Research on Online Stream Media Marketing Strategy Based on Data Mining

Yihan Jia
Beijing Institute of Fashion Technology, China

Keywords: Data Mining, Marketing Strategy, Machine Learning, Live Streamers.

Abstracts: With the rapid development of science and technology and economy in China, the online stream industry relying on the Internet platforms is becoming more and more prosperous. As a new media industry, online stream industry has characteristics such as low entry thresholds, image virtualization, function positioning, and entertainment positioning. This paper uses two machine learning algorithms to analyze the potential relationship between the characteristics of network anchors and the sales volume, and makes descriptive statistical analysis and correlation test with the sales volume. A sales evaluation model is then built to adjust the precision marketing strategy based on these characteristics. It is expected to provide a more effective way to increase the income of enterprises and promote the economic development of online stream industry.

1 INTRODUCTION

"In 2021, the number of webcast users in China has reached 635 million, and it continues to rise." The statistics of China's Internet Network Information Center are enough to prove that the online stream industry is creating myths. The success of Li Ziqi, Li Jiaqi, Feng Timo and others has made more and more Chinese people imagine starting from scratch to becoming rich. At present, live streamers have become a "phenomenal" profession. Streamers in different fields transmit and disseminate relevant content information. It can be said that the commercial value of the live streamers online stream industry is becoming increasingly apparent, and it has become an important link in the business chain of enterprises to carry out brand building, consumption insight and marketing promotion, and is a sharp tool for contemporary corporate marketing.

This article will study the influence of the behavioral characteristics of live streaming e-commerce on its sales volume. On the other hand, starting from sales and behavior data, combined with the differentiated needs of groups, segmenting public customer groups, and finally quantifying the evaluation index system. The primary research topic is how to measure and obtain behavioral characteristics and sales data of live streaming e-commerce [1]. The second thing to do is to extract and analyze the behaviors of goods, activities and topics released by live streamers, and conduct in-depth research on the logic and laws behind the behaviors; The third is based on the discovery of the second point, a behavioral feature database is constructed, and a machine learning model is constructed based on the basic characteristics of Internet celebrities to explore the relationship between each behavioral characteristics of Internet celebrities and sales levels.

Finally, rationalization suggestions are put forward, hoping to provide some theoretical references for online celebrities' own operations, platforms and merchants when choosing resources for cooperation.

2 ELEMENTS AND CHARACTERISTICS OF ONLINE LIVE STREAMING PLATFORMS

2.1 Analysis of the Characteristics of Online Live Streaming Platforms

As a live streaming carrier, the live streaming platforms is a basic and important part of the online stream industry. The live streaming platforms uses the media characteristics of digital media to combine

Jia, Y.
Research on Online Stream Media Marketing Strategy Based on Data Mining.
DOI: 10.5220/0011913500003613
In *Proceedings of the 2nd International Conference on New Media Development and Modernized Education (NMDME 2022)*, pages 440-444
ISBN: 978-989-758-630-9

live streaming, small video, pictures, animation and audio into a new form of streaming. In terms of communication effect, it is more immersive and interactive, and it is more likely to be loved by everyone. It is worth mentioning that the development and improvement of any emerging things requires a long-term process, especially in today's prevailing digital media era, online live streaming as an emerging thing still has a lot of room for development, and the country's regulatory mechanism is relatively not mature and perfect.

2.2 Image Characteristics of Live Streamers

Live streamers are the main force in the online stream industry of live streamers from the perspective of digital media, and the image construction of streamers plays an important role in enhancing their commercial value. The streaming content of entertainment streamers is biased towards real-time entertainment consumption, mainly through the streamers's appearance, talent and fan interaction to attract the attention of users, obtain rewards from fans, and make profits by commercial activities and product promotion after accumulating a certain fan base. E-commerce streamers are similar to the Key Opinion Leader of the segment, carry out product promotion in the live streaming room, and comprehensively show the advantages of the product and promote sales through the streamers's personal tasting, try-on, and trial (GUO 2020).

