

# Analysis of Influential Factors of Fund Purchasing Behavior Based on Multimodality--Based on College Students in Shanghai at Pudong New Area

Jianrong Li<sup>1,\*</sup>, Zhaobo Lin<sup>2</sup>

<sup>1</sup>Department of Economic statistics, Shanghai Polytechnic University, Shanghai, China, 201209

<sup>2</sup>Fuzhou University, Fuzhou, China, 350108

\*Corresponding author: 18077265201@163.com

**Abstract.** The purpose of this paper is to study the behavior of college students in purchasing funds and their influencing factors. By distributing online and offline paper questionnaires to college students in Pudong New Area, Shanghai, and at the same time, obtaining first-hand information about college students' fund purchasing through the Internet. According to the survey results of the questionnaire, followed by the analysis of the characteristics of the fund purchasing behavior of college students in Pudong New Area, using the decision tree model to analyze the fund purchasing behavior and purchasing decision, through the questionnaire research and the Zhihu comment crawler to draw the word cloud diagram to derive the different influencing behaviors of college students on the purchase of funds, and the use of LDA sentiment analysis model to derive the double comments of college students in Pudong New Area on the attitude of fund purchasing. The following main conclusions were obtained through this research study:1) Social media communication has become a key factor influencing fund purchases, with college students prioritizing news media and online bloggers for initial understanding before purchasing funds;2) College students' purchase of funds is highly enthusiastic but mostly experimental. Due to their insufficient understanding of the fund market and small sources of income, they choose low-risk funds for experimentation. Undergraduates' financial knowledge is shallow and their ability to bear economic risks is weak, and they lack sufficient knowledge and experience in financial literacy. Finally, the correlation suggestions are put forward: refuse to blindly follow a single source of information to buy funds; some undergraduates have little financial knowledge and buy more impulsively; keep a good mindset of long-term investment, not caring about the momentary gains and losses.

**Keywords:** Fund purchasing behavior; Influencing factors; CART Decision Tree Model; Latent Dirichorean Sentiment Analysis Model; Cardinality Analysis.

## 1. Introduction

### 1.1. Background

In the era when college students are becoming more and more aware of investment and finance, fund as a low-risk investment method has become the priority choice of contemporary college students [1]. In addition, due to the decline in market interest rates, everyone is looking for new investment areas, a large amount of money into the fund market, set off a "fund fever", attracting more college students to invest in the fund market [2]. However, college students do not have a high level of understanding of the fund market, and their knowledge of funds is relatively narrow, which makes them prone to improper purchasing behavior, or losses caused by wrong information [3].

Therefore, by studying the behavioral factors affecting college students' fund purchasing and making suggestions to all parties, this paper is used to promote college students to establish a perfect fund purchasing and investment thinking as well as to improve college students' investment and financial literacy [4], and hopes to positively guide college students' investment and financial management through the influence of college students' fund purchasing behaviors [5].

## 2. Materials and methodologies:

### 2.1. Data

This study obtains primary data through questionnaires as well as literature surveys, and in terms of questionnaires [6], offline and online questionnaires are distributed to university students in Pudong New Area, and the sample size is determined and distributed as shown in Table 1:

**Table 1.** Sample size determination and allocation

Pudong New Area	Main universities	Main universities
port neighborhood	Shanghai Maritime University, Shanghai Ocean University, Shanghai Institute of Mechanical and Electrical Engineering, etc.	500 people
Jinhai Road area	Shanghai Second Polytechnic University, Shanghai Sanda College, Shanghai Lixin Accounting College, etc.	150 people
Other areas	Shanghai Customs Institute, Shanghai University of Traditional Chinese Medicine, etc.	160 people

### 2.2. Methods of analysis

#### 2.2.1 Decision Tree Models

A decision tree is a predictive model that represents a mapping relationship between object attributes and object values. Each node in the tree represents a particular object, while each forked path represents a possible attribute value [7], and each leaf node corresponds to the value of the object represented by the path traveled from the root node to that leaf node. In the algorithm, the Gini coefficient represents the uncertainty of the sample [8]. As the number of each category converges to the mean, the Gini coefficient becomes larger and more uncertain.

