European Journal of Business, Economics & Management

Vol. 1 No. 2 2025

Article Open Access

Research on the Application of Machine Learning in the Pricing of Cash Deposit Products

Xiao Jing [1,*]

[1] Wealth Management, JP Morgan Chase, New Jersey, 07310, USA
* Correspondence: Xiao Jing, Wealth Management, JP Morgan Chase, New Jersey, 07310, USA

Received: 27 May 2025
Revised: 05 June 2025
Accepted: 30 June 2025
Published: 08 July 2025

Abstract: In the context of interest rate marketization and the vigorous development of financial technology innovation, the pricing of cash deposit products has a dilemma of real-time responding and personalization. Based on machine learning technology, this paper designs a price model path of data acquisition, attribute construction, algorithm screening and model practice. Through neural network, supply and demand curve optimization and other means, it quantitatively analyzes consumer behavior and product fit, empirically analyzes the effectiveness and operability of the model, and provides technical support for banks to optimize capital pricing efficiency and customer loyalty.

Keywords: cash deposits; machine learning; intelligent pricing; financial technology; customer behavior modeling

1. Introduction

In recent years, with the deepening of the electronization degree of the financial industry, the traditional method of determining the deposit price based on the experience value and the average interest rate has been unable to meet the requirements of the market fine pricing for each individual or institution entity. The pricing of monetary deposit products determines the source of the bank's core liabilities and directly affects the cost of funds and customer stickiness. Using machine learning's strong nonlinear modeling capabilities and data analysis capabilities can offer the possibility of an intelligent, agile price setting model. This paper focuses on the specific ways and effectiveness of applying machine learning to the price setting of money-based savings products.

2. Theoretical Basis of Cash Deposit Product Pricing Driven by Machine Learning

2.1. Basic Concepts and Characteristics of Cash Deposit Products

Cash deposit product is a very common deposit product, which refers to the high liquidity and low risk products provided by banks and other financial institutions to individuals and companies to meet the allocation requirements of people's short-term financial needs. Deposit products are mainly divided into term depsoit, daily deposit, structural deposit such as notice deposit and so on. Its main characteristics are ease of use and security of funds [1]. There is usually no specific time limit, and there is less interest than regular, but it is flexible and convenient to use. It is a common financial allocation method used by people in daily life, while managing the liquidity for each deposit product holder.

With the intensification of market competition and the continuous advancement of interest rate liberalization, traditional cash deposit products are faced with the "dilemma" of not only reducing product rates while still retaining customers, but also taking into account multiple requirements of customer benefits and services. Therefore, fine but flexible price determination has become an important means for banks to ensure cash deposit volume, expand cash deposit volume and improve customer satisfaction. It is also a key step in developing customer loyalty.

2.2. Basic Principles of Machine Learning and Its Financial Application Value

Artificial Intelligence (AI) is a technology that relies on data-driven information to self-learn the problem itself and predict future events, categories, etc. It is to analyze and predict new data by establishing and using mathematical models based on existing historical data. Specifically, it can be divided into four categories: supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning [2]. The most common one is the regression and classification of supervised learning, which is widely used in financial modeling [3].

In all aspects of banking business, the feature extraction capability and nonlinear simulation characteristics relying on machine learning are widely used in credit assessment, fraud detection and other aspects. Compared with traditional data analysis technology, machine learning has accurate prediction ability and automatic operation ability, which is especially suitable for dealing with financial big data with complex structure, more variables and large dynamic changes of data. In terms of the pricing of fixed deposit products, machine learning is used to mine historical data such as customer behavior and market interest rates to train a dynamic computational pricing method, so as to achieve a "customer-centric" interest rate adjustment mechanism and improve the attractiveness of products and the efficiency of market response. Using tools such as XGBoost, banks can divide large customer groups into several types and then design interest rates to meet customer needs while balancing income and expenditure (Figure 1).

