

# Research on the Youth Group's Expectations for the Future Development of self-Media while in the Digital Economy

Xuan Wang<sup>1, a</sup>, Zhihui Yang<sup>1, b, \*</sup>

<sup>1</sup>Anhui University of Finance and Economics, Bengbu, 233000, China

<sup>a</sup>xuan\_wang0918@163.com, <sup>b</sup>zhi-hui-yang@163.com

**Abstract:** Based on the survey data on the influence of self-media on youth values, this article analyzes the variance of different variables with academic qualifications as the main factor. Use the Sklearn tool to establish a decision tree classification model. The results show that the educational level factor is inversely proportional to the proportion of learning behaviors accumulating knowledge through self-media platforms. Different educational backgrounds have significant differences in the judgment of the authenticity of information cognition. Respondents' views on the future development of public opinion-oriented self-media mainly depend on their daily browsing frequency. Respondents with a very high frequency will make different judgments because of the inverse proportion of academic level and credibility. Finally, this article gives opinions and suggestions on the development of the self-media from three perspectives: the government level, the social level, and the youth group themselves. Encourage youth groups to actively face the opportunities and challenges brought about by developing the self-media industry in the digital economy.

**Keywords:** Analysis of variance, Self-media, Digital economy, Decision tree classification.

## 1. Introduction

With the advent of the digital age, people's lives are becoming increasingly intelligent and personalized. Massive data networks support the construction and development of artificial intelligence, blockchain, big data, and cloud computing applications. If the digital age is divided according to action progress, it can be distinguished explicitly into a primary digital age and a secondary digital age. Information is mainly disseminated via newspapers, radio, and television in the direct digital era. The secondary digital age includes the Internet, satellite television, and mobile communications. The third information age consists of the primary and secondary digital ages [1].

As a unique product of the digital age, the digital economy is also the most dynamic part of the country's economic development. It plays a vital role in stimulating demand and consumption, boosting internal and external investment, creating employment opportunities, etc. It is also generating and aggregating various innovative elements, injecting new momentum into the traditional economy and becoming an important driving force for national economic development [2]. Digital finance focuses more on the scale effect and tail effect than conventional financial models, with massive databases and optimization algorithms reducing the marginal costs of related businesses in traditional finance and an internet-based business model that allows digital finance to reach beyond the limits of time and space [3]. Freedom from these constraints means that the low productivity of labor and the restrictive nature of trade distances are diminished. The size of the audience for disseminating information and the range of audience areas have almost the exact cost [4].

After years of development, the digital economy has become relevant to people's lives. As a representative of the new financial model, the development of digital finance provides unique opportunities for solving the financing of

small and medium-sized enterprises and promoting the development of innovation and entrepreneurship [5]. In comparison with the impact of the SARS epidemic in 2003, the smooth running of anti-epidemic activities during the New Crown epidemic has also clearly benefited from the rapid development of the digital economy in recent years [6]. The story of digital finance has helped to improve the entrepreneurial behavior of rural residents and has brought about an equalization of entrepreneurial opportunities [7].

In its most common form, the digital economy is present in people's daily lives through big data and information technology, combined with social networks. Although policy resources and economic support have been tilted towards newspapers, magazines, radio, and television as traditional media in the form of support, the new media, which includes online digital media and self-publishing platforms, still has a considerable communication advantage. As the director of media development in the digital age, new media has also been widely recognized [8]. In 2012, the popularity of the 3G and 4G era and the intellectual development of mobile devices promoted the further development of the self-media industry. The massive volume of users drove a large amount of capital to enter the self-media market. As a product of the information revolution, the arrival of the self-media era has transformed our way of life, making access to information faster and more convenient [9]. With the arrival of the 5G era, the self-media industry will enter the following new stage of development, and the development of the self-media field will be more closely integrated with the industry field.