#### 2.2.2 Latent Dirichlet Allocation

Latent Dirichlet Allocation is a method of extracting potential topics in documents, the basic idea of the method is: a number of documents are automatically encoded into a certain number of topics [9], greatly reducing the human intervention and burden, when the human set the number of topics and then run the LDA model to get the distribution probability of the words under each topic as well as the probability of the corresponding topic of the document. The working principle is equivalent to a projector, by determining the number of topics and then setting the values of  $\alpha$  and  $\beta$  parameters on the machine to control the working state of the machine, and finally randomly generating a new document that has no relationship with the original document, and then comparing the correlation between the new document and the original document to illustrate the goodness of the model, and finding the optimal model and the values of  $\alpha$  and  $\beta$  among many models with different parameters. In this,  $\alpha$  and  $\beta$  control a Dirichlet distribution respectively,  $\alpha$  randomly generates a polynomial distribution  $\theta$ ,  $\theta$  follows to generate a topic  $z$ , while  $\beta$  randomly generates a polynomial distribution  $\Phi$  of the corresponding words of the topic, and after that  $z$  and  $\Phi$  generate the word  $w$ , and so on to generate a document containing  $m$  words, and finally generates  $n$  documents under  $k$  topics, and the process is shown in the following figure, and after arriving at the document by calculating the Confusion are determined the optimal number of topics, other initial parameters of LDA are determined empirically, and the topic-word probability distribution is obtained by topic extraction, as shown in Figure 1[10].

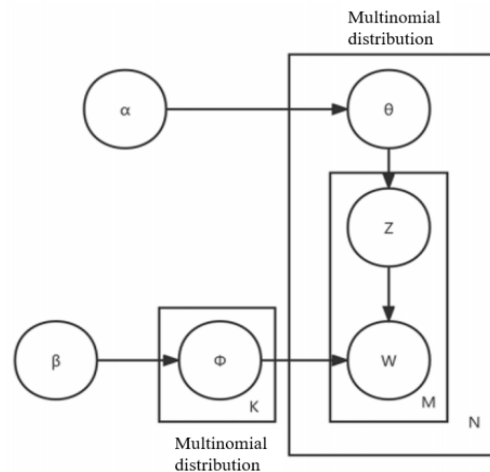


Figure 1: Flowchart of LDA model

### 3. The characteristics of fund purchasing behavior

#### 3.1. Fundamental Analysis of Buying Behavior

Before purchasing funds, college students will collect investment information through different channels, extracting the information from the questionnaire as shown in Figure 2 below: this study found that news and media are the main sources of information for consumers to purchase funds, with the highest selection rate, followed by the number of times of choosing funds through personal experience is 162; at the same time, we also found that the fund information provided by the financial reports and professional financial websites is often the most complete, but only a minority of college students will choose the two. However, only a small number of college students will choose both, we speculate that it may be because college students have little knowledge of professional financial websites, and at the same time, they rarely choose to view financial reports in their daily lives, the number of times they choose the two is 76 and 39 respectively.