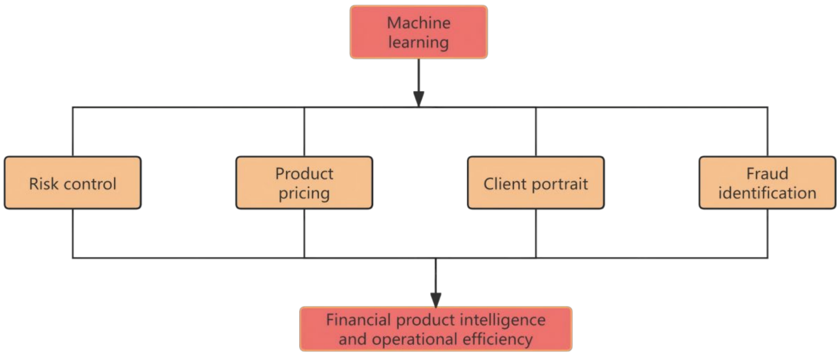

Figure 1. Typical Application Scenarios of Machine Learning in the Financial Industry.

New technologies such as Random Forest, XGBoost, and neural networks can be used by banks to efficiently personalize analytics and pricing for large scale customers. This can improve customer satisfaction while ensuring a relative balance between revenue and expenditure.

3. Technical Path Design of Machine Learning Model in Deposit Pricing

3.1. Data Acquisition and Preprocessing Methods

In the construction of cash savings product pricing algorithm, the dimension, granularity and accuracy of data play a key role in the expected stability and effectiveness of the model, and even determine the application value of the model. Therefore, careful planning and holistic consideration of data sources and construction methods are required [4]. Generally speaking, the data used for model training and estimation can be divided into two categories: one is internal label data, such as customer account data, account term, past interest rate trend, account operation activities and other data; The other is external macroeconomic data, which covers a wide range and has clear trend characteristics, and competitor rates provided on similar deposit products. For example, the adjustment of the deposit rate of the central bank is the reference of the pricing benchmark of various financial institutions, and its change often plays a bellwether role; The growth rate of Consumer price index (CPI) and gross National product (GNP) can reflect the overall economic growth state, and has a very important impact on consumers' preference for asset allocation and savings willingness.

The first step is to process the collected data to ensure that the input characteristics do not change and that the model is easy to understand. The calculation method is as follows:

$$zi = \frac{xi-\mu}{\sigma} \tag{1}$$

Among them, xi represents the original eigenvalue, μ is the sample mean of this feature, σ is the standard deviation, zi is the normalized value. Moreover, in order to improve the model's ability to extract nonlinear features between variables, min-maximum normalization is usually adopted, in the form of formula (2):

$$x'i = \frac{xi-xmin}{x\,max-xmin} \tag{2}$$

During processing, outliers are eliminated and missing values are filled (average method, complement method and KNN method can be used). Then variables related to deposit interest rate are selected according to information gain theory and PCA based principal component analysis method to form a relatively stable attribute set. For time series data such as transaction amount, dynamic encapsulation can be achieved by sliding window method (such as weekly, every 30 days average) to adapt to changes in modeling requirements [5]. The level of data processing has a key influence on the stability and discriminability of the modeling results.

3.2. Model Selection and Algorithm Comparison

Regression problem is the focus of cash deposit product pricing. The purpose of modeling is to predict the future interest rate level through multi-source data. Common models include linear regression, random forest, XGBoost, neural network, etc. The following is a comparative analysis of several mainstream algorithms:

As can be seen from Table 1, except for the poor performance of linear regression model in terms of company equity, XGBoost performs well in all businesses. XGBoost is more suitable when the data dimension is high and there is strong nonlinearity between data dimensions. Random forests are more suitable for stable operation when the amount of data is relatively small. Neural network can train relatively complex models, but it has a large demand for large-capacity data and complicated calculation, so it is weak to interpret the output results of the data. Therefore, in practical application, the main model and the basic model can be combined, such as XGBoost as the main model, and the linear regression model to assist the analysis of the results, in accordance with the legal requirements. A model system with high precision and strong interpretability is formed.