Youth groups are widely active in self-media platforms, and this generation is very closely connected to self-media. Today, when the construction of self-media platforms is not yet perfected, young people should become the masters of self-media platforms, draw nutrition from them and develop their horizons instead of being influenced by their undesirable factors at will, blindly following the crowd and losing themselves in the ocean of information. According to Liu

Wenrong's [10] survey report, the value structure of young people places more emphasis on the interests of others and society, maintaining social order and harmony, and restraining their individuality, with a higher percentage of "moderate and conservative" and "self-transcendent" values being the dominant values. The rate of "middle-of-the-road conservative" and "self-transcendent" values, which emphasize the interests of others and society, maintain social order and harmony, and moderate self-individuality, is higher. Liao and Chen [11] summarise the fundamental youth values: dynamic, self-aware, pioneering, romantic, realistic, marginal, and central.

In order to understand the youth group's perception of self-media, analyze the youth group's views on the future role of the self-media industry in guiding public opinion development, and promote self-media to lead youth to establish correct values. This paper uses ANOVA and decision tree models to explore the views of youth groups on the future orientation of public opinion in new media based on existing survey data. Compared with the existing literature, the innovation of this paper is mainly reflected in two aspects: 1) Exploring hotspot innovation. In the era of the digital economy, more and more people are putting their eyes on the new media industry due to the disappearance of spatial restrictions and the increasing demand. Although the new media industry has brought certain economic benefits, it has also generated a lot of problems in the process of its development. This paper makes up for the lack of relevant research to a certain extent and is innovative and practical; 2) innovation in research content. The survey questionnaire results were quantitatively coded and combined with machine learning to analyze the habits and perceptions of contemporary youth towards the new media industry and to study the attitudes of youth groups towards the future development of public opinion guidance in the self-media sector in the digital era.

## **2. Research on Young People's Attitudes Towards the Future Development of Public Opinion Guidance in the Self-media Industry**

### **2.1. Descriptive Statistics of Current Perceptions**

The data for the analysis of this paper comes from the questionnaire on the influence of self-media on youth values, and the main target of the study is the youth group. Among the respondents, 64.2% were women, and 35.8% were men.

The proportion of post-00 surveyed users was 20.57%, post-90 users accounted for 54.29%, post-80 age group users accounted for 11.57% of the total surveyed respondents, post-70 accounted for 11.86%, and the audience of other age groups only accounted for 1.71% of the total number of respondents. It can be seen from the data that the proportion of the post-90 age group was the highest, followed by post-80 users, post-70 and post-80 surveyed. The number of users is basically the same, and other age groups have the least number of monitored users. The proportion of specialized education is 10.18%, the balance of surveyed users with bachelor's degrees is 64.1%, the ratio of master's degree and above is 17.73%, and the proportion of other component education part is 7.99%, which can be seen that the proportion of bachelor's degree and master's degree is more than four-

fifths, and the balance of higher education among surveyed users is significant.

In addition to the basic information of the respondents, there are also some respondent tendency characteristics after cleaning the data. Firstly, there is a tendency to find frequency and length of viewing in the available data. Combining frequency of use and time peruse interactively, we can see that: the number of people who watch self-published content regularly and occasionally are close to each other, at 261 and 241, respectively. One hundred eighty-six of the respondents have been in the habit of watching self-published content, and of those who care occasionally, the number of those who use it for four hours or more (142) is the largest of all those who use it for four hours or more, at 68.93% ( $142 \div (142 + 25 + 39)$ ).

Among those who are always in the habit of watching self-published content, those with a viewing time of 2-4 hours made up the most significant proportion of all respondents with a viewing time of 2-4 hours, at 46.04%. Those who constantly watch self-published content and care for 2-4 hours make up the most significant proportion of all respondents who always watch, at 52.15%. It can be seen that those who care occasionally have a high probability of watching more extended periods of time when watching self-published content. The surveyed audience who watch more frequently are more likely to have 2-4 hours of viewing time.

Secondly, respondents believe that most of the information is accurate and can expand their knowledge. Among the surveyed users, the highest number of respondents, 429 and 449 respectively, chose that the information on the Internet can develop learning and that it is primarily factual, accounting for 62.35% and 65.26% of the respondents in their respective fields. Of all the respondents, only one thought that all information on the Internet was untrue, with 685 people favoring the truthfulness of data on the Internet.

