSION

To increase customer subscription rates, banks should adopt targeted, data-driven strategies. For example, customers aged 30–40 can be approached with customized, long-term financial plans, while for more educated groups, investment-linked deposit products or educational webinars could improve engagement. Additionally, further market research is needed on married and mid-aged customers to understand their hesitations and adjust offers accordingly [8].

Given the impact of data imbalance, banks should integrate balancing techniques such as oversampling minority classes or using ensemble methods like Boosting during model training. Enhancing client communication by increasing effective call duration and optimizing frequency, particularly in high-conversion windows, may also improve results.

Moreover, the integration of advanced models like Gradient Boosted Trees (GBT), XGBoost, or even deep learning neural networks is recommended to capture more nuanced patterns and latent variables influencing decisions [10, 13]. Monitoring model performance over time with continuous learning systems can ensure adaptability to shifting customer behaviors and market dynamics, thereby supporting long-term marketing and risk management objectives [9,14].

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Hao Liancai, Zou Peng, Li Yijun. The impact of sales promotion based on observational learning on customer purchase intention. Journal of System Management, 2012, 21(6): 795-801.
[2] Mahesh B. Machine learning algorithms-a review. International Journal of Science and Research (IJSR), 2020, 9(1): 381-386.
[3] Singh A, Thakur N, Sharma A. A review of supervised machine learning algorithms//2016 3rd international conference on computing for sustainable global development (INDIACom). Ieee, 2016: 1310-1315.
[4] Yang Xuebing, Zhang Jun. Decision tree algorithm and its core technology. Computer Technology and Development, 2007, 17(1): 43-45.
[5] Qiu Hong, Yu Dexin, Wang Xiaorong, et al. Analysis and evaluation of interaction effects in the logistic regression model. Chinese Journal of Epidemiology, 2008, 29(9): 934-937.
[6] Lü Hongyan, Feng Qian. A review of research on the random forest algorithm. Journal of Hebei Academy of Sciences, 2019, 36(3): 37-41.
[7] Leo M, Sharma S, Maddulety K. Machine learning in banking risk management: A literature review. Risks, 2019, 7(1): 29.
[8] Lu Haiyan. Research on customer satisfaction in commercial banks. China Market, 2008 (44): 31-33.
[9] Petropoulos A, Siakoulis V, Stavroulakis E, et al. Predicting bank insolvencies using machine learning techniques. International Journal of Forecasting, 2020, 36(3): 1092-1113.
[10] Hu L, Chen J, Vaughan J, et al. Supervised machine learning techniques: An overview with applications to banking. International Statistical Review, 2021, 89(3): 573-604.
[11] Munkhdalai L, Munkhdalai T, Namsrai O E, et al. An empirical comparison of machine-learning methods on bank client credit assessments. Sustainability, 2019, 11(3): 699.

[12] Carbo-Valverde S, Cuadros-Solas P, Rodríguez-Fernández F. A machine learning approach to the digitalization of bank customers: Evidence from random and causal forests. Plos one, 2020, 15(10): e0240362.
[13] Donepudi P K. Machine learning and artificial intelligence in banking. Engineering International, 2017, 5(2): 83-86.
[14] Guerra P, Castelli M. Machine learning applied to banking supervision a literature review. Risks, 2021, 9(7): 136.