Table 1. Comparison of the Applicability of Mainstream Machine Learning Models in Deposit Pricing.

Model name	Advantage	Inferior position	Application scenario
Linear regression (LR)	The model is simple in structure, easy to interpret and fast in training	The linear assumption of variable relationship requires high accuracy and is limited	Establish the benchmark model, variable screening stage

Random forest (RF)	It can handle high dimensional and nonlinear features and has strong anti-overfitting ability	The structure of the model is complex and the training time is long	Multi-variable interaction, high-dimensional feature prediction task
XGBoost	High prediction accuracy and strong ability to deal with missing values	Parameter tuning is complicated and the training cost is high	Financial time series, behavioral variable modeling
Neural network (DNN)	Strong nonlinear fitting ability, suitable for large-scale samples	The explainability is weak and it is easy to overfit	Long-term trend modeling, big data scenarios

3.3. Modeling Process of Deposit Interest Rate Pricing

For commercial banks, the pricing of deposit products demand is related to the external macroeconomic situation, such as the central bank's benchmark interest rate and inflation rate; Customer's personal consumption habits and capital flow patterns; As well as interbank rate strategy, product competitiveness and regional financial ecological environment and other aspects of the factors. Therefore, modeling should include the whole process from collecting data to applying the model (Figure 2).

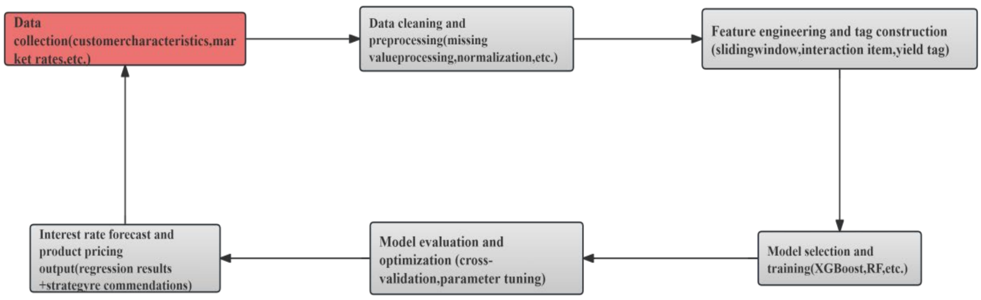

Figure 2. Modeling Flowchart of Deposit Interest Rate Pricing.

In addition, it is necessary to introduce several dynamic attributes (such as the activity in the last 30 days, the rise and fall total size of the account balance within 720 days, etc.) to construct the feature set. Cross-validation strategies (such as 5-fold cross-validation) are used to ensure that the model has good generalization ability, and the precision of parameter setting is further optimized through grid search and Bayesian optimization. In addition, the above forecast results can be superimposed into the business strategy according to the demand, and after combining the risk weight, the final price recommendation results can be generated for banking institutions to adjust the floating interest rate policy reference. In this way, the data-driven, model updated, application closed-loop "data-model-application" logical link is implemented to achieve a more accurate and intelligent attempt to set the price of savings products.

3.4. Model Deployment and Dynamic Iteration Mechanism

To ensure the successful application of machine learning models in real-world business scenarios, the process of going live, monitoring, evaluating, and iterating needs to be fully managed. Therefore, an "executable, observable, modifiable" automatic deployment iteration mechanism is required. The architecture design for deployment and update is shown below (Figure 3).

