

Analysis on User Satisfaction on Helping Farmers Live based on Text Mining and Structural Equation Modeling

Xinyi Li¹, Zuxin Meng², Zhuoyi Li³, Yuhui Li^{4*}

¹2795452610@qq.com, ²mzxx0906@163.com, ³920655209@qq.com, ⁴515231978@qq.com

¹Applied statistics, Guilin University of Electronic Technology, Guilin, China

²Department of Financial Engineering, Guilin University of Electronic Technology, Guilin, China

³Applied statistics, Guilin University of Electronic Technology, Guilin, China

⁴Academy of Business, Guilin University of Electronic Technology, Guilin, China

Abstract: Live video streaming has become a popular means of merchandising, more and more farmers are trying to use commerce live streaming to sell their agricultural products. This paper uses text mining technology to design a questionnaire, identifies important and potential user groups through K-means clustering, and uses structural equation modeling to verify the structural relationship between the impact of live streaming, anchor, product and service on the satisfaction of farming live streaming users. The results show that: the satisfaction of users on Helping Farmers Live stay at a basic level, there exist enough space to improve it; their loyalty on Helping Farmers Live is still low, and the frequency of repurchase behavior is not high; the easy-to-use groups for Helping Farmers Live are young, middle-income and rural consumers.

Keywords: Helping Farmers Live; agricultural products; text-mining; K-means clustering; structural equation modeling.

1 Introduction

During the prevention and control of the New Crown epidemic, in order to solve the problem of stagnant agricultural products, major e-commerce companies and social media platforms, relying on the advantages of their platforms, have constructed a new Internet sales mode with live content production as the carrier, effectively integrating the sales demands of traditional agricultural production areas into new consumption scenarios, becoming an important force to help agricultural products go upstream and promote the upgrading of the agricultural industry.

In view of this, this paper uses Python information to extract 267,679 comments from the e-commerce platform and conduct text analysis from the perspective of live farming users, collects relevant data through a questionnaire survey on the consumption status and satisfaction of the

"agricultural products + live streaming" marketing model, and then uses K-means cluster analysis to K-means cluster analysis was used to identify important and potential user groups, and structural equation modeling was used to verify the structural relationship between live streaming, anchors, products and services on the satisfaction and loyalty of users of live farming. Theoretically supported suggestions are provided for the healthy development of the "agricultural products + live streaming" model at various levels, including government, enterprises and consumers.

2 Review of the literature

Based on the questions posed in this survey and the purpose of the research, this paper briefly reviews and summarizes the research findings in the following areas: the current state of development of live farming assistance, satisfaction with farming assistance, and consumer attitudes and needs.

Jian Chengyuan and Feng Sichao (2021) analyzed the chaos of live e-commerce in 2020 and believed that mainstream media, relying on their years of professional experience, could play to their strengths to further improve the standards in the field of live e-commerce [1].

Liu Bingli (2021) explored the problems of the live-streaming with goods model and raised the issues of false sales and product quality that arise in the sales process in response to the characteristics of agricultural products [2].

The duo of Meng Na and Zhao Fengqing (2020) argue from the perspective of the subject and from a holistic perspective that the live e-commerce model and a better connection between agricultural products and markets can change the way agriculture increases income in rural areas [3]

Based on stimulus organism response theory, Xu Yue et al (2021) constructed a chain theory model in which technology availability and anchor characteristics influence customer engagement and consumer trust, which in turn promote consumers' willingness to purchase agricultural products [4].

Zhuang Litong and Tang Antelope (2022) conducted a SWOT analysis of live agricultural products sales, explained the prospect of live agricultural products and proposed strategies related to vigorously developing diversified products, enhancing the professional quality of anchors and multi-channel development [5].

A summary of the above-mentioned literature reveals that it is difficult to reveal the characteristics and behavioral mechanisms of consumers' preferences for the marketing channel of "agricultural products + live streaming" at the micro level, and to portray the overall satisfaction and loyalty levels of consumers. It is difficult to establish a logical and consistent theoretical link between consumers' purchasing decisions, enterprises' marketing strategies and government support policies under the "agricultural products + live streaming" model. Therefore, this paper will investigate the above deficiencies.

3 Preparation of research

3.1 Questionnaire design based on Python information extraction and text analysis

Information extraction: Through Python, the comment data of agricultural products aiding agriculture on the e-commerce live broadcast platform is captured, and 267679 comments captured are written into the txt file, and the most effective user comment data is obtained by data cleaning and deduplication process, and comment effectiveness rate is 45.6%.

