

## Classification Model of Household Saving Behavior Based on RF

Jiayan Tang, Linqi He, Jing Cao, Yuxi Li, Cichun Liao, Zirong Ji, Qiahong Peng and Yan Zhou\*

College of Mathematics, South China Agricultural University, Guangzhou, China

\*Corresponding author. Email: [zhouyan198021@126.com](mailto:zhouyan198021@126.com)

**Abstract.** In order to stimulate residents savings and realize the balance of savings-investment, this paper selected the data of household savings in China's household finance survey as the main research object of the Behavior Classification Model, and mainly investigated the investment behavior, household consumption and other factors as the constraints of household savings decision-making. The establishment of RF Model for decision classification prediction and feature extraction, improve the interpretability of behavioral differences and high prediction rate, and further analyze the correlation between savings-investment balance and social value cycle with Solow Model, provide a new thinking line for the improvement of regional economy, and then put forward targeted suggestions and strive to bring about better social effects.

**Keywords:** RF model, characteristic extraction, decision-making classification.

### 1. Introduction

The savings habits of nationals vary considerably between countries with seemingly identical economic sizes and political systems, a phenomenon that was discovered as early as the Adam Smith era and raised the associated economic issues that were re-emphasized after the global financial crisis. As one of the main evaluation criteria of the country's overall economic level, the savings rate of residents, which was previously high, can provide abundant funds for investment and become an important factor for the sustained and rapid development of the national economy.

For savings study, Y. Mugerma[1] et al. suggested that residents savings decisions were more closely related to social and investment behavior, and J. van de[2] believed that the diversity of influencing factors that need to be considered can lead to greater decision-making contradictions. Therefore, it is suggested that targeted strategy design is a good way to consider behavioral response to make allocation decisions. Therefore, in order to put forward a good design of strategies to stimulate residents savings, it is possible to classify the behavior in the consideration factors involved. In behavioral decision analysis, K. L[3] et al. explored the interaction factors with energy-saving behavior by applying the comprehensive behavior decision model, although the model had a high interpretation of behavior difference, but its predictive ability in the behavior process was low.

Therefore, this paper will take a Random Forest Classification Model instead of Behavior Decision Model, combined with the actual data for the multi-factor analysis of residents' savings behavior, for its behavior classification has good predictive ability, at the same time, at a certain accuracy rate can successfully predict the residents savings decision-making on the basis of, and then combined with the

Solow Model to put forward targeted economic recommendations, and promote the overall level of the economy.

## 2. Basic data analysis

This paper selected the data of Guangdong Province of China's Household Finance Survey updated by SWUFE in December 2019 as the research object. The basic personal information, savings, capital allocation and other variables of the respondents were selected for research and analysis. This paper mainly focuses on the real data distribution of demand deposit, time deposit and cash flow in savings, and carries out a preliminary descriptive statistical analysis of the data.

**Table 1.** Descriptive statistics and skewness, kurtosis analysis of the data.

	Demand Deposits	Time Deposits	Cash Flow
Average	30441.227	20249.492	8751.964
Median	1000.000	0.000	2000.000
S.D.	1693.935	1402.303	684.9782
Skewness	12.414	10.081	24.075
Kurtosis	223.897	139.433	776.540

We can get that most of the residents in Guangdong province had a habit of savings and cash with the high liquidity of the money. But regular savings is not high, which would be related to daily income and investment transformation capital turnover of high activity, also be related to low-interest and low-risk wealth management products such as bank annual interest. And, it can be seen that the three variables are significantly greater than zero, indicating that there is a large gap between the data distribution and the normal distribution, resulting in the "long tail" skewness effect. Meanwhile, the right-skewed distribution will increase the risk of the decline of savings rate in the process of saving selection and saving amount decision.

## 3. Resident savings decision prediction based on RF model

### 3.1. Analysis of influencing decision factors

Household savings referred to the voluntary deposit of part of capital income by individuals or families, which was mainly divided into demand deposit and time deposit according to the purpose.

Therefore, the reason, plans and purpose of saving will be used as constraint conditions, which was a breakthrough point in addition to the liquidity of the above factors, internal factors should also be observed residents.

### 3.2. Variable definition and measurement

In order to predict residents deposit preference, a Prediction Judgment Model based on random forest RF algorithm was established. Since the residents savings decision can be divided into three parts: whether or not to deposit, demand deposit amount, and time deposit limit, this kind of variable will be defined as a qualitative variable according to the numerical range, as the interpreted object of the model.

