Data Analysis Based on Knowledge Payment Behavior of College Students

Qimiao Ma, Bo Tang*, Ou Lan, Zihao Wang, Xinbo Li, Jingke Yin

* Corresponding author: apriltb@163.com

School of Mathematics and Computing Sciences,
Guangxi Colleges and Universities Key Laboratory of Data Analysis and Computation,
Guilin University of Electronic Technology, Guilin, 541001, CHINA

Abstract: In view of the existing willingness to pay for knowledge, we carried out multiple Logistic regression analysis on the influencing factors of college students' ability to pay for Internet knowledge, and obtained the correlation between college students' willingness to pay for Internet knowledge and various factors. The results show that college students tend to choose content with better quality, more flexible service, high public evaluation and prominent brand effect when paying for online knowledge.

Keywords: knowledge payment, impact factors, Multinomial Logistic Regression Analysis

1. INTRODUCTION

In recent years, the knowledge economy industry has developed vigorously, and the online knowledge payment platform has emerged as the times require[1-4]. With the development of the era of knowledge payment, online knowledge payment has become a key industry in China. As the largest knowledge demander and the largest potential user in the knowledge payment industry, college students have an important impact on the development of online knowledge payment. Through the research of this project, it can provide reference for users to choose knowledge payment products in practice, and put forward some reasonable suggestions on how to improve college students' willingness to pay.

2. RESEARCH SCHEME

Through the research of this project, we want to give suggestions can be made for the knowledge payment platform and product operation, which will help improve the willingness of college students to pay for knowledge. At the same time, the research conclusion can also provide reference standards for college students' payment knowledge and enrich the theoretical knowledge of relevant research.

The research covered the following areas. First, we find out the factor with the highest frequency as the independent variable through data statistics, and analyze the main factors affecting college students' willingness to pay for online knowledge. Then, based on the theory of multivariate analysis, we analyze the relationship between the independent variable and the willingness to

ICBBEM 2023, May 19-21, Hangzhou, People's Republic of China
Copyright © 2023 EAI
DOI 10.4108/eai.19-5-2023.2334338

pay for knowledge, select measurement indicators, and analyze the attributes of influencing factors. According to the theoretical model constructed in the previous step, we took a social questionnaire to determine the content of the questionnaire and collected the questionnaire data. Finally, by analyzing the data and getting the results, we put forward the corresponding countermeasures and suggestions on the factors affecting the process of college students' knowledge payment.

3. RESEARCH PROGRESS

The group first conducted a questionnaire survey, designed a questionnaire and distributed it among college students. Through the analysis of the questionnaire data and hypothesis test, we found the factors that affect the willingness to pay for knowledge, and discussed the relationship between these factors and the role of college students in the purchase decision.

The reliability test of the questionnaire design quality refers to the analysis of the accuracy of the questionnaire measurement results[5-7]. We used the same questionnaire for the same person at different time points. The two questionnaires are separated by two weeks. The maximum difference between the results of the problem is 0.018, which is not more than 0.05, so it is more reliable.

The Cronbach's Alpha Coefficient method is used in this paper[8,9], and the range of reliability coefficient is [0, 1].

The formula 1 is as follows:

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum {S_i}^2}{{S_t}^2}\right) \tag{1}$$

It can be seen that Cronbach's Alpha coefficient method evaluates the consistency of each item in the scale. The greater the reliability coefficient, the higher the reliability of the measurement. A well-designed questionnaire should have a reliability coefficient of 0.80 or higher, and the acceptable range is 0.70 to 0.80. We should consider modifying the scale or adding or deleting items. The Cronbach's alpha coefficient of the questionnaire are shown in the table below (Table 1).

Table 1. Cronbach's alpha coefficient of the questionnaire.

Level	Cronbach's Alpha	Number of Items	Reliability Evaluation
1	0.986	11	Good
2	0.957	5	Good

Validity Analysis:

Validity is an important factor to measure the questionnaire. Reliability and efficiency are not necessarily high, but must be high. The standard validity is the "gold standard" based on the results of the classic scale. If the correlation coefficient P is high, the validity of the standard is good. The standard validity of this scale is as follows (Table 2).

Table 2. Scale of criterion validity.

Level	1	2
Correlation Coefficient	0.970	0.932
Confidence Level	0.000	0.000

In order to observe the structural validity, the KMO and Bartlett's tests are used to study the relationship between factors and measured items.

Table 3. Scale of KMO and Bartlett's tests.

Level	Coefficient	Validity Evaluation
1	0.903	Good
2	0.793	Appropriate

From above table (Table 3), KMO and Bartlett's tests coefficients of the two scales are both up 0.7, which is not the best, but the validity is still high, so it is more suitable.