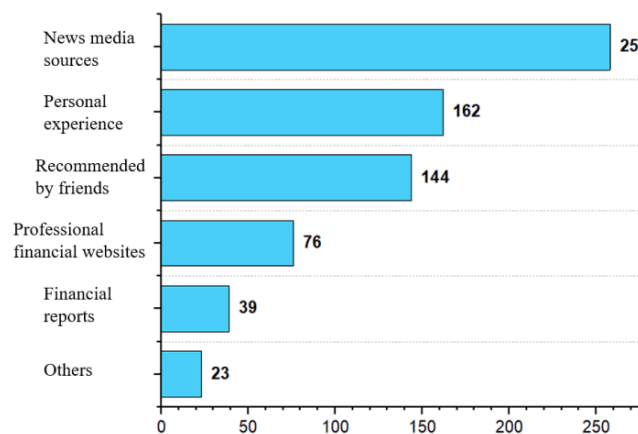
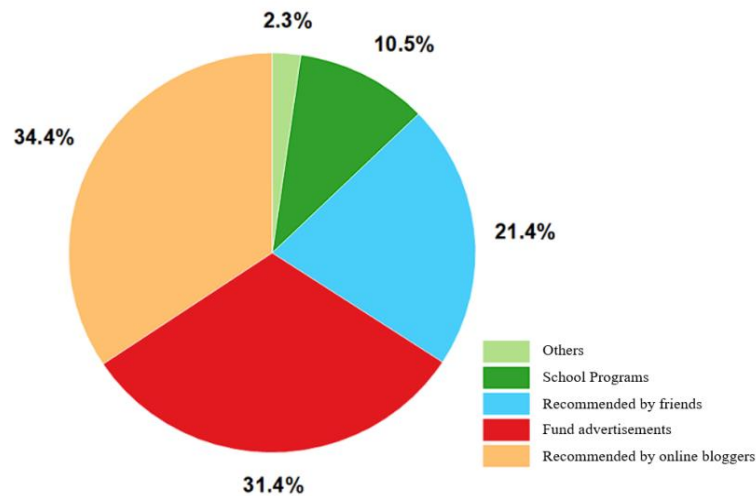


Figure 2. Sources of investment information bar chart

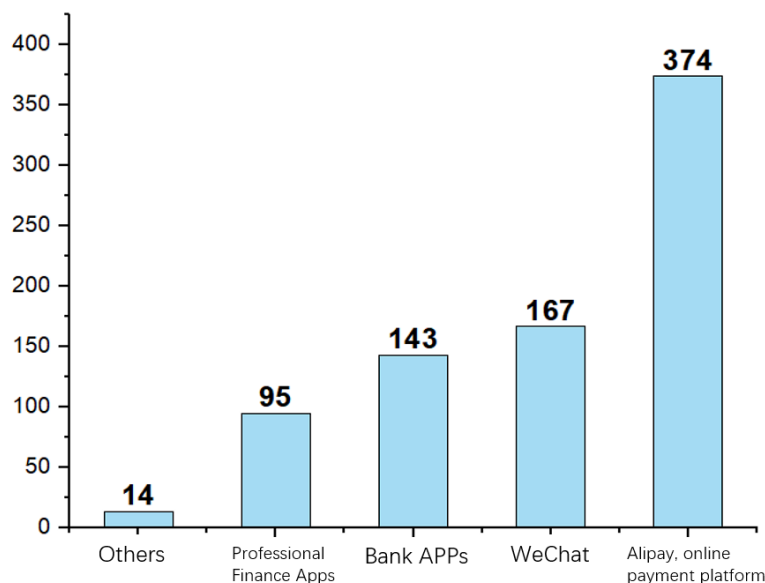
By exploring the reasons for college students in Pudong New Area to buy funds for the first time, we get the following figure 3, we find that the vast majority of consumers learn about funds for the first time through the recommendation of Internet bloggers, accounting for 34.4%. With the continuous development of the Internet, the term "fund" is tied to "profit", and through the publicity of Internet bloggers, more college students who lack economic knowledge have learned about funds, which promotes the trend of college students' fund purchasing. Fund-related advertisements also magnify the advantages of funds and attract new students to enter the fund market. Therefore, 31.4% of the college students who bought funds for the first time learned about the fund market from fund-related advertisements. Although school finance and economics courses mention a lot of fund content,

college students tend to use this content as a knowledge point to prepare for exams without actually learning about the fund market, so only 10.5% of college students made their first purchase after learning about funds through their school finance and economics courses.