### **2.2. Analysis of Variance (ANOVA) of Academic Level and Other Factors**

Analysis of variance (ANOVA) is an essential tool in the analysis of experimental data, which focuses on the correlation between some factors (independent variables) on an indicator (dependent variable) and whether the effect of the studied factors on the hand is significant [12]. The idea behind the construction of a statistic is to decompose the sum of squares and degrees of freedom of the total variation of the observations as a whole into two or more components according to the different sources of variation and to obtain the mean square and error mean square of the other sources of variation; by comparing the mean square and error mean square of the different sources of variation, one can determine whether the effect of a factor is significant at each level. The questionnaire included two options for respondents' perceptions of data and information on self-media platforms, namely, whether they could increase their knowledge base through self-media and the credibility of self-media information. A crosstabulation table created through data cleaning revealed that both options were associated with educational attainment, and the results of the ANOVA are shown in Table 1.

**Table 1.** Cardinality test

	Statistic		Degrees of freedom		Bilateral asymptotic significance	
	Knowledge base	Authenticity of information	Knowledge base	Authenticity of information	Knowledge base	Authenticity of information
Pearson	22.151	36.303	9	9	0.001	0
Likelihood ratio	21.741	38.131	9	9	0.001	0
Linear correlation	1.844	9.463	1	1	0.175	0.002

The p-value of 0.001 for the Pearson chi-square test of independence was less than the significance level of 0.05, so the original hypothesis that education is not correlated with the ability of self-media to increase knowledge reserves was rejected, and a significant correlation was found between the two, meaning that there was a substantial difference in the perception of education on whether self-media could increase knowledge reserves. The higher the education level, the smaller the proportion of those who think that self-media can increase knowledge reserve. The percentage of those who believe that self-media can increase the knowledge base is 74.5% among others, 74.3% among specialists, 61.7% among bachelors, and 52.5% among masters and above.

For the authenticity of the information, the p-value of Pearson's chi-square test of independence is less than the significance level of 0.05, so the original hypothesis that education is not relevant to the view of whether the information in self-media is accurate is rejected, and a significant correlation exists between the two, that is, there is a substantial difference between the views of different education levels on whether the information in self-media is accurate. According to the statistics, 0.7% of those with a specialist degree think that all information in self-media is untrue, while others do not; as the education level increases, the proportion of those who believe that most or all information in self-media is accurate becomes higher. Among the users surveyed, there is a clear distinction between the perception of the authenticity of online information and the perception of increasing knowledge. The most significant number of respondents, 429 and 449 respectively, chose that knowledge could be expanded, and that online information mainly was accurate, with 62.35% and 65.26% in each area. Of all the respondents, only one thought that all information on the Internet was untrue, with 685 people preferring it to be true. It can also be seen that there are significant differences in the perception of whether information on the Internet is accurate or not by education.

### 2.3. Decision Trees

A decision tree is a decision support tool that uses a tree model of decisions and their possible consequences, including the outcome of chance events, resource costs, and utilities. It is a way of displaying algorithms that contain only conditional control statements. Each internal node represents a 'test' of an attribute (e.g., whether a coin flips heads or tails), each branch represents the result of the test, and each leaf node represents a class label (the decision made after computing all attributes). The path from the root to the leaves represents the classification rules. In decision analysis, decision trees, and closely related influence diagrams are used as visual and analytical decision support tools. The expected value (or expected utility) of competing scenarios is

calculated. Decision trees consist of three nodes: decision nodes, chance nodes, and end nodes [13].

Decision trees can be used as classification or regression models. By constructing a tree structure, the data set is decomposed into smaller subsets for the ultimate purpose of prediction [14]. In a decision tree, the decision nodes divide the data, and the leaf nodes make predictions. Traversing from the root node to the leaf nodes enables the conversion to logical rules in IF-ELSE-THEN format. The root node (the first decision node) is the most effective node for dividing the dataset [15]. There are usually two methods for evaluating the ability of a node to split a dataset, information gain and the Gini coefficient. One is information gain, where the feature with the most significant information gain is chosen as the root node to divide the dataset. The information gain is the difference between the parent node's entropy and the child nodes' average entropy, by which the importance of a given feature value can be judged. Entropy is a standard measure of the impurity of the target class.

$$Entropy = -\sum_i p_i \log_2 p_i \quad (1)$$

where  $p_i$  is the proportion of samples in the category,  $I$  in the data set.