Description analysis:

The basic information of the respondents is analyzed. The respondents are college students. The survey is divided into two parts: preliminary survey and formal survey. A total of 163 questionnaires were distributed in the school market, 152 of which were valid, with an effective response rate of 93.3%. The survey data shows that the proportion of men and women basically conforms to the ratio of 1:1, and there is no significant difference in the use of gender needs. Among them, the number of participants in each grade and major is basically balanced, reflecting the randomness of this experiment.

The next analysis of college students first use of knowledge to pay the reasons. According to the data collected, the reason is mainly reflected in the rich content resources. Application members can enjoy free content that ordinary users cannot access, you can enjoy content that ordinary users don't have access to. For college students, they often need to consult various materials. In the process, a lot of times will encounter search content quality is not good, can not meet their needs, or to find the information they need when the platform requires a fee to continue to browse. In such a paradoxical situation, most users will choose to pay a fee to unlock the information they need. Such a situation is exactly "Other products are not satisfactory" this kind of situation, which also shows that the survey results are true and credible.

It is also worth noting that the option "accidentally see, attract content" is checked the least. More than 75% of college students are more willing to accept recommendations from friends and paid products promoted by social media reflected in these two options "Friend recommendation" and "social media promotion", which just reflects the phenomenon that college students pay more attention to word of mouth and are more willing to follow the trend.

As for the use of knowledge payment platform, college students use online education, entertainment and social media the most, which also suits this group of college students. Study, social and recreational life dominate. According to the result of the survey, the biggest demand of college students is homework demand, which corresponds to the rich content resources in the factors of the first-use knowledge payment platform mentioned before. All the major platforms

have grasped the short-term urgent and effective demand of college students and launched the online knowledge payment model in various forms.

Chi-Square Test:

According to the proportion distribution of the multiple-choice questions, the chi-square goodness of fit test was used to analyze. This paper analyzes the reasons why users use the knowledge payment platform for the first time, and obtains the following table (Table 4):

Table 4. Summary of response rates and penetration rates for first-time user reasons for using the knowledge pay platform.

Project	Response		Universalization Rate (*n*=152)
	n	**Responsibility**	
1	103	37.60%	67.80%
2	90	32.80%	59.20%
3	4	1.50%	2.60%
4	54	19.70%	35.50%
5	23	8.40%	15.10%
Summary	274	100%	180.26%

Goodness of fit test: $\chi^2 = 130.562$ $p = 0.000$

The above table (Table 4) shows that goodness of fit test shows significant (chi = 130.562, p = 0.000.05), which means that the selection proportion of each item has significant difference, which can be compared by response rate or popularity rate. Specifically, options 1 and 2 had significantly higher response rates and penetration rates. The online education has a higher response rate and penetration rate, and the impact is more obvious.

Based on the analysis of the user's drive to use the knowledge-based payment platform, the following table:

Table 5. Summary of response rates and penetration rates for user reasons for using the knowledge pay platform.

Project	Response		Universalization Rate (*n*=152)
	n	**Responsibility**	
1	132	45.36%	86.84%
2	85	29.21%	55.92%
3	29	9.97%	19.08%
4	45	15.46%	29.61%
Summary	291	100%	191.45%

From the above table (Table 5), the goodness of fit test shows significant difference (chi=87.213, p=0.000.05), which means that the selection rate of each item is significantly different, which can be compared by response rate or popularity rate. Specifically, the response rate and permeability of alternatives 1 and 2 are significantly higher.

Through the above analysis, we can conclude that college students are more inclined to pay for knowledge with rich content resources, and are more willing to pay for knowledge for online

education. Below we take an English online learning APP as an example, the analysis of its user data, further determine the impact of the factors of knowledge payment.

Figure 1. A cloud of user comments about the APP

In this paper, the word cloud statistics of the student evaluation of the English learning APP is carried out, and the following word cloud map is made as Figure 1. As you can see, people are more willing to pay for easy, fun, and informative learning. For the students who have paid, we have carried out evaluation statistics, as can be seen the main reason for everyone to pay is that you can use fragmented time to learn spoken language anytime and anywhere, which is the choice of over 80% of students. This performance reflects the attention to learning time is high. Because the APP makes use of fragmented time, its billing rate is at a high level in similar paid apps. For all students who have tried or paid to study, for the purpose of analysis, we can find that most students aim to improve their oral ability.

4. BASED ON LOGISTIC REGRESSION ANALYSIS OF FACTORS INFLUENCING COLLEGE STUDENTS' WILLINGNESS TO PAY FOR ONLINE KNOWLEDGE

This paper studies the correlation between the willingness of college students to pay for online knowledge and some factors, so we can use the regression model. Because "Whether college students are willing to pay for online knowledge" is a binary discrete variable, so we choose the Logistic regression model for analysis.