**Figure 3.** Reasons for Purchasing Funds Fan Chart

Choosing the platform for purchasing funds is also one of the factors affecting consumers' purchase of funds, and in the survey process, we got the following figure 4 about the choice of platforms for college students to purchase funds; first of all, Alipay, as a necessary payment software, there is a certain convenience in the way of payment for purchasing funds from the platform, and at the same time, Alipay has created a special topic of funds for the convenience of the users to understand and purchase, so this also makes Alipay has become the choice of some college students who pursue simplicity, and we found that 374 or 70.3% of college students choose Alipay as the platform to buy funds, followed by various types of bank APPs; WeChat is the main chatting platform for college students, but the fund topics it opens also attract some college students to use WeChat to buy funds, which accounts for 31.4% of the total number of college students; compared with the other three platforms, professional Compared with the other three platforms, professional financial management software is chosen by fewer people, only 17.9% of college students choose it. We speculate that college students seldom take the initiative to know about professional financial management APPs, and they mainly figure out the easy interface of Alipay, and professional financial management software is not their main choice.



**Figure 4:** Bar chart of fund purchasing platforms

In order to explore the college students' choices of different types as well as risks of funds, we analyzed the cross-tabulation table of the two choices of the questionnaire's fund's risk preference choices and fund type choices, and the obtained matrices are shown in Table 2 below:

**Table 2:** Selection of different types of funds as well as risks

Risk appetite/Type of fund	Currency-based funds	Bond Fund	Equity funds	Hybrid funds	Total
Low risk, low return	32	86	20	112	250
Risk neutrality	34	64	35	15	148
High risk, High yield	17	18	69	30	134
Total	83	168	124	157	532

Firstly, it can be found that college students with high risk appetite are more willing to choose equity funds, because equity funds have higher volatility in the market but high returns, which satisfy the risk appetite needs of this kind of college students in terms of risk and return; secondly, risk-neutral college students will choose more money and equity funds, the objective of money funds is to maintain the stability of the capital and to provide relatively high liquidity; finally, the number of mixed funds chosen by college students with low risk and low return is the highest among these three categories. The risk and return of money funds is consistent with the expectations of this group of risk-averse college students; finally, college students with a low-risk, low-yield risk appetite choose hybrid funds in the highest numbers of all three groups. Hybrid funds have risks and returns that are intermediate between bond and equity funds, and they are designed to balance risk and return by diversifying across different markets. Hybrid funds are suitable for college investors who are looking for a diversified portfolio in one investment vehicle.

### 3.2. Purchase Decision Behavior Analysis

In order to further study the influencing factors of college students in purchasing funds, we use the decision tree model to further explore the college students' decision to purchase funds.

Take "whether to buy funds" as the dependent variable, take other questions as input variables, set the Gini coefficient as the judgment criterion, the smaller the Gini coefficient is, the smaller the decision instability is, and it is considered to determine the certainty of the decision. Using Jupyter Notebook to draw the decision tree and prune the branches, in the optimization process, the three variables that account for more weight are filtered out for the drawing of the decision tree, which are: the information selection channel, platform selection, and the investment amount of three variables. The obtained decision tree is shown in Figure 4 below. It can be seen from the decision tree diagram:

- 1) Paypal is a key factor influencing college students' decision-making on fund information;
- 2) The reasons why college students choose to buy funds are influenced by the factors of "optimistic about the fund manager's reputation" and "hedging financial risks", and they also choose to invest between 2001-3000 RMB in funds;
- 3) College students who buy funds are also influenced by diversifying their finances and specialized financial software, while investing 1,001-2,000RMB;
- (4) "All kinds of bank APP" will also affect college students' choice of fund purchase, and most of the college students usually do not choose all kinds of bank APP in the case of low investment amount.