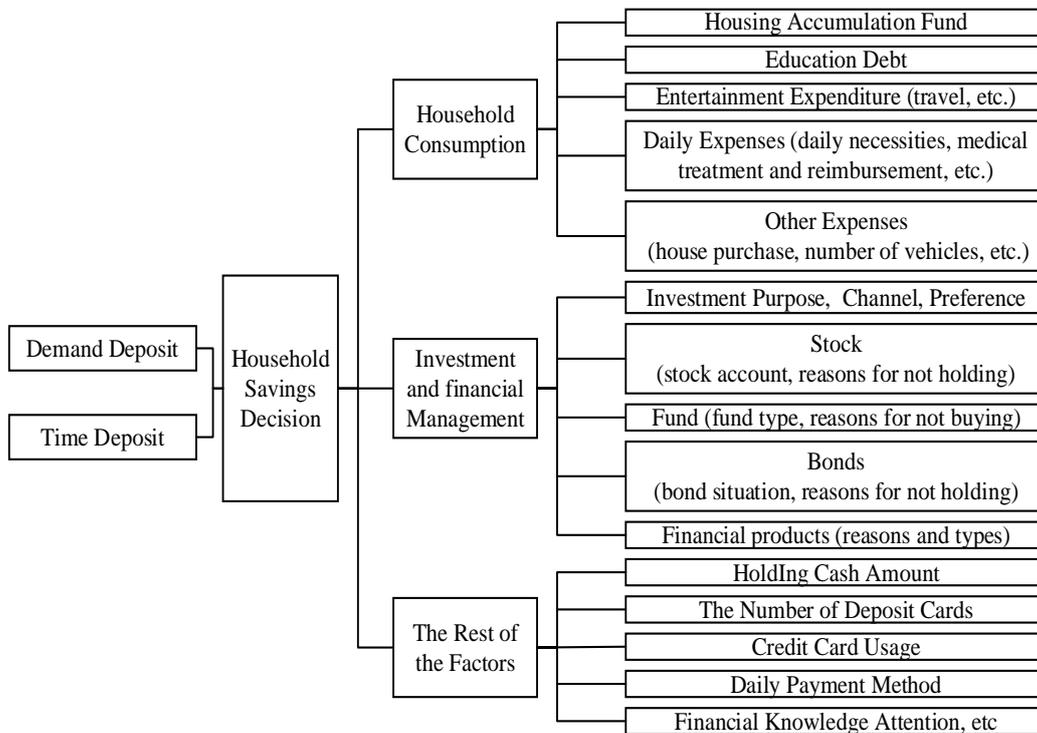
**Table 2.** The variables are defined as the measures.

Variable	Conditions	Number
Time Deposits Behavior $x_1$	No demand deposits	0
	Demand deposits	1

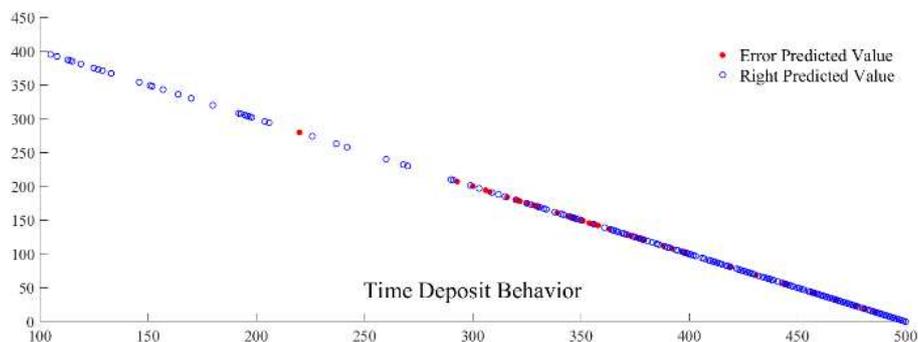
Demand Deposits Behavior $x_2$	No time deposits	0
	Time deposits	1
Time Deposit Limits $x_3$	Annual deposit of less than 10,000	0
	Annual deposit of 10,000 to 50,000	1
	Annual deposit of 50,000 to 100,000	2
	Annual deposit of more than 100,000	3
Demand Deposit Limits $x_4$	Annual deposit of less than 10,000	0
	Annual deposit of 10,000 to 50,000	1
	Annual deposit of 50,000 to 100,000	2
	Annual deposit of more than 100,000	3