The choice of explanatory and interpreted variables:

(1) To analyze the rationality of the problem

The problem of the ROC curve drawing, get as follows table (Table 6):

Table 6. ROC curve ANALYSIS TABLE

The result variable of the test	**AUC**	**Standard error**	**P**	**95%Confidence interval**	
				upper limit	**lower limit**
I think it's important for me that the paid-for-knowledge model can provide more quality content.	0.598	0.071	0.000 ***	0.459	0.737

I can get help from customer service if I have problems in the process of using knowledge to pay.	0.524	0.07	0.000 ***	0.387	0.662
My trial experience before deciding whether to buy a knowledge product will motivate me to buy it.	0.465	0.068	0.000 ***	0.332	0.598
Knowledge-based paid products with good word-of-mouth will motivate me to use them.	0.502	0.07	0.000 ***	0.366	0.639
Subscribing to expert recommendations in the column on the platform will motivate me to use it.	0.633	0.071	0.000 ***	0.494	0.771
People who influence me recommend that I use knowledge to pay for it.	0.511	0.07	0.000 ***	0.374	0.648
I think paying for knowledge is more than worth it.	0.555	0.071	0.000 ***	0.416	0.694
I think it's important to me that the paid-for-knowledge model provides a more flexible, alternative service.	0.48	0.069	NaN	0.345	0.615
I think paying for knowledge can satisfy my desire to a certain extent.	0.416	0.065	0.000 ***	0.288	0.544
I think the paid-for-knowledge model is very convenient.	0.58	0.071	0.000 ***	0.441	0.719
The people around me are paying for knowledge, so I'm paying for knowledge, too.	0.566	0.071	0.000 ***	0.427	0.705
I'm worried that paying for knowledge will put more pressure on my budget.	0.405	0.071	NaN	0.266	0.545
I'm afraid I can't get useful knowledge by paying for it.	0.521	0.069	NaN	0.387	0.656
I'm worried that I don't have enough time to digest the knowledge products I buy.	0.498	0.07	NaN	0.361	0.634
I'm worried that the knowledge I pay for has no impact on society.	0.349	0.07	0.000 ***	0.211	0.487
I'm afraid that paying for knowledge will add pressure to my life	0.47	0.07	0.000 ***	0.332	0.608

By analyzing the data, we could see that the AUC of the negative influence factors was much lower than that of the positive influence factors, so we excluded the negative influence factors in the follow-up study, start from the front for more accurate analysis.

(2) The explanatory variable

The explanatory variable is whether college students are willing to pay for online knowledge.

Whether college students are willing to pay for online knowledge reflects the degree of their acceptance of online knowledge. The binary choice behavior of the first individual, "Whether they are willing to pay for online knowledge", is expressed as the explanatory variable, when you select yes, Yi takes the value 1, and when you select no, it takes the value 0.

Explained variable is as follows formula 2(dependent variable):

$$y = \begin{cases} 1 & \text{Have paid for knowledge before} \\ 0 & \text{Didn't pay for knowledge} \end{cases} \quad (2)$$

(3) Explanatory variables

The explanatory variables is the basic situation of college students and the things they care about when they pay for online knowledge.

The basic personal information of the students is divided into three variables: gender, grade and monthly living expenses Identify "What you care about when you pay for online knowledge" as the fourth variable. In the questionnaire design, the four variables under personal information are given in the form of interval and "Yes" or "No", and a large number of collected data are qualitative variables. For example, the professional into 3 files, directly coded as 1,2,3. According to the research goal and the summary of the questionnaire data, the variable name and the variable setting table of the model are obtained (Table 7):

Table 7. Table of Explanatory variable (independent variable)

Explanatory variable	Variable symbol	Variable definition
Gender	gender	$gender = \begin{cases} 1 & Male \\ 0 & Female \end{cases}$
Grade	grade	$grade\_1 = \begin{cases} 1 & freshman \\ 0 & else \end{cases}$
		$grade\_2 = \begin{cases} 1 & sophomore \\ 0 & else \end{cases}$
		$grade\_3 = \begin{cases} 1 & junior \\ 0 & else \end{cases}$
		$grade\_4 = \begin{cases} 1 & senior \\ 0 & else \end{cases}$
Major	major	$major\_1 = \begin{cases} 1 & Science\ and\ engineering \\ 0 & else \end{cases}$
		$major\_2 = \begin{cases} 1 & Literature\ and\ history \\ 0 & else \end{cases}$
		$major\_3 = \begin{cases} 1 & Art \\ 0 & else \end{cases}$
Provide high quality and flexible service	10_2	
convenient charging mode	10_4	
Public influence	10_6	
Brand effect	10_9	
The constant term	intercept	