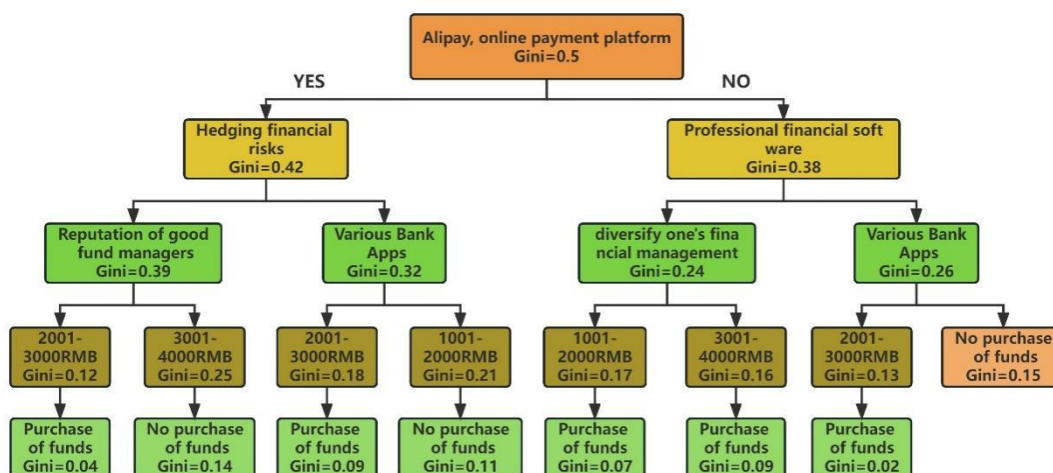


Figure 5: Decision Map for Fund Information Selection

### 3.3. Data Crawling

In this study, the steps of crawling Zhihu comments are divided into four parts: requesting data, obtaining data, parsing data, and keeping data, and the Python automation program is used to crawl all the comments and topics and generate CSV table files and TXT text files to save the obtained data. The browser chosen for crawling is Google Chrome, and the chromedriver program is used to facilitate the extraction of web page information. Import selenium, webdriver package, the use of python automation to realize the computer from the web page request data to obtain data process, the use of time.sleep statement to set the interval between crawling, to prevent the web page is not loaded in time to prevent the automation of the program error at the same time, and set up if there is an error, then print the statement of 1 to check the process of crawling to check whether there is an error in the process of crawling in order to facilitate the timely stop of the Automation program. When the crawl is successful, the automation stops, and the crawled comments are saved in TXT files and CSV tables for subsequent analysis.

### 3.4. Data pre-processing

In order to ensure that the collected data is valuable to the research problem, some useless data need to be processed to reduce the interference of irrelevant data. By analyzing the comments of Zhihu fund purchase, we can find that the article needs to exclude the comments of a user repeated many times and the comments of different users which are exactly the same, the reason is that in order to earn cash back and platform points and other interests, some users will fill in the false comments into the comment area, which will cause "false high" interference to the statistics of real positive and negative comments, therefore, we will exclude the repeated comment records and will not keep one of the repeated records as the record of the user as the user. The reason is that in order to earn cash back and platform points and other benefits, some users will fill in false comments to the comment area, which will cause "false high" interference to the statistics of real positive and negative comments, so we will exclude the duplicate comment records, and will not keep one of the user's duplicated records as the user's comments on a certain topic data. The pre-processed data will be saved as TXT text and CSV table files for further use. A total of more than 1,000 comments were counted.

### 3.5. Text Segmentation and Word Cloud Analysis

Through the Chinese Academy of Sciences' lexical tool ICTCLAS, these comments were lexically separated and labeled, and the sentences were divided according to the Chinese period, exclamation mark, question mark, and the sentence breaks at the end of a comment without a sentence break were added as sentence breaks and numbered with the HowNet lexicon. Compare the word labeled



can reflect that consumers will be affected emotionally by loss and crash when buying funds, and some consumers will also be affected emotionally by their loss after choosing fixed investment, and the handling fee is also one of the points leading to the negativity of some consumers; the keywords of Theme 2 are recommendation, try, free money, etc., which can reflect that consumers will be more satisfied with the effect of using free money to test the water through the recommendation of bloggers on Knowledge and other information channels, and even some consumers will make a lot of money from it. As well as other information channels to use free money to test the water, the effect makes consumers more satisfied, and even some consumers will make a lot of money from it, and at the same time, the third-party financial institutions will also make consumers feel more satisfied.