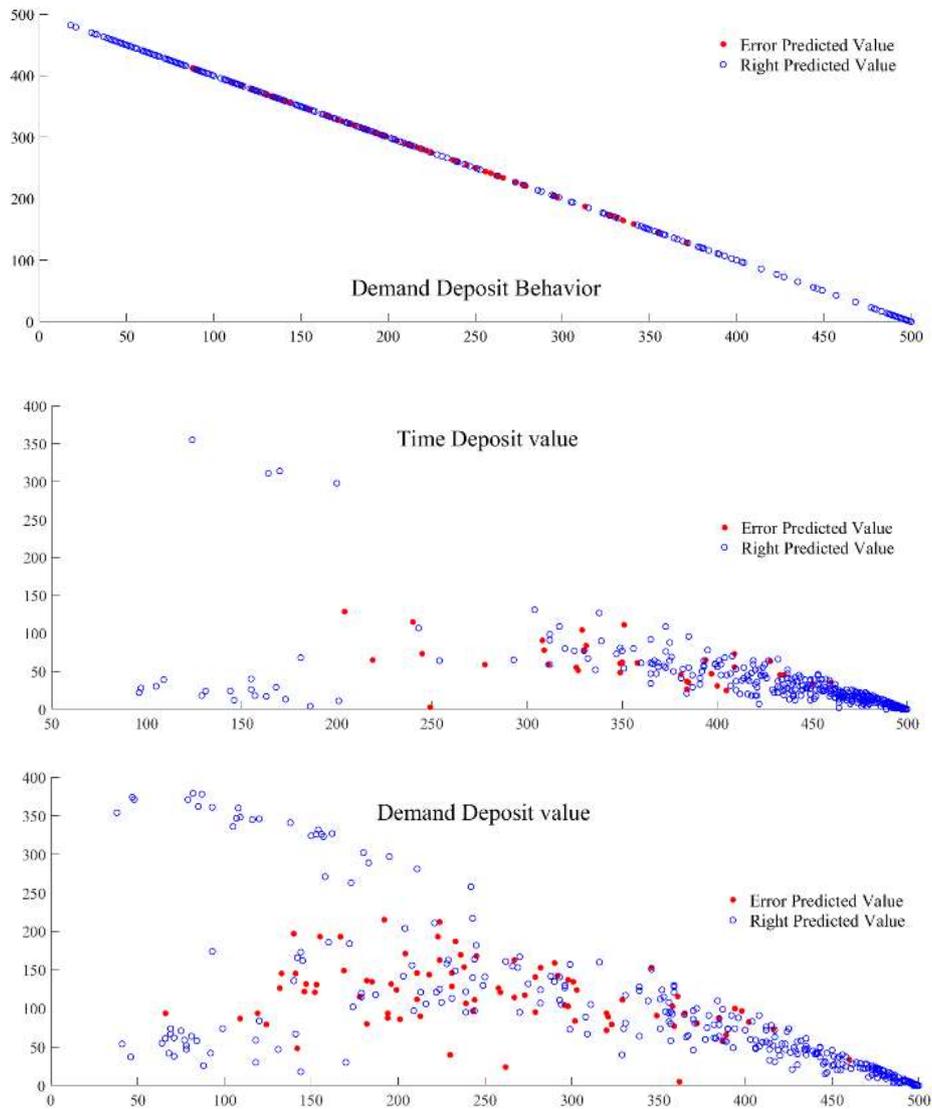
### 3.3. Model building

In view of the random forest response characteristics are still serious under the premise can still maintain a good accuracy advantages, the random selection of 8% of the sample data was multi-classified to 4 kinds of residents savings decision-making behavior, and the Decision Tree Evaluation Model (Random Forest) based on the integrated algorithm was realized and solved. The results are shown in Figure 1:



**Figure 1.** The influencing factors of resident savings decision.





**Figure 2.** Sample tests.

Note: The upper left is the time deposit behavior, the upper right is the demand deposit behavior, the bottom left is the time deposit limit, the bottom right is the demand deposit limit.

The accuracy rate of the above four decision-making behaviors were based on RF Model:

$$(Y_1, Y_2, Y_3, Y_4) = (88.25\%, 81.75\%, 90.50\%, 81.25\%)$$

Under the condition that the average prediction rate was higher than 80%, the characteristic extraction was analyzed and the main characteristics affecting the decision-making behavior of residents savings were studied.

### 3.4. Feature selection

By predicting the above model, the decision factors that affect residents savings can be sorted according to the characteristic score:

**Table 3.** Score table.

Impact Characteristic	Household Consumption	Investment and Finance Management	The Rest of the Factors
Score	2935.84	1291.06	2485.12
Rank	1	3	2

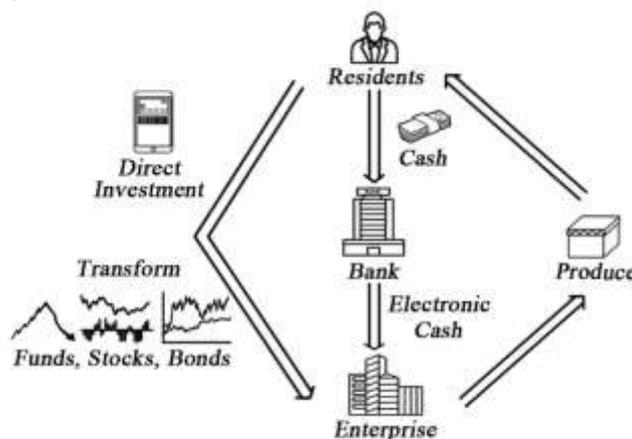
According to the model, the amount of savings and the amount of cash held as the main impact characteristics, followed by household consumption of daily expenditure, of which medical and health insurance consumption accounted for the main part, and finally, investment and financial management part, more than 80% of residents did not hold any investment and wealth management products, nearly half of which due to the lack of relevant knowledge and gave up financial management, As a result, savings and investments cannot be successfully converted, followed by the level of daily expenditure as a representative of the household's actual spending power, thus, the expression of the family's well-being, will be the main assessment characteristics.