The powerful development of live streaming is indispensable to the boost of brokerage companies. MCN (MCN, Multi-Channel Network) is an emerging business model and Internet format, which is a new thing brought by online media live streaming marketing and live streaming e-commerce, and is also the inevitable result of Internet specialization and segmentation.

3 RESEARCH ON THE MARKETING STRATEGY OF LIVE STREAMERS LIVE MEDIA FROM THE PERSPECTIVE OF DIGITAL INTELLIGENCE

3.1 Data Acquisition and Feature Extraction

Sources of Data: Considering the selection of market segments and the definition of research problems, this article will select the top 102 streamers who appeared on the daily sales list of Douyin in September from the third-party data detection platform as the analysis objects, match their Douyin data, and analyze their personal data information and behavioral data information (JIN 2020).

Acquisition of Data: Combined with the list of Internet celebrities that have been obtained, this article collects and organizes the data through Python scripts (Serdar 2017).

3.2 Sales Evaluation Model

3.2.1 Construction of Live Streamers Features

Before constructing the model, the correlation between each feature quantity and sales volume is directly analyzed, and after the normality test of the sales data, the sales data does not conform to the normal distribution, so the Spearman correlation coefficient is used in the correlation test of each feature quantity (Xie 2019). (See Table 1)

Table 1 Correlation test between the characteristic quantity of goods and sales

Feature classification	Feature volume	Correlation coefficient
Characteristics of the behavior of bringing goods	Number of videos	-0.015
	Video ratio	-0.089
	Number of video reviews	-0.125
	Number of video forwards	-0.114
	Number of video likes	0.078

3.2.2 The K-NN Model Evaluates Sales

K-NN algorithm model construction

The K-NN model is mainly used in the classification of feature space, so the K-NN model mainly has the following three points: distance measurement, k-value selection, and classification decision rules (Xi 2017).

Distance metric: Suppose the feature space X is the n-dimensional real vector space Rn, $x_i, x_j \in$ X,

$$K_P(x_i, X_2') = \left(\Sigma_L^n = \left\|X_i^1 l' - x_j(l)\right\|^3\right)^{\frac{1}{P}} \quad (1)$$

For the adjacent k training instance points forming the set N, the misclassification rate is:

$$\sum xi \in Nk(x)\ I(yi \neq cj) = 1 - k1\ \sum xi \in Nk(x)\ I(yi = cj) \quad (2)$$

In practical applications, if the value of k value selected is small, it is equivalent to selecting the training data in a smaller field for classification prediction, and the approximate error of machine learning will be small（Yu2004）; Therefore, only data sets with inputs close to the actual amount of data will have an effect on the predicted results of the variables; If the value of k value is large, the learning error will decrease, but the approximate error will increase, so that the prediction effect will not play a role in predicting the distant instance points, making the prediction result wrong.

3.2.3 K-NN Model Evaluation Results

In table 2, the accuracy of the simple marketing feature model and the personal attribute feature model are 0.69 and 0.68, respectively, indicating that relying solely on marketing characteristics and personal characteristics of Internet celebrities cannot comprehensively measure the ability of Internet celebrities to bring goods (Sheng 2019).

Table 2 K-NN Evaluation indicator output results

	Full-time full-feature model	Personal attribute characteristic model	Daily behavior characteristic model	Full feature model of marketing month	Single marketing feature model
The recall rate	0.44	0.54	0.41	0.39	0.54
Accurate rate	0.71	0.78	0.68	0.65	0.61
F1	0.41	0.36	0.49	0.41	0.39
AUC	0.39	0.41	0.55	0.42	0.50

The daily behavior characteristics model of Internet celebrities, the accuracy rate is 0.75, indicating that the daily sharing behavior and past behavior characteristics of Internet celebrities can also act on their sales at the same time, so e-commerce Internet celebrities should not only pay attention to their marketing behavior, but also pay attention to the accumulation of their daily behavior and the maintenance of fan relationships.