The other is the Gini coefficient, which is calculated as follows.

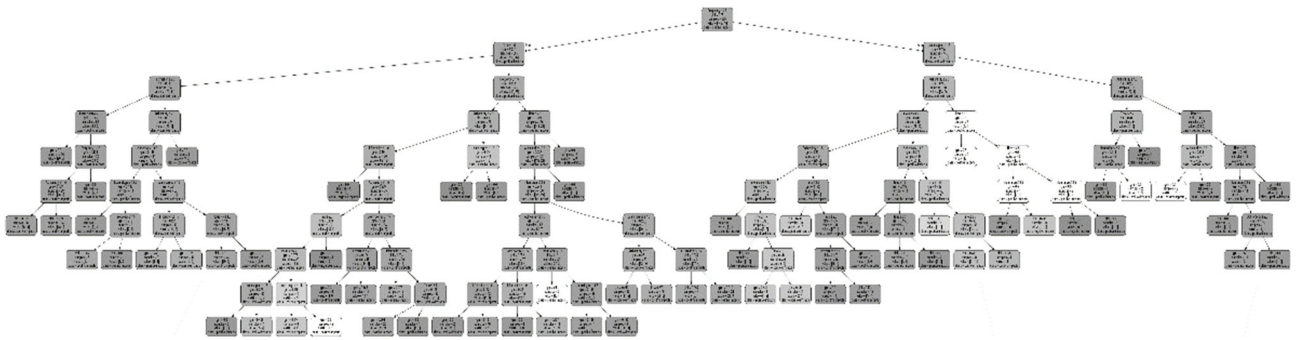
$$Gini = 1 - \sum_i p_i^2 \quad (2)$$

The Gini coefficient is faster to calculate because it does not require the calculation of a logarithmic function. In practice, the information gain or Gini coefficient will not have much influence on the results—the decision tree forecasting model. Based on the previous descriptive statistics, reference is made to the principles of systematicity, typicality, dynamism, and scientificity. The selection of indicators and update coding are shown in Table 2.

In this paper, using the Sklearn tool, respondents' level of education (Learn degree), daily browsing time (Time), daily browsing frequency (Frequency), respondents' income level (Income), increased knowledge base (Information), and information authenticity (Authenticity) were selected as independent variables. Attitude towards future-oriented development was chosen as the predictor variable. The training and testing sets were divided, and the SOM model was used to train the existing sample data. 70% of the data was used as the training set and 30% as the test set. The decision tree model was also built. In order to prevent the decision tree model from overfitting in the training set, additional model parameters such as the maximum height of the tree, the maximum

**Table 2.** Indicators and updated coding

Title	Description	Range	Code (integer)
RespondentId	Respondent No.	688 valid data	1—688
Learn degree	Respondent's education level	Secondary education and below, Higher education	1,2
Time	Daily browsing hours	Shorter, fair, longer	1,2,3
Frequency	Daily browsing frequency	Less frequent, average, frequent	1,2, 3
Income	Respondent's income level	Secondary and below, higher	1,2
Information	Increase in knowledge base	Not, not always, yes	1,2,3
Authenticity	Authenticity of information	Not much, mostly, all	1,2,3
Attitude	Future-oriented development attitude	Negative, positive	0,1

