Variables that Influence Customers' Choice to Purchase Term Deposits

Yuanzheng Yang[1,a,*]

[1]Department of Statistics, China University of Petroleum-Beijing at Karamay, Xinjiang, 834000, China
a. 2022016621@st.cupk.edu.cn
*corresponding author

Abstract: This paper focuses on how to use data mining technology to improve the success rate of bank periodical deposits subscriptions, under the premise of considering the data imbalance problem, processing customer data and making targeted subscription recommendations according to the final classification results to reduce the cost of the bank, ultimately to increase the benefit of the bank. This paper investigates the potential factors that can influence a customer's decision to subscribe to a time deposit at a bank. To do this, a decision tree algorithm was used to analyze 21 different factors that may have an impact on this decision. Through this analysis, a decision tree model of influencing factors of bank customer subscription to time deposits was constructed. This model provides useful insights into how customer behavior may be influenced by various factors, and can be used to inform decisions around marketing and product development. Research has revealed that there are three key factors which have a major impact on customers subscribing to time deposits: the employee index, the duration, and the month. By taking these three factors into account, banks can optimize the range of customers they market time deposits to, and increase their efficiency in this regard. This can lead to an increased number of subscriptions, and more successful campaigns due to the targeted approach. Additionally, banks can use this research to identify areas of improvement and focus on aspects that can further increase the success rate of their campaigns.

Keywords: data mining, customer segmentation, decision tree, statistical decision making

1. Introduction

In order to remain competitive in China's banking industry, banks have been forced to look for ways to improve their customer management efficiency. This has necessitated the establishment of a customer classification system that is able to accurately and effectively sort customers according to their needs and preferences. This system enables banks to better understand their customers, tailor services to their needs, and provide more personalized experiences that lead to increased customer loyalty and satisfaction. Additionally, the system allows banks to identify areas of potential growth and develop strategies to tap into them. By using customer classification, banks can better manage their customers and ensure their long-term success in a competitive market.

A study aimed to investigate the effects of bank-specific factors on the total deposits of banks in Ghana from 2008 to 2017 using a Random Effects Model. The results of the analysis showed that

under the control of macroeconomic factors, the profitability, bank size and liquidity of a bank were key factors in determining the deposits. In addition, the macroeconomic instability caused by inflation had a notably negative impact on the deposits. Surprisingly, the increase in bank capital adequacy did not translate into an increase in deposits [1]. Utilized Indian banking data from 1998-2014 to examine whether expanding non-interest income activities can significantly improve the profitability of banks. The results show that a higher proportion of non-interest income can significantly increase a bank's non-deposit assets, with foreign banks and those with lower asset quality benefiting more from diversifying their revenue activities [2].

For the resident savings deposits, the macroeconomic factors such as market interest rates, gross domestic product, consumer price level, urban and rural residents income level are mainly used as explanatory variables, and the urban and rural residents savings income of the city commercial bank is taken as the explained variable to establish a linear regression model and a logarithmic linear regression model for analysis and research [3]. For a time series data of the total deposit income variable of a certain domestic financial institution, a combined prediction model based on the grey model GM(1,1), the third-order exponential smoothing model and the BP neural network model was established [4].

The logistic regression model can be used to identify the relationships between the different factors related to deposits of commercial banks, such as the size of the bank, the interest rate of the deposit, the risk appetite of the bank, and other external factors. The model can also be used to forecast the impact of any changes in the factors on the deposits of the bank. Additionally, the model can be used to identify any hidden patterns or trends in the data that can be used to make better decisions about the management of the bank's deposits [5]. Reviewing the econometric methods of the bank customer dynamic panel data model and illustrating the use of these methods, the author focuses on the GMM estimation method of the single equation model of the autoregressive dynamic panel with non-strict exogenous explanatory variables in the case of small T and large sample size N (which is applicable to the microeconomic data faced by this paper) [6].

In addition to convenience, trust and confidence, personal preferences, and product features, customer choice of bank deposits can also be influenced by the availability of various banking services, the bank's reputation and customer service, the terms and conditions of the deposit products, the interest rates, the fees, and other factors. Furthermore, customers may also be swayed by the bank's digital capabilities, such as mobile banking apps, online banking services, and other technological innovations [7]. Additionally, customer experience and satisfaction, as well as competitive pricing, have been identified as important factors influencing customer choice. Moreover, customer loyalty and retention are also important considerations, as customers are more likely to return to a company that they are satisfied with in terms of service or product quality. Additionally, customer feedback and reviews can also prove to be invaluable in helping a company understand how customers perceive their services or products, and make changes or improvements accordingly. Finally, the availability of new technologies and innovative products can also be a major factor that affects customer choice [8].