3.3 Logistic Regression Model to Evaluate Sales

3.3.1 Logistic Regression Model Construction

$$\text{Sigmoid}\ (p) = log e\ (p1\text{-}p) \quad (3)$$

Sigmoid(p)=m, there is:

$$p = \frac{1}{1+e^{-m}} \quad (4)$$

The correlation test of the 55 eigens obtained in the previous article found that the correlation coefficient between some independent variables was greater than 0.8, and there was obvious collinearity, so the factor analysis was used to reduce the dimensionality of the data, the value of KMO was 0.714, according to the KMO metric, the original variable was suitable for factor analysis, and the probability of Bartlett's sphericity test was 0.000, rejecting the null hypothesis. Based on the total explanatory variance, when the number of factors is 12, the cumulative percentage of variance is 84.459%, so new variables are generated to be placed into the logistic regression model for sales level prediction.

After each feature is multiplied by a regression coefficient, the result of multiplying the feature and the regression coefficient is added into the Sigmoid function to obtain the value between 0 and 1. If the data result is greater than 0.5, it is divided into class 1; if the data result is less than 0.5, it is class 0. (See table 3)

Table 3 KMO and Bartlett Test

KMO sampling suitability quantity	The approximate chi-square	0.714
Bartlett's sphericity test	Degrees of freedom	17653.148
	Significant	578 0

3.3.2 Logistic Regression Model Evaluation

Through the introduction of the principle and algorithm of logistic regression above, this section still uses python's sklearn to put the feature quantity after dimensionality reduction into the model to estimate the sales level and evaluate the predicted results. The evaluation indicators are as follows:

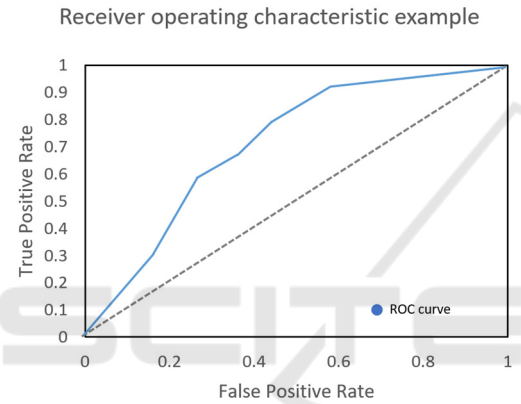

Fig 1 Logistic regression ROC curve

1. The accuracy rate, recall rate and F1_score of dichotomy were 0.85, 0.64 and 0.68 respectively. The accuracy rate of the model is slightly lower than that of the full-feature model of the marketing month with the highest accuracy in the KNN evaluation model. It can be seen that the evaluation effect of KNN is better than that of the logistic regression model when there is a large number of features.

2. The ROC curve of the logistic regression model is shown in Figure 1 below, and its AUC value is 0.85, slightly lower than the full feature of the marketing month in the KNN model. Therefore, under this classification, the KNN evaluation effect of the data cited in this paper is better than that of the logistic regression model.

Through the K-NN algorithm and logistic regression to evaluate the sales of Internet celebrities, and give the evaluation effect, collated to say that the classification effect of logistic regression under this classification is higher, the KNN model for each group of comparative experiments finally reached the same conclusion, that is, the marketing characteristics of Internet celebrities and personal characteristics information complement each other and act together on the sales of Internet celebrities.

3.4 Problems in the Marketing Process of Bringing Goods

Data falsification, loss of audience trust: Nowadays, business transactions such as powder increase, brush volume, interaction, and likes have become a gray industry in the industry. And developed functions such as commenting, liking, and interacting on the corresponding software according to the rules of each live streaming platforms, and it was even called "shaping the perfect live streaming room data". Brushing traffic and popularity are of great benefit to streamers and platforms to some extent, so some platforms will also take the initiative to carry out capital operations to improve the traffic data in the live streaming room, so a large amount of spam data and fake streamers will appear frequently, and will eventually lose the trust of the audience.

