Hindawi
Wireless Communications and Mobile Computing
Volume 2020, Article ID 8861207, 14 pages
https://doi.org/10.1155/2020/8861207

Research Article

Research on the Evaluation Model of Rural Information Demand Based on Big Data

Yanfeng Jin,[1,2] **Gang Li**,[1] **and Jianmin Wu**[2]

[1]*School of Economics and Management, Beijing University of Posts and Telecommunications, Beijing 100876, China*
[2]*Shijiazhuang Posts and Telecommunications Technical College, Shijiazhuang 050021, China*

Correspondence should be addressed to Gang Li; jinyf@bupt.edu.cn

Received 17 June 2020; Revised 7 July 2020; Accepted 20 August 2020; Published 8 September 2020

Academic Editor: Bingxian Lu

In recent years, the imbalance of rural information supply and demand has seriously hindered the process of rural informatization. Rural information demand is a decisive factor in the relationship between rural information supply and demand. Therefore, research on the influencing factors of rural information demand has attracted much attention. The traditional rural information demand factor analysis does not consider the correlation between factors. The factors themselves carry a lot of repeated information, which seriously interferes with the objectivity of the analysis results. Proceeding from the complexity and diversity of influencing factors of rural information demand, based on the selected subjective and objective factors, based on the forward partial correlation analysis and post-ROC test, a probit discriminant model of influencing factors of rural information demand was constructed, and the relationship with Lingshou was determined. There are 24 factors that are significantly related to county rural information needs. The research results show that this method not only eliminates the factors that carry highly repetitive information and the correlation is not significant but also makes the results more reliable. At the same time, it also found that rural information supply is related to farmers' information cognition ability, acceptance awareness, and acceptance ability. This study provides new methods and new ideas for solving related problems.

1. Introduction

With the rapid development of science and technology and the advent of the big data era, the construction of smart cities at home and abroad has made remarkable achievements. At the same time, rural informatization also ushered in new opportunities for development. Digital rural areas and intelligent rural areas have become the hot spots of scholars. Of course, there are great challenges as well as opportunities. Due to the unbalanced development of regional economy, there are many problems in rural informatization. First, the collection, processing, integration, and sharing of rural information are difficult. Second, data mining cannot be carried out effectively and does not provide the information farmers need. Third, there is a contradiction between the diversity of farmers' information needs and the unity of platform information. Fourth, lack of dynamic maintenance and update mechanism, data outdated, cannot play its due role. This requires big data technology and big data thinking to improve and solve the current difficulties. The application of big data in various industries has achieved good results. The idea of big data has gradually penetrated into the process of rural informatization. With the help of big data technology, we can build a comprehensive rural information platform based on farmers' information needs.

Whether in developed countries such as Europe and the United States, or in developing countries such as Asia and Africa, there are many studies on the information needs of rural residents. Scholars' research shows that farmers' demand for information is more and more extensive, and the types of demand and access channels show a variety of characteristics [1]. Kaniki's survey of two rural communities in South Africa found that the main information needs of farmers are information needed to seek jobs or increase income, vocational or skills training opportunities, information about grants, medical and health information, legal counseling services, and so on [2]. In Asia, Raju's study found that the most common information needs of Indian farmers

were medical and health information, infrastructure information, crop improvement and yield information, product sales and market information, policy, and service information [3]. Vevrek thinks the daily information needs of the rural population in the United States are information about local government decisions, information about health services, and local news [4–6]. Domestic researchers have found that farmers pay more attention to specialized information related to agricultural production and operation. Zhang Ying (2017) based on the rural information service platform, from the perspective of farmers, found that farmers' demand for labor market information, agricultural market information, agricultural policy information, and agricultural production information decreased in turn. Li Lu (2016) surveyed the demand for agricultural technology social services and found that farmers' age, education level, and whether they went out to work would affect farmers' demand for agricultural technology social services. Young and experienced peasants paid more attention to information services related to the circulation of agricultural products. Zhou Fengtao (2016) studied the farmers' demand for information services and found that educational level and whether to participate in rural cooperatives had a significant impact on their demand for technical services, agricultural services, and information services. Lu Xinru and Li Zhigang (2017) explored the unique information needs and behaviors of farmers through questionnaires. Farmers' information demand had three characteristics: the tendency of market purchase and sale information, the necessity of policies and regulations, and the particularity of meteorological forecast. Information behavior was restricted by educational level and the overall channel was narrow. Ma Chunyan (2016) carried out an investigation and research on poverty-stricken areas. From the questionnaire, through the analysis of demand types, information access channels, personal literacy, and other aspects, it provided suggestions and countermeasures for speeding up the development of local agriculture and reversing the backward development situation in remote areas. Pan Yuchen and Huo Yucan (2018) analyzed the concept of rural information consumption, the level of demand, and the motivation of consumption, especially in the field of emotional demand, which was also a further reflection of the demand level theory. Provided guidance for the development of the whole society and related enterprises helped enterprises to improve the pertinence of information services and achieve steady growth. Wang Xiaoning and Wang Ming (2018) empirically analyzed the main channels for farmers to obtain information under the background of mobile Internet by issuing questionnaires. Through the analysis, it was concluded that mobile micromessaging, mobile QQ, and mobile microblogging are the absolute dominant advantages in information dissemination, while agricultural information website platform was not generally known to farmers. Guan Lili (2017) analyzed the information needs and constraints of farmers through questionnaires, especially the five characteristics of local farmers: the increasing variety of demand categories, the diversification of access methods, the depth of demand levels, the strong internal motivation of demand, and the strong ability of information research and judgment. It provided experience in understanding the level of rural informatization and promoting the construction of information frameworks. Cui Kai and Feng Xian (2017) combed and analyzed the relevant literature at home and abroad, and studied the significance of information dissemination, the information needs of rural residents, and the information supply in rural areas. From the perspective of information poverty alleviation, Li Gang and Qiao Haicheng (2017) proposed that the government should pay attention to information poverty alleviation through the construction of rural poverty-stricken area model and analysis of relevant data.