#### 4. Conclusions

In this study, the research target is determined as college students in Pudong New Area, Shanghai, and the data are obtained by distributing as well as collecting questionnaires. Based on the multimodal form to study the behavioral habits of the target audience about the direction of fund purchasing and the influencing factors, the following conclusions were finally obtained:

1) Social media communication has become a key factor influencing the purchase of funds. When purchasing funds, college students prioritize the initial understanding and analysis of funds from the level of news media and online bloggers, learn about the required fund products from the aspect of social media, and then choose the type of fund through the recommendation of online bloggers.

2) the shallow financial knowledge of college students to bear the ability to low, due to undergraduates for the financial knowledge of the involvement of the narrower, in the absence of adequate knowledge of the fund to understand the situation, most undergraduates follow their own will or in the case of less understanding of the situation of blindly purchasing funds, but also in favor of medium and high-risk investments, ultimately leading to the loss of property.

With the above conclusions, we can give the following recommendations from them:

1) Refuse to blindly follow a single source of information to buy a fund. After obtaining different sources of information, compare and analyze them to find out the similarities and differences and the key points, and then buy the relevant fund products according to your own needs and abilities.

2) Supplementary financial knowledge, the purchase of funds to return to rationality, first of all, to organize their own personal financial situation, to ensure that there is enough budget to buy funds; familiar with different types of funds, fund types are varied, you need to understand in detail the risk level and the expected return; regular attention to the fund market news, according to the market trend of the funds held to make adjustments to the strategy; to avoid undue anxiety, to maintain a rational avoid excessive anxiety, maintain rational thinking, do not cause internal conflict due to unknown external guidance, adopt a calm analytical approach, and should make decisions based on personal thinking and judgment.

#### References

- [1] Liu Yubin. Research on the Financial Behavior of Small Investors Based on the Internet Financial Era-- Taking Balance Treasure and Baidu Wealth Management as an Example [J]. *Economic and Trade Practice*, 2017(05): 113.
- [2] XU Sisi, YANG Binren, PING Na. Analysis of small amount financial management of college students in the era of internet finance [J]. *Value Engineering*, 2019, 38(27): 41-42. DOI: 10.14018/j.cnki.cn13-1085/n.2019.27.016.
- [3] Shou Zhenkun, Zheng Zhili, Bai Wenrong. Research on financial management behavior in the era of big data finance - based on college student group[J]. *Education and Teaching Forum*, 2020, No.452(06): 100-101.
- [4] Xu Wei, Han Yanru, Zhai Huimin. Research on investment behavior characteristics of college students and their demand influencing factors of Internet financial management products [J]. *Journal of Economic Research*, 2022, No.507(13): 99-101.

- [5] (2021). What influences the willingness of university students to invest in stocks? Godfred Matthew Yaw Owusu, Fred Kwasi Anokye, James Kwame Otieku Godfred Matthew Yaw Owusu, Fred Kwasi Anokye, James Kwame Otieku and Gabriel Sam Ahinfu Published Online:February 1, 2021pp 116-135 <https://doi.org/10.1504/IJEED.2021.114366>
- [6] Mandell, Lewis and Schmid Klein, Linda, The Impact of Financial Literacy Education on Subsequent Financial Behavior (2009). Journal of Financial Counseling and Planning, Vol. 20, No. 1, 2009, Available at SSRN: <https://ssrn.com/abstract=2224231>
- [7] Xiao Han, Mao Xuesong, Zhu Zed. A text similarity detection method based on HybridDL model [J]. Electronic Technology Application, 2020, 46(06): 28-31+35. DOI: 10.16157/j.issn.0258-7998.191257.
- [8] Zheng Yue, Qin Yangyang, Shi Yonghui et al. Analysis of the current status of online public opinion research [J]. Industrial Control Computer, 2023, 36(07): 100-102.
- [9] Zhang Xiaohu. Analysis of factors for confirmatory diagnosis of heart disease based on logistic regression and decision tree [J]. Modern Information Technology, 2023, 7(07): 117-119+123. DOI:10.19850/j.cnki.2096-4706.2023.07.029.
- [10] Wang ZY, He LL. Application of integrated machine learning to predictive maintenance datasets [J]. China Science and Technology Information, 2023(14): 112-114.