Fakes are frequent: Some streamers one-sidedly pursue commercial interests in the live streaming room, especially on some small-scale and poorly regulated live streaming platforms, and do not fully verify the quality and efficacy of the promoted goods, resulting in the continuous emergence of news about the sale of "three no products" and "fake goods" in the live streaming room, and in the short term, businesses and streamers have profited, but in the long run, it has seriously affected the image of live streamers in the minds of the audience, disintegrated the trust of the audience, and adversely affected the harmonious and orderly development of the Internet business environment. This has led to damage to the overall image of the online stream industry.

The number of fans, set the amount of praise, praise powder ratio are ranked in the top ten characteristics of the importance of live streamers, the number of fans represents the personal influence of live streamers, set the amount of praise, praise powder ratio represents the quality of the work of live streamers, and the activity of fans, therefore, enterprise marketing should pay more attention to the personal influence of live streamers when choosing cooperative resources, Internet celebrities and MCN organizations should also pay attention to the quality of their works in the daily operation process, and create more works with high originality and high quality, so as to enhance the interaction with fans and enhance the activity and stickiness of fans.

4 CONCLUSIONS

This paper classifies the characteristics of Internet celebrities' marketing behavior, personal attribute information and other characteristics, and analyzes the correlation between each characteristic quantity and sales volume. It is found that the number of livestream-related videos and the number of fans in the marketing month have a high correlation with the sales volume of e-commerce Internet celebrities, but the correlation is still weak, which is not enough to use a single variable to predict sales volume.

Based on various data reports and third-party monitoring platforms, the list of e-commerce influencers on Douyin platform that meet the research conditions was obtained. Based on data crawler and manual recording methods, the personal information and behavioral data of these influencers on Douyin platform were obtained as the main research objects.

According to the evaluation results of the previous model, the marketing characteristics and personal attributes of the month complement each other and jointly affect the sales volume of live streamers. Therefore, when choosing resource cooperation, enterprises or businesses should not only consider the current marketing methods of Internet celebrities, but also refer to their historical behavior data and personal influence. Internet celebrities and MCN organizations should also pay attention to the accumulation of daily behaviors and the improvement of work quality in the daily operation process, so as to increase the fans' stickiness and promote the growth of their commercial value.

It is necessary to strengthen the image construction of the streamers themselves in order to make great progress in live streaming. Only by adhering to professional ethics, adhering to the bottom line of morality, disseminating high-quality content, establishing a positive image among netizens, and guiding the public to establish healthy consumption concepts and aesthetic appeals, can streamers positively promote the green development of the online stream industry, and can they also maximize the commercial value of live streamers and form a good Internet "business ecosystem".

REFERENCES

Guo Peilin, Lian Yueqi. Analysis of the impact of online celebrity e-commerce on the purchase decision process in the Internet environment--to Douyin platform as an example [J]. Modern Marketing (Business Edition), 2020, 4:76-77

Jin Lin. Research on User Behavior Analysis Based on Data Mining [J]. E-Commerce, 2020(04):41-42.

Sheng Ye. Analysis and prediction of business impact of social behavior of live streamers [D]. Beijing: Beijing Jiaotong University, 2019.

Serdar Oktay .An analytical study to identify and determine the usage frequency of sales and marketing strategies for 5 star hotels in the Antalya region [J]. ProcediaComputer Science, 2017, 120.

Xie Huimin. Research on Marketing Management of E-commerce Enterprises Based on Data Mining [J]. Brand, 2019, 000(016):35-36.

Xi Y .Leung, Billy Bai, Mehmet Erdem .Hotel social media marketing: a study on message strategy and its effectiveness [J]. Journal of Hospitality and Tourism Technology, 2017, 8(2)

YU Li, LIU Lu, LUO Zhanghua. Comparative analysis of e-commerce recommendation strategy in China [J]. Systems Engineering-Theory & Practice, 2004(8):96-101.)