Analysis on emotional dictionaries by text mining: Using Python's library, a large-scale corpus is statistically analyzed, representative words are pre-selected as benchmark words by manual annotation, and then semantic similarity is calculated for the candidate words to obtain new sentiment words, thus extending the coverage of the sentiment lexicon.

After the pre-processed text comment is analyzed by discrimination, the text sentiment analysis results are obtained as figure 1:

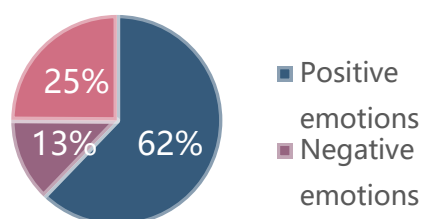


Figure 1: The results of emotional dictionaries by text mining

On the basis of literature review and Python information extraction and text analysis, a structured questionnaire of this study was formed. The questionnaire is divided into six sections: the first part is demographics, including gender, age, income, education, occupation, and hometown. The second part is the division of consumers, including whether they often live-stream purchases, the types of goods purchased, and why. The third part is correlation analysis, including the correlation between product attributes, product types, live venues and product seasons. The fourth part is the perception assessment of live broadcasts, anchors, products and services. The fifth and sixth parts of the balance are the measurement and exploration of user satisfaction and loyalty.

3.2 Data collection

In the formal survey, 1325 questionnaires were obtained through a combination of online and offline methods, and after excluding questionnaires with a duration of less than 30s, 1233 valid questionnaires (questionnaire validity of 93.06%) were obtained. After reliability test ($\alpha = 0.952 > 0.9$), validity test (KMO coefficient = $0.939 > 0.9$), random inspection (progressive significance of $0.319 > 0.05$), the results show that the data samples are reliable, valid, random and of good quality.

4 Analysis on user satisfaction

4.1 Potential user mining based on K-means clustering

Selection of clustering factors: Cluster analysis is conducted for potential users, and the indexes for cluster analysis are determined as follows: gender, age, educational background, monthly income and purchase factors (shown as table 1). These five indicators are represented by S, A, I, H and P respectively as the indicators to identify customer value and are recorded as SAIHP model.

Model construction and testing

Table 1: Cluster center of potential customer (PC) types

Types of PC	I	II	III	IV	V
purchase intention	five	three	four	one	three
gender	female	male	male	female	female
age	under 22 years	over 50 years	22-35	22-35	over 50 years
Income level	under 1500	under 1500	5000-8000	1500-3000	more than 8000
hometown	rural area	rural area	rural area	town	town

Table 2: ANVOA analysis results

	cluster		error		F	P
	mean square	d.f	mean square	d.f.		
Gender (S)	4.22	4	.212	625	19.9	0.00
Age (A)	143.8	4	.215	625	667.7	0.00
Income level (I)	122.7	4	.300	625	409.1	0.00
Hometown (H)	3.5	4	.229	625	15.4	0.00
Purchase intention (P)	42.0	4	.202	625	201.9	0.00

Analysis of clustering results: According to the clustering results shown as table 2, the characteristics of each cluster group are summarized by analyzing the situation of five user groups in each attribute.

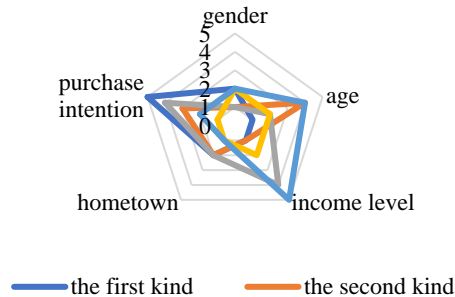


Figure 2: Analysis diagram of potential users' characteristics

Based on this characteristic description, we classify potential customers into five categories: important leads, important development leads, secondary leads, and general and low value leads. The characteristics of each potential customer category are shown as figure 2:

Potential users are low-income, rural, female students, development users are middle-income, rural areas, young and middle-aged men. The value analysis of potential users are as follows: first, Help Farmers Live can refine the positioning of potential customers by analyzing the characteristics of user groups, and also make development direction for potential customers to better meet their needs. Second, easy to use crowd characteristics clear. Third, Help Farmers Live companies should increase the advertising into more energy and money, increase publicity, so more people pay attention to Help Farmers Live.