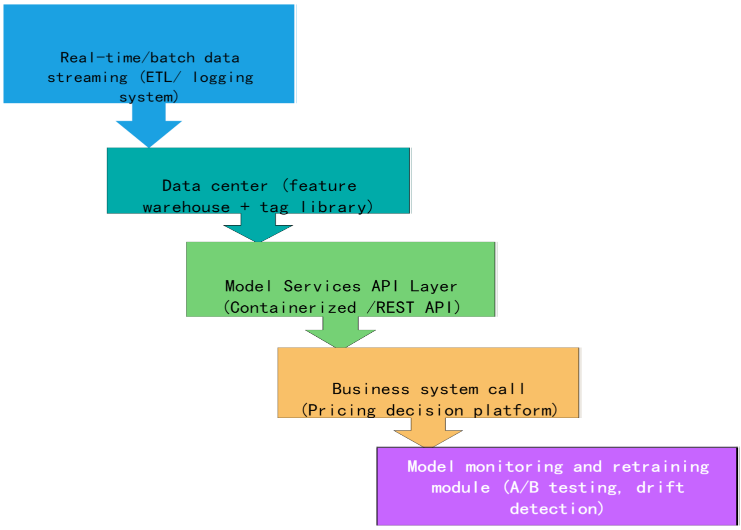

Figure 3. Machine Learning Pricing Model Deployment and Dynamic Iteration Framework.

Usually, the model is deployed as a Docker or Kubernetes container, and after running, it provides services to the application in the form of apis, achieving the prediction result of "input and feedback". After the model is started, the model monitoring platform is used to continuously monitor the prediction deviation and input data changes to judge the concept drift and data drift problems. When the predicted results gradually deteriorate, retraining is triggered, and a cyclic iteration mechanism is formed.

In addition, through the A/B testing mechanism, the effect of different versions of pricing models can be compared in specific scenarios, and the optimal model can be selected and put into use. Such a dynamic learning mechanism can help financial institutions adjust the pricing strategy of savings products in the increasingly competitive financial market, improve the efficiency of capital cost management and customer satisfaction.

4. Model Application Scenario and Empirical Research Analysis

4.1. Specific Application Cases of the Model in Cash Management Products

Taking the "stepped income fixed rate passbook" issued by Bank A in 2024 as an example, XGBoost algorithm is used to train the interest rate fluctuation prediction model for intelligent pricing. According to the input variables of deposit time in the past three years, account changes, and customer star marks (such as gold and silver customers), the model will predict the response degree under the interest rate range of corresponding customers. Optimize pricing strategies based on forecast results to improve product competitiveness and deposit retention (Figure 4).

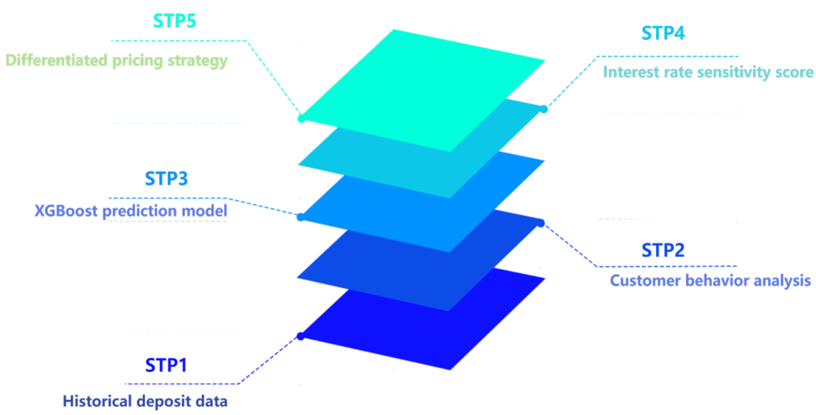

Figure 4. A Schematic of the Deposit Product Pricing Process Based on Machine Learning.

The actual promotion data of the new product proved that the successful conversion rate of users reached 18.3% of the target user group after only 3 months using this algorithm, and the average duration of their funds on site increased by 22 days, which was significantly improved compared with the previous time period compared with the product (without this algorithm). The effective use of this algorithm greatly shortens the price determination time, so that the "rapid response-precision sales" form a closed loop.

4.2. Customer Behavior Prediction and Product Matching Optimization

The choice of cash deposit products is predictable to a certain extent, and banks can accurately match different types of customers and corresponding best products by establishing a loss rate prediction system and a preference scoring system, so as to achieve differentiated pricing and precise marketing.