(4) Model establishment

According to the definitions of the explained variable and the explained variable, the binary Logistic regression model is constructed as follows:

$$\text{Logit}(Y) = \ln \frac{Y}{1- Y} = b_0 + b_1 x_1 + \mathbf{L} + b_{13} x_{13} = b x^T \tag{3}$$

$$Y = \frac{1}{1+ e^{b_0 + b_1 x_1 + \mathbf{L} + b_{13} x_{13}}} = \frac{1}{1+ e^{b x^T}} \tag{4}$$

In the above formula 3 and formula 4, Y_i is the interpreted variable and $\widehat{Y_i}$ is the fitted value of the interpreted variable, taking values between [0, 1] and usually with a threshold of 0.5. This paper compares the magnitude of the fitted value with the threshold value. The predicted Y_i 1 is assumed if the calculated fitting value is greater than 0.5, and the predicted Y_i= 0 is assumed if the fitting value is less than 0.5. X is the explanatory variable, and x_k corresponds to 13 dummy variables under the four variables of gender, grade, major, and Care, respectively. *i* (*i* = 1,2, ... ,*n*) represents the observed value, θ represents the regression coefficient to be estimated.

The table below (Table 8) is the parameter estimation, significance test and result of Logistic binary regression of college students' willingness to pay for online knowledge. According to the output of regression coefficients, most of the regression coefficients are significant ($p < 0.05$) . The positive and negative coefficients of the variables show that there is no significant difference between gender and willingness to pay for online knowledge, and there is a positive correlation between the grades of freshmen and seniors. Senior students are faced with graduate entrance examination or job hunting after graduation. They urgently need to improve all aspects of their personal abilities in a short time, so they turn to better quality online paid content; In professional fields, students in science, engineering, literature and history are more likely to pay for online knowledge, while art students are less willing to pay for online knowledge. In terms of college students' interest in online knowledge payment, college students are more likely to pay for online content with higher quality, more flexible service, high public evaluation and prominent brand effect, and vice versa.

Table 8. Table of parameter estimation and significance test results.

	coef	**P>|z|**
gender_1	1.9208	0.000
grade_2	1.0998	0.008
grade_3	0.9472	0.031
grade_4	2.8231	0.000
major_2	0.2311	0.472
major_3	-0.4990	0.614
10_2	0.3156	0.042
10_4	-0.2091	0.164
10_6	0.0514	0.689
10_9	0.1154	0.328
intercept	-2.8250	0.003

5. INSPIRATION AND ADVICE

According to the experimental results, college students' willingness to pay for knowledge tends to high-quality content and word-of-mouth. College students are a large group of online knowledge payment. Referring to college students, we can conclude that online knowledge payment platform should pay more attention to the quality of product content, and improve their competitiveness in the industry, in the quality of the promotion of good word-of-mouth formation.

ACKNOWLEDGMENTS: This work is partially supported by Guangxi Natural Science Foundation (No. 2020GXNSFBA238022), Guangxi college student Innovation Project (No. S202110595218) and Guangxi Key Laboratory of Cryptography and Information Security (No. GCIS201928).

REFERENCES

[1] Yu, Du. Will You Pay for Knowledge[J]. China's Foreign Trade, 2020, No. 579(03):50-51.

[2] Li Wu. Online payment platform for knowledge: Where to go [J]. Library and information science, 2018,000(003) : 2.

[3] Zhang Shuai, Wang Wentao, Jing Ulrich. A study on the influencing factors of users' online knowledge payment behavior [J] . Library and information work, 2017, 10(v. 61; No. 575): 95-101.

[4] Li Gang, Lu Yanqiang, Teng Shuyuan. Research on the payment behavior of users' online knowledge based on the theory of planned behavior [J] . Research in library science, 2018, 000(010): 49-60.

[5] Cai Shun, Shi hairong, Fu Xin, etc. A study on the factors influencing the sales of knowledge-based paid products: taking Zhihu Live as an example [J]. Journal of Management Engineering, 2019, 033(003): 71-83.

[6] Huang Yuting, Lu Qing, Susan Pan. Study on the influencing factors of knowledge payment willingness based on regression analysis [J] . Shanxi agricultural economics, 2019(24): 13-14.

[7] Hui ka-ming. Analysis of influencing factors of urban residents' willingness to participate in community based on Logistic regression analysis [J]. Statistics and management, 2018, No. 253(08): 103-108.

[8] Wei Xie, ZhiHao Zhang, Min Xiang. Research on the Perceived Value of College Students for Knowledge-Paying Products [J].Journal of Physics: Conference Series, 2019, 1168(3): 2-4.

[9] Jing Jia. Analysis and Research on Operating Mode of Voice Platform Based on Knowledge Payment[J].Open Journal of Social Sciences,2019, 07(8):397-404.