4.2 Customer satisfaction and loyalty based on structural equation model

Model assumption: According to the research object and content, we set 6 potential variables in the model: anchor, product, live broadcast, service, satisfaction and loyalty, which influence and restrict each other, and establish the following hypotheses:

- H₁: "Anchor" has an impact on "satisfaction";
- H₂: "Live broadcast" has an impact on "satisfaction";
- H₃: "Product" has an impact on "satisfaction";
- H₄: "Service" affects "satisfaction";
- H₅: "Satisfaction" has an impact on "loyalty";

Structural Equation Modeling: Preliminary model fitting was carried out, inappropriate relevant dominant indicators were removed, relevant parameter estimation was continuously revised, and the following structural model was finally determined (shown as figure 3):

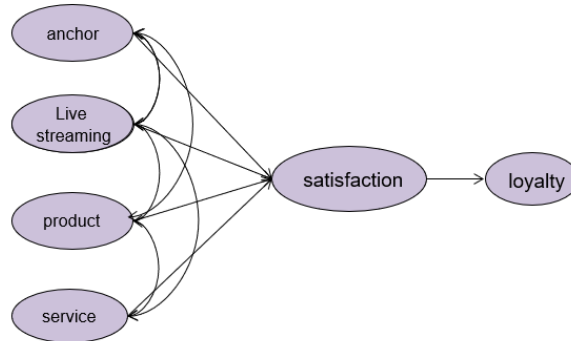


Figure 3: Estimation results of structural equation model

After standardization, the load coefficients between potential variables and observed variables are mostly above 0.7, which proves that there is a strong correlation between potential variables and observed variables.

Model test: The Critical Ratio (CR) and P value obtained by AMOS can be used to test the significance of path coefficient or load coefficient. The fitting results of model parameters are shown in the table 3.

Table 3: Regression coefficients of the model

X	→	Y	SRC	SE	CR	p
anchor	→	satisfaction	0.67	0.243	3.78	0.000
live streaming	→	satisfaction	0.68	0.116	12.52	0.001
product	→	satisfaction	0.63	0.324	8.99	0.000
service	→	satisfaction	0.60	0.132	17.88	0.000
satisfaction	→	satisfaction	0.74	0.395	22.70	0.000

According to the analysis of the results, it can be found that when P value <0.05 and CR>2, the two factors are considered to have an influence on each other, and the path set by the model is reasonable.

5 Conclusions

The results show that:(1) the satisfaction of users on Helping Farmers Live stay at a basic level, there exists enough space to improve it;(2) their loyalty on Helping Farmers Live is still low,

and the frequency of repurchase behavior is not high; (3) the easy-to-use group for Helping Farmers Live are young, middle-income and rural consumers.

With these results, we provide suggestions for improvement from three perspectives as follows. From the perspective of the government, it is helpful to broaden the new model of live broadcast by government officials to help farmers, with the collaborative participation of multiple types of subjects; create a complete industry chain and improve infrastructure construction; promote the standardized development of the live broadcast with goods industry and establish a long-term supervision mechanism. From the perspective of the anchor, they should accelerate the cultivation of live farming talents to continuously increase value creation; improve social interaction between live broadcasters and showcase commodities in multiple dimensions. From the perspective of enterprise, they need to identify the target consumer group and carry out precise marketing with local characteristics; establish a reasonable after-sales service management system and system to ensure the integrity of the shopping chain; improve the supply chain system for live-streaming with goods to help farmers and enhance the level of information technology.

Acknowledgements: This work was supported by the Guangxi Science and Technology Base and Talent Program (No.AD19245100), the Guangxi Key Laboratory of Cryptography and Information Security (No. GCIS201818), and the Graduate research and innovation project of Guilin University of Electronic Technology (No.2021YCXS089).

References

- [1] Jian Chengyuan, Feng Sichao. (2021). Network integrity and ecological self-cleaning in the new media environment[J]. *News Lovers*. 10, 33-35.
- [2] Liu Bingli. (2021). Problems and governance countermeasures of direct broadcasting with goods[J]. *Management and Administration*. 12, 10-12.
- [3] Meng Na, Zhao Fengqing. (2020). A study on the role of live prefects' e-commerce in rural consumption poverty alleviation[J]. *New Media Research*. 9, 47-48.
- [4] Xu Yue, Zheng Fuyuan, Chen Weiping. (2021). The Influence of Technology Availability and Anchor Characteristics on Consumers' Willingness to Purchase Agricultural Products[J]. *Rural Economy*. 11, 104-113.
- [5] Zhuang Litong, Tang Antelope. (2022). Research on the development strategy of live sales of agricultural products in the context of rural revitalization strategy[J]. *China Market*. 14, 139-140.