Found that convenience, trust and confidence, and product features were the most important factors influencing customer choice of bank deposits. Other factors such as customer service, competitive pricing, and bank reputation also played a role in customer decision-making. Additionally, the study found that customers were more likely to choose a bank deposit if they felt the bank was reputable, provided competitive pricing, and had good customer service. Furthermore, customers also tended to be more likely to choose bank deposits if they felt the bank was convenient, trustworthy, and had features that were attractive to them [9]. Trust and confidence are two key factors when customers are considering bank deposits. Customers want to feel secure and confident in their chosen bank, and that the bank is trustworthy and reliable. Convenience is also an important

consideration, as customers want to be able to access their bank deposits quickly and easily. Technology plays an important role in customer choice as well, with customers often preferring banks that have the latest technology available. Finally, product features are also important, as customers want to make sure that the bank deposits they choose offer the features and benefits that they need [10]. Research has found that trust and confidence, convenience, technology, and product features are the most important factors influencing customer choice when it comes to selecting bank deposits. Customers want to be sure that their deposits are secure, that the bank offers the features they need, and that the technology used is reliable and up-to-date. Additionally, customers want to ensure that the product features match their preferences, such as interest rates, fees, and other terms and conditions.

In this paper, we take the customer data provided by Portuguese banking institutions as the research object to investigate the factors influencing the subscription of fixed-term deposits by bank customers, so that banks can more easily identify customers who are likely to make such deposits and provide them with better services. This problem is essentially a classification problem, meaning that we need to identify profitable and high-quality target customers based on multiple research factors. The data provided includes both discrete and continuous variables, binary and multi-value variables, and so on. Traditional modeling and prediction methods, as well as classification methods, are not applicable in this case. For example, linear regression models do not meet the model assumptions; non-parametric regression methods will face the problem of high dimensionality; neural network models are too complex and have poor computing efficiency; discriminant analysis methods are difficult to determine appropriate and complex data type appropriate distance functions to construct discriminant criteria. Thus, in this paper we have adopted decision tree methods due to their weak assumptions on data types, high computing efficiency, and effective handling of discrete or complex classification data. We believe that the results of this research will prove useful for Portuguese banking institutions to better understand the factors influencing the subscription of fixed-term deposits by bank customers, and to improve their services to these customers.

2. Methods

2.1. Data and Variables

This paper investigates the various factors that influence customer subscription to regular deposits of a Portuguese bank institution. To better understand the customer's decision-making process, the decision tree method is employed in this study. This method is chosen for its flexibility in dealing with discrete and complex data types, its efficiency in calculations, and its effectiveness in processing classification data. By analyzing the customer data from the bank, this paper tries to identify and quantify the factors that affect customer subscription to regular deposits, so that the bank can better understand and predict customer behavior.

Using data collected from customers of a bank, predictive models can be used to identify whether customers are likely to subscribe to term deposits and classify them into different categories. As seen in Table 1, 21 attributes are used to determine this prediction, which can be divided into four categories: customer situation, bank relationship, contact bank activity status, and economic and social environment status. Customer situation consists of age, gender, employment status, marital status, education level, house loan, and personal loan. Bank relationship consists of credit arrears status and account balance. Contact bank activity status consists of method of contact, contact date of the previous month, contact month of the previous year, duration of contact, number of contacts during this activity, number of contacts before, result of previous activities, and number of days since last contact. Economic and social environment status consists of quarterly index of the number of

employees, employment change rate, consumer confidence index, and consumer price index. By utilizing these attributes, predictive models can be created to accurately identify whether customers will subscribe to term deposits and categorize them accordingly. This allows banks to more accurately target potential customers and target their marketing campaigns more effectively.

Table 1: The 21 attributes indices.

Sequence number	Attribute name
X1	age
X2	employment status
X3	marital status
X4	level of education
X5	mortgage loans
X6	personal loans
X7	Credit arrears status
X8	account balance
X9	subscription fixed deposit situation
X10	the ways of contact
X11	the date of contact in recent months
X12	the months of contact in recent years
X13	the duration of contact
X14	the number of contacts during this activity period
X15	the number of contacts before
X16	the results of the previous activities
X17	the number of days since the last contact
X18	employment change rate
X19	consumer confidence index
X20	consumer price index (CPI)
X21	the quarterly indicator of employee numbers

2.2. Data Processing and Conversion

When analyzing data, it is important to consider the accuracy of the results, especially when the data being analyzed does not accurately reflect real-life situations. To ensure that our research results are valid and applicable, we need to take into account the underlying relationships between data points. To do this, we can use Bayesian principles to add prior information to the target variable before conducting data extraction. This will help to reduce the potential bias caused by the data sample and make the results more accurate and useful for practical applications.