### 3.5. Demonstration and suggestions

P. R. Mahalingam[4] and others forecasted the maximum balance of savings through the current account month combined with the user transaction log database, and set the influencing factors according to the occurrence time to predict the amount of residents savings. The method set the time, and its actual consideration coefficient was limited, so it was difficult to achieve the diversity of RF Model practical considerations.

The RF Model was based on the average accuracy of decision number classifiers in the case of behavior synthesis prejudgment was higher than 80%, and was related to the number of valid data sets, and it proved the wide application of the model in economic decision-making under the conditions of distinguishing the main characteristics of resident behavior decision-making and the prejudgment. At the same time, the research showed that residents savings decisions are mainly influenced by daily expenditure and investment behavior.

According to the high savings rate of Solow Model of the deposit scale side showed the narrow ness of the national investment channel, and its abnormal savings rate will lead to the imbalance of economic structure, on the contrary, the diversification of social funds will greatly stabilize the current economic system, and achieve the golden law level of economic welfare.



**Figure 3.** Savings-investment cycle.

The policy measures to stimulate residents savings, residents will convert savings assets. Through the banks and other commercial institutions, the funds are put into the development of small and medium-sized enterprises. The enterprises get enough funds to maximize the social output, the overall social welfare rises, and the residents savings further increases, which is the social value cycle

phenomenon under the factors of saving stimulation. The savings investment conversion mechanism affects the investment direction and efficiency of the whole society's funds, and the significance of residents' savings is not only as consumption savings, but mainly as the conversion of current investment production.

- The financial industry began to innovate in various areas of consumer financial services with its large customer base, thus gradually gaining the advantages of big data technology[5], under the condition of data conditions related financial institutions should be based on the dynamics of financial markets to carry out accurate marketing, according to the real-time situation of different types of residents to adjust the structure of products to meet specific needs.

- The random forest model uses unbiased estimation to calculate the error value, can run in parallel, can balance the error, maintain the accuracy in the case of the loss of characteristics, the combination of high-tech and new model, is conducive to improving the service efficiency of banking, and selectively stimulate spending and investment decision-making.

- Encourage similar institutions to try innovation, adopt information technology under the condition of insufficient information of residents, establish effective models to improve residents' credit database, and use local characteristics to propose solutions for local culture.

#### 4. Summary

The decision of residents savings referred to the trend of the consumption surplus diversion of residents, which mainly included the saving behavior and the investment transformation, the essence was to stimulate the residents savings to achieve saving-investment balance.

The resident saving behavior is affected by more factors, not just by the time influence factor setting, this paper selects the rf model with strong generalization ability to carry out the classification test and feature extraction of residents' comprehensive savings behavior, and the research shows that daily expenditure, the number of savings cards and insurance will be the main research object of the financial institution. At the same time, Solow Model further analysis of the social gold rate level will be the balance of savings-investment mechanism, the paper puts forward that financial institutions need to use information technology to establish behavior decision-making model, supplement the resident information base, reduce the risk of information asymmetry.

#### References

- [1] Y. Mugerman, O. Sade and M. Shayo, Long term savings decisions: Financial reform, peer effects and ethnicity [J], *Journal of Economic Behavior & Organization*. 106 (2014) 235-253.
- [2] J. van de Ven, SIDD: An adaptable framework for analysing the distributional implications of policy alternatives where savings and employment decisions matter [J]. *Economic Modelling*. 63 (2017) 161-174.
- [3] K. L. van den Broek, I. Walker and C. A. Klöckner, Drivers of energy saving behaviour: The relative influence of intentional, normative, situational and habitual processes [J]. *Energy Policy*. 132 (2019) 811-819.
- [4] P. R. Mahalingam, S. Vivek. Predicting Financial Savings Decisions Using Sigmoid Function and Information Gain Ratio [J]. *Procedia Computer Science*. 93 (2016) 19-25.
- [5] Z. Y. Zhou. Analysis of Marketing Strategies for Related Products in China's Financial Market from the Internet Perspective—Taking Ant Credit Pay as an Example [P]. *Proceedings of the 2nd International Symposium on Social Science and Management Innovation (SSMI 2019)* (2019).