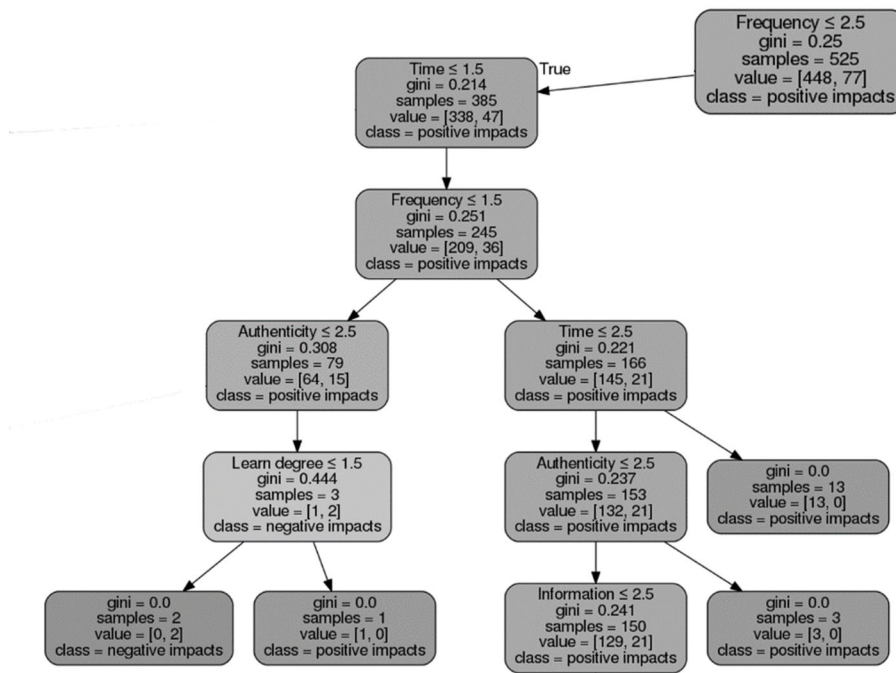


**Figure 1.** Decision tree classification results

number of leaf nodes, minimum division impurity, and a minimum number of samples of leaf nodes were also set. In the evaluation of the effectiveness of the future-oriented development attitude prediction model, the prediction accuracy of the type in the test set was 80.57%, which was much higher than that of random guesses (50%). Finally, the decision tree model obtained from the training was visualized with the help of Graphviz software, and the figure shows the visualization results of the decision tree.

Due to the large size of the decision tree, a portion of the

results of the central nodes of the decision tree have been intercepted for illustrative purposes in this paper to see the results. In the model building, the feature with the most significant Gini coefficient at the root node corresponds to the length of time spent browsing (Frequency), a node that illustrates that the factor that most influences respondents' expectations of the future development of self-publishing are the frequency of browsing on new media platforms by the respondents themselves on a daily basis.



**Figure 2.** Scaled-up results for some cases of decision trees

By zooming in on the above diagram and looking at the

paths from the root node to the two leftmost sets of leaf nodes

in the chart, two typical respondent characteristics can be seen, and the consistency includes that the daily browsing hours and frequency are both low and average (less than 2 hours). It is also considered that the authenticity of information on new media platforms requires a certain degree of judgment and is context-specific. The final decision of the future-oriented development attitude of the two types of such characteristics towards the self-media platforms depends mainly on education. Respondent types with an educational level of secondary education and below have a negative attitude, while those with an academic level of higher education and above have a positive attitude.

### 3. Conclusions and Recommendations

#### 3.1. Specific Conclusions

From the overall sample, almost all of them habit using self-media. After analyzing the specific questions of the questionnaire, we have mined the current situation of the survey audience's awareness of self-media and their views on the future development of public opinion orientation through descriptive statistical analysis, analysis of variance, and decision tree model. The following conclusions were drawn.

Firstly, the respondents of this paper are primarily young people, and there is a specific correlation between their daily browsing frequency and the length of their browsing time. It can be found that the number of respondents who often and occasionally watch self-published media is close to each other, and the frequency of browsing is inversely proportional to the length of browsing for most respondents. Respondents who view frequently are more likely to use their time in fragments, while those who view occasionally are more likely to view for more extended periods.

Secondly, education significantly impacts the learning behavior of knowledge accumulation and the judgment of the perceived authenticity of information through self-marketing platforms. There was an inverse relationship between education level and knowledge accumulation through self-media platforms. The percentage of those with secondary education level who believe that self-media can increase their knowledge base is 74.3%, and this percentage decreases as the level of education rise. There is a significant difference between education levels in terms of whether the information on self-media is perceived to be true or not. As education levels increase, the proportion of those, who believe that most or all of the information in self-media is actual increases.

Finally, respondents' perceptions of the future development of public opinion on self-media depend mainly on their daily browsing frequency. People who browse frequently make different judgments about the credibility of information on self-media platforms due to their medium education level, which is inversely proportional to their credibility. Among this group, those with a medium education level and a confident attitude towards the authenticity of information all think that self-media will play a positive role in developing public opinion guidance in the future.