Let the historical behavior feature vector of customer i be $Xi = \{x1, x2, \dots, xn\}$, and the goal is to predict its preference score Sij on product j, which can be modeled using the following functions:

$$Sij = \sigma(W \bullet Xi + b) \tag{3}$$

Where σ is the activation function (such as Sigmoid), W is the trained weight vector, and b is the bias term. Combined with customer history retention rate and transaction frequency, the expected revenue function can be further calculated:

$$Uij = pij \bullet rj - cj \tag{4}$$

Where, pij represents the successful conversion probability of customer i on product j, rj is the product interest rate, and cj is the marketing cost. The final system can carry out the optimal matching of customer and product according to the principle of maximizing Uij.

In addition, we also tried to test the effectiveness of our strategy on a group of customers known as the "three high", that is, an area with a high turnover rate in banking. After six months of implementation, the overall customer retention rate has increased by 4.5%, and the budget cost that we retained of those valuable customers as been decreased by 10.2%.

4.3. Comparative Analysis of the Effect of Different Algorithm Models

By comparing the linear model, random forest model, XGBoost model and neural network model in the same data set, the efficiency difference of each model in the application of linear product pricing is verified. The overall evaluation chart is as follows (Figure 5 and Table 2):

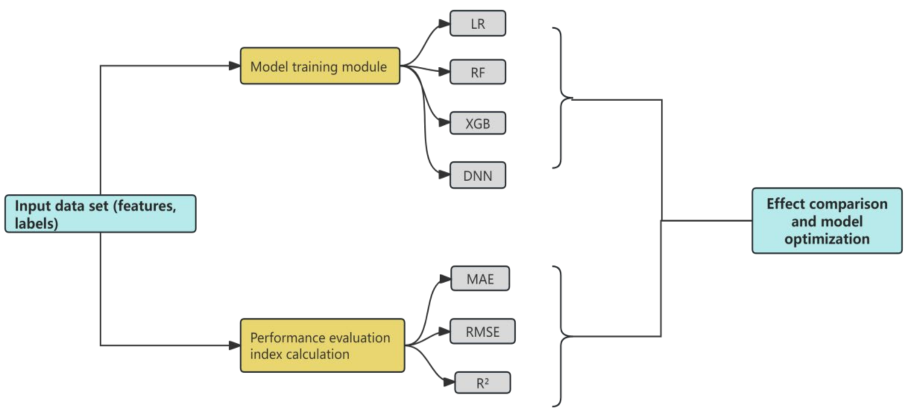

Figure 5. Model Evaluation and Comparative Analysis Frame Diagram.

Table 2. Comparison of Effects of Different Models.

Model	MAE	RMSE	R^2	Training time (seconds)
Linear regression	0.134	0.201	0.55	0.5
Random forest	0.096	0.150	0.71	3.2
XGBoost	0.089	0.138	0.78	4.5
Neural network	0.086	0.136	0.67	9.1

It is found that the prediction ability of the model is better than that of the neural network model and XGBoost model, but in the environment with limited resources, the random forest model can be selected according to the trade-off between efficiency and performance. Linear models, on the other hand, are straightforward but less predictive, and are only used for preliminary analysis or in environments where a high degree of explanatory model regulation is required.

4.4. Model Adjustment under Risk Control and Compliance Constraints

Since the implementation of financial models must meet the requirements of risk control and compliance, it is necessary to introduce model interpretability and risk constraint function to avoid being regarded as a "black box". The risk control constraint function of constructing interest rate adjustment is as follows:

$$\Delta r \leq \varepsilon,\ \forall r \in R \tag{5}$$

Where, Δr represents the rate change range of pricing in adjacent cycles, and ε represents the maximum fluctuation limit set within the bank to ensure the stability of the pricing strategy and avoid customer transactions. In order to enhance the transparency of the model, SHAP value or LIME method is introduced to explain the contribution of each feature to the prediction results (Table 3):

Table 3. Model Feature Importance Diagram (Based on SHAP).