In many classical fitting models, it is assumed that the variables follow a normal distribution to get the best fitting effect and analytical properties. To improve the accuracy of modeling and estimation efficiency, it is often necessary to merge categories when the number of variable categories is large and the observed samples are concentrated in a small number of categories. When dealing with scattered continuous variable data, it can be transformed through a function transformation to make its distribution closer to the normal assumption. Similarly, for multiple classification variables,

they can be sorted and merged to improve the accuracy of the model. This process is often used to optimize the data before fitting the model in order to get the best results.

3. Results and Discussion

3.1. Results

The binary data utilized for the logistic regression study also showed no significant effects of customer age, income level, education level, occupation, or marital status on time deposit subscription. This indicates that these factors do not have a major influence on the customer's decision to subscribe to a time deposit, suggesting that the customer's gender is not a major factor in this decision. Additionally, the decision tree model used in the study demonstrated that customer gender was not a major factor in time deposit subscription, as seen in Table 2. This further supports the notion that customer gender is not a major factor in this decision, and suggests that other factors such as customer age, income level, education level, occupation, or marital status may have a more significant impact on this decision.

Table 2: Rules for the decision tree model.

Sequence number	The content of the rule
X1	If the number of employees is 5161 or lower, the subscription rate was 72.8%.
X2	If the number of employees in a company exceeds 5,161, the probability of customers subscribing to time deposits is 26.9%
X3	If the number of employees is less than or equal to 5161 and the duration is greater than 373.6 seconds, there is a high chance of customers subscribing for time deposits, with a probability of 90.3%
X4	If the number of employed individuals is greater than 5161 and the duration of the call is longer than 373.6 seconds, there is a high likelihood of 69.1% that the customer will subscribe for time deposits
X5	If the number of employees is greater than 5161 and the duration is less than or equal to 373.6 seconds, the likelihood of customers subscribing for time deposits is only 2.2%
X6	If the number of employees is less than or equal to 5161 and the duration is less than or equal to 373.6 seconds, there is a 63.1% chance that the customer will subscribe to time deposits
X7	If the number of employees is equal to or less than 5161 and the duration is less than or equal to 373.6 seconds, the likelihood of customers subscribing to time deposits in months other than May is 76.5%
X8	If the number of employees is less than or equal to 5161 and the duration is less than or equal to 373.6 seconds, the likelihood of customers subscribing to time deposits in May is 19.4%

The initial analysis of the 41188 bank customers of this Portuguese banking institution revealed a wealth of data that could be used for further investigation. However, using all of this data for analysis would be inefficient and would lead to low calculation accuracy. A simple solution to this problem is to randomly select a portion of the data that is representative enough to produce accurate results in a timely manner. To further improve the efficiency and accuracy of the model, the data can be divided into a training dataset (70%) and a validation dataset (30%). By building the model with the training dataset and then validating it with the validation dataset, overfitting of the model can be avoided and

the model can be more flexible and accurate. This can ultimately improve the quality and prediction effectiveness of the model.

Due to the initial data indicating that only 12.7% of customers ultimately subscribed while 87.3% did not respond, there may be significant bias in the model's performance if the data is modeled without consideration of this discrepancy. To improve the accuracy of the model, a test set was formed with 2060 samples of customers who subscribed and 2060 samples of customers who declined, denoted YES and NO respectively. These two datasets were then combined into a single test set NEW such that the proportion of subscription and declination are roughly equal so that the characteristics of each dataset can be better reflected.

Through the analysis and processing of data, the analysis results are shown in Tables 3 and 4.

Table 3: Simulation results for validation data.

actuality	prediction	
	yes	no
yes	78%	5%
no	22%	95%

Table 4: Simulation results for training data.

actuality	prediction	
	yes	no
yes	78%	8%
no	22%	92%

The results of the training and validation sets indicate that a reliable prediction model has been developed to accurately predict whether a customer will subscribe to regular deposits. The probability of predicting "Yes" and being actually "Yes" in the training set (TRAIN) was 78%, and the probability of predicting "No" and being actually "No" was 95%. The probability of predicting "Yes" and being actually "Yes" in the validation set (VALID) was 78%, and the probability of predicting "No" and being actually "No" was 92%. This indicates that the prediction model is able to generalize well, and can be used to accurately predict whether a customer will subscribe to regular deposits in different contexts.