#### 3.2. Suggestions

First, at the government level. Government regulators should innovate management ideas, explore new ways to rectify cyberspace, use modern internet and internet technology to establish a grading mechanism for self-media users, set entry thresholds for the self-media industry according to the levels of different self-media users, and

implement territorial monitoring and complete process management for self-media users; for the time being, the most effective approach is to fully promote the real-name authentication system implemented by large self-media platforms. In this way, when self-media users publish information, they can carry out real-time identity conscientiousness and, to a certain extent, regulate the online behavior of the masses and reduce the pollution of inaccurate and harmful information.

Secondly, the moral level forms a "soft binding force." For the source, we require the self-media to publish positive information, encourage the establishment of self-media industry associations, improve the self-media industry's ethical standards, and improve the professionalism of self-media people; for the audience, self-regulation and other regulation should be combined. On the one hand, the audience should be restrained from a legal perspective. On the other hand, the audience should be strengthened in their ability to judge the value of information, choose and discern the authenticity of the information, establish correct values and choose information on self-media platforms rationally.

Finally, youth groups need to take the best of the best and remove the worst of the worst. Self-media is a "double-edged sword" for the development of young people. When using self-media, young people should learn to take advantage of their strengths and avoid weaknesses. Young people should learn to broaden their minds and horizons, exercise their ability to think independently and analyze and judge things; likewise, young people should learn to abandon the temptation of undesirable factors on self-media and establish a correct view of the Internet so that the positive guidance role of self-media can be entirely played on young people and meet the opportunities and challenges brought by the development of self-media industry in the digital economy.

### References

- [1] Iranga Suroshana. *Social Media Culture*. Colombo: S. Godage and Brothers. 2016. ISBN 978-9553067432.
- [2] Liu, Hang, Fulin, Li, Tao, Sun, Baowen. Research on the Internet and digital economy based on Chinese practice - An overview of the first Internet and digital economy forum [J]. *Economic Research*,2019,54(03):204-208.
- [3] Lu L. Promoting SME Finance in the Context of the Fintech Revolution: a Case Study of the UK's Practice and Regulation.[J]. *Banking and Finance Law Review*. 2018.33(3).317~343.
- [4] Jiang Xiao Juan, Luo Li Bin. Globalization of services in the network era - new engines, acceleration and great power competitiveness[J]. *China Social Science*,2019(02):68-91+205-206.
- [5] Yin Z. Gong X. Guo P. Wu T. What Drives Entrepreneurship in Digital Economy? Evidence from China [J]. *Economic Modelling*, .2019.82(3).66~73
- [6] He Zongyue, Song Xuguang. Mechanisms and insights of the digital economy in promoting employment - Reflections after the epidemic [J]. *The Economist*,2020(05):58-68.
- [7] Zhang Xun, Wan Guanghua, Zhang Jiajia, He Zongyue. Digital economy, financial inclusion and inclusive growth[J]. *Economic Research*,2019,54(08):71-86.
- [8] Xia Deyuan. Media interaction in the digital age and the symbolic significance of traditional media [J]. *Academic Monthly*,2011,43(03):25-31.

- [9] Deng Xinmin. Self-media: the latest stage of new media development and its characteristics[J]. Exploration, 2006(02):134-138.
- [10] Liu Wenrong. Characteristics and influencing factors of basic values of urban youth[J]. China Youth Studies,2019(03):52-59.
- [11] Liao Xiaoping, Chen Jian Yue. The basic characteristics of youth values [J]. Journal of China Youth Politics College, 2006 (04).
- [12] Wang Songgui, Chen Min. Chen Liping. Linear statistical models: linear regression and analysis of variance [M]. Beijing: Higher Education Press, 1999.
- [13] Kamiński B. Jakubczyk M. Szufel P. A framework for sensitivity analysis of decision trees. central European Journal of Operations Research. 2017. 26 (1): 135-159. doi:10.1007/s10100-017-0479-6.
- [14] Zhou C. H. Machine learning [M]. Beijing: Tsinghua University Press, 2016.
- [15] Cieslak D A. Chawla N V. Learning decision trees for unbalanced data. Machine Learning and Knowledge Discovery in Databases. 2008.