Characteristic variable	SHAP value contribution degree	Weight ranking
Customer past 6-month account balance size	0.35	1
Deposit Historical Pricing	0.22	2
Client Type of size	0.18	3
Competitor Pricing	0.12	4
Central Bank Benchmark rate SOFR	0.08	5

Finally, to ensure that the valuation process of cash savings products by the machine learning model complies with laws, regulations, and professional ethics in the financial field, we need to design a systematic compliance review mechanism that follows the principles of "strong control, traceability, and transparency". This mechanism ensures that the model does not include sensitive variables and other factors that may lead to discriminatory outcomes. And complete the legal and risk control due diligence process (Note: See the Figure 6 below).

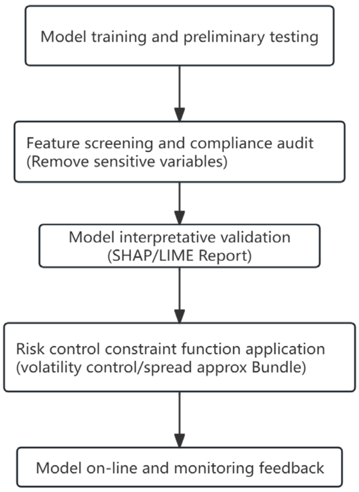

Figure 6. Flow Chart of Compliance Model Review and Adjustment.

With the help of the technology and institutional collaboration model, we can not only greatly enhance the accuracy and fit of the machine learning model to the cash class duration pricing model, but also comply with the requirements of financial regulation in the process of promoting differentiated pricing. On the one hand, we use high-level algorithms, technical transparency and model iteration paths to ensure the scientific and applicability of the model; On the other hand, the legal evaluation path, risk evaluation criteria, model application monitoring system ensures that it is always under legal supervision. The two cooperate to form a new smart pricing model of "technology-based and legal", which is not only conducive to the banking industry to better achieve sound operation, development and progress in a highly competitive environment, but also helps to build a smart financial ecosystem in the future.

5. Conclusion

With the continuous development of artificial intelligence technology, the application of machine learning is more and more widely applied in the financial industry, and it has also brought great achievements in the pricing of savings cash products. This paper mainly uses data pre-processing, model construction and experimental analysis to explain the existing improvement means of machine learning in improving pricing accuracy, enhancing customer matching, optimizing risk control ability, and satisfying problems in risk regulation, and looks forward to the future development trend of machine learning.

References

1. T. Alhalabi, V. Castro, and J. Wood, "Bank dividend payout policy and debt seniority: Evidence from US banks," *Fin. Mark. Inst. Instrum.*, vol. 32, no. 5, pp. 285–340, 2023, doi: 10.1111/fmii.12183.
2. I. N. Widyanti, "The influence of liquid collateral types and credit processes through digital marketing on customer decisions to take cash collateral credit at Bank ABC," *Jurnal Indonesia Sos. Teknol.*, vol. 5, no. 3, 2024, doi: 10.59141/jist.v5i3.981.
3. M. B. Pasha, R. N. Rao, S. Ismail, *et al.*, "Tribo-informatics approach to predict wear and friction coefficient of Mg/Si_3N_4 composites using machine learning techniques," *Tribol. Int.*, vol. 196, 2024, doi: 10.1016/j.triboint.2024.109696.
4. C. K. Ch'ng, "Hybrid machine learning approach for predicting e-wallet adoption among higher education students in Malaysia," *J. Inf. Commun. Technol.*, vol. 23, no. 2, pp. 177–210, 2024, doi: 10.32890/jict2024.23.2.2.
5. C. Tang, Y. Qin, F. Wu, and Z. Tang, "Dynamic demand-aware power grid intelligent pricing algorithm based on deep reinforcement learning," *IEEE Access*, vol. 12, pp. 75809–75817, 2024, doi: 10.1109/ACCESS.2024.3406338.