3.2. Discussion

This trend appears to be consistent across different economic and social environments. For example, when the number of employee indicators is below the 5161 threshold, the subscription rate of regular deposits is higher in rural areas (75.2%) than in urban areas (47.9%) and in areas where the number of employees is above the 5161 threshold, the subscription rate of regular deposits is lower in rural areas (27.6%) than in urban areas (31.2%). This indicates that the number of employees in the economic and social environment has a significant impact on customers' decisions to subscribe to regular deposits.

The findings of the study on gender disparities were surprising to many, as it was previously believed that gender would have a significant impact on financial outcomes. However, the results showed that, while gender may have some influence, it is not enough to significantly affect the outcomes when all other factors are held constant. Furthermore, the distinction between a deposit and a loan is fairly obvious, and the results of the study were not significantly impacted by default status. This suggests that the independent variable, which is a history of default or not, is a reasonable

rationale for the results. Thus, it can be concluded that gender disparities do not have a significant impact on the present financial climate.

4. Conclusion

The duration of the last contact with the bank is an important factor influencing customers' decision to subscribe to regular deposits. Customers who have a longer duration are more likely to subscribe to regular deposits compared to those with a shorter duration. This indicates that the duration of the last contact is a significant factor affecting customers' subscription to regular deposits. Moreover, when the duration exceeds 373.6 seconds, the proportion of customers subscribing to regular deposits with smaller quarterly indicators of employee numbers increases to 90.3%. This further supports the idea that quarterly indicators of employee numbers have an impact on the subscription to regular deposits. Additionally, the latest month of contact also affects customer subscriptions for term deposits. In May, the proportion of clients who subscribe to term deposits decreases, which may be attributed to the warming of the real estate market, the rise of various investment and financial management options, and a series of directional discount policies by banks that directly result in a reduction in the proportion of bank subscriptions to term deposits.

Furthermore, the consumer's marital status, personal status, and housing situation are also important factors that affect the purchase of fixed deposits. The consumer's age and personal economic situation, such as the amount of current savings and the amount of capital invested, are also factors that can influence the consumer's decision to purchase a fixed deposit. In addition, the consumer's risk preference, whether or not they have previously purchased a fixed deposit, and the consumer's understanding of fixed deposits are also important factors that can affect their decision to purchase a fixed deposit. Furthermore, the consumer's communication channel preferences, such as whether they prefer to communicate through the internet or through traditional channels, can also influence their decision. In addition, the consumer's awareness of the advantages of fixed deposits, such as their safety and the potential for higher returns, can also influence the purchase of a fixed deposit. Lastly, the consumer's perception of the bank's reputation and its service quality can also be a factor that affects the purchase of a fixed deposit. Overall, the factors influencing the purchase of a fixed deposit are varied and complex, and can be both internal and external to the consumer. Understanding these factors and their relative importance is key to developing effective strategies to promote the purchase of fixed deposits.

References

[1] Yakubu I N, Ünvan Y N. Do bank-specific factors drive bank deposits in Ghana. *Journal of Computational and Applied Mathematics*, 2020, 28: 376-383.

[2] Mostak Ahamed M. Asset quality, non-interest income, and bank profitability:Evidence fromIndian banks. *Economic Modelling*, 2017, 1-14.

[3] Hu Zhuang. Empirical Analysis of Factors Influencing Bank Savings Deposits Under Measurement Model. *Yin Shan Journal (Natural Science Edition)*, 2018, 137-141.

[4] Feng Y. Research on the composite forecasting model of financial institution deposits. *Economic Research Guide*, 2018, 140-144.

[5] Lei. Empirical Research on Deposit Growth Model. *Journal of Nanjing University of Finance and Economics*, 2006, 46-48.

[6] Bond S R. Dynamic panel data models: a guide to micro data methods and practice. *Portuguese Economic Journal*, 2002, 1(2):141-162.

[7] Haroon A, Sharma A, Mir M. Factors Influencing Customer Choice of Bank Deposits: A Review. *International Journal of Research in Business and Social Science*, 2020, 9(12).

[8] Kebede B M, et al. Factors influencing customers' choice of deposit products: Evidence from the banking sector of Ghana. *International Journal of Economics and Financial Issues*, 2014, 473-481.

[9] Li Y, et al. Factors influencing customer choice of deposit products: Evidence from the Chinese banking sector. *International Journal of Bank Marketing*, 2020, 38(3): 563-590.

[10] Tsiros A, Karathanasopoulos, A. Factors influencing customer choice of deposit products: Evidence from the Greek banking sector. *International Journal of Bank Marketing*, 2022, 40(1): 87-112.