In summary, it has been found from the existing research that the information needs of farmers in China are increasingly strong and the demand structure is increasingly diversified, but the specialized information related to agricultural production development is still the most important component of farmers' information consumption. Affected by income levels and cultural quality, mass media such as television and broadcasting are still the main channels of information dissemination, but the proportion of computers and mobile phones is increasing, especially in economically developed areas [10–13]. Researchers summarized and analyzed the influencing factors of farmers' information demand from various angles, but the correlation analysis between the influencing factors is relatively small, and the statistics and descriptions of the factors are not comprehensive enough [14–16]. At the same time, the significant impact of various factors on rural information demand is insufficient. Aiming at the above problems, the concept model of farmer information demand of "source-flow-use" was put forward. Based on the discrete selection model of econometrics, the probit model of rural information demand was constructed. Firstly, the partial correlation analysis of the influencing factors of rural information demand was carried out, and the high coincidence factor was removed. The probit model was used for the second test. Finally, the ROC curve was used for discrimination. Eight factors with no significant influence, such as the proportion of fixed-line administrative villages, were removed. At the same time, 24 significant influencing factors were ranked according to the degree of influence. The results prove the feasibility of the method.

2. Model Building

2.1. Evaluation of Influencing Factors Based on Partial Correlation Analysis. In a system consisting of multiple elements, when studying the influence or correlation of one element on another, the influence of other elements is regarded as a constant, i.e., the close relationship between the two elements is studied separately without considering the influence of other elements, which is called partial correlation analysis [17, 18]. That is the partial correlation coefficient. In the study of rural information demand, there are many factors involved. There may be some correlations between the factors, which leads to the duplication of information reflected by two or more factors, which leads to the system being too complicated because there are unrelated factors [19]. Through partial correlation analysis, factors with repeated

information that affect rural information needs can be removed.

(1) Calculation of partial correlation coefficient.

Suppose t_{ij} is the data value of the i index of the selected village j in the region, t_{kj} is the data value of the k index of the selected village j in the region, and r_{ik} is the partial correlation coefficient between the k index and the first index. The formula is as follows:

$$r_{ik} = \frac{\sum_1^n \left(x_{ij} - \overline{x_i}\right)\left(x_{kj} - \overline{x_k}\right)}{\sqrt{\sum_1^n \left(x_{ij} - \overline{x_i}\right)^2}\sqrt{\sum_1^n \left(x_{kj} - \overline{x_k}\right)^2}}. \quad (1)$$

Among them, n denotes the number of villages in the study area, $\overline{x_i}$ denotes the average value of the i factor, and $\overline{x_k}$ denotes the average value of the k factor.

Suppose R is a correlation coefficient matrix composed of partial correlation coefficient r_{ik}, where m is the number of influencing factors, then.

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mm} \end{bmatrix}. \quad (2)$$

Let S be the inverse matrix of the correlation coefficient matrix R.

$$S = R^{-1} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1m} \\ s_{21} & s_{22} & \cdots & s_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ s_{m1} & s_{m2} & \cdots & s_{mm} \end{bmatrix}. \quad (3)$$

According to the formula of partial correlation coefficient, the partial correlation coefficient r'_{ik} between the i factor and the k factor can be obtained.

$$r'_{ik} = \frac{-s_{ik}}{\sqrt{s_{ii}s_{kk}}}. \quad (4)$$

The greater the partial correlation coefficient r'_{ik} is, the greater the correlation between the and the k influencing factors is. And the smaller the r'_{ik} is, the smaller the correlation between the i and the k influencing factors is.

(2) Calculation of F value

When the correlation between the two factors is high, in order to avoid the subjective deletion of the significant factors, we can solve this problem by calculating the F value of the two factors. Assuming that F_i is the F value of the i factor, Equation (5) can be used for calculation.

$$F_i = \frac{\left(\bar{x}_j^{(0)} - \bar{x}_j\right)^2 + \left(\bar{x}_j^{(1)} - \bar{x}_j\right)^2}{\left(1/\left(n^{(0)} - 1\right)\right)\sum_{y_j=0}\left(x_{ij} - \bar{x}_j^{(0)}\right)^2 + \left(1/\left(n^{(1)} - 1\right)\right)\sum_{y_j=1}\left(x_{ij} - x_j^{(1)}\right)^2}, \quad (5)$$

$$\overline{x_i}^{(0)} = \frac{1}{n^{(0)}}\sum_{y_j=0} x_{ij}, \quad (6)$$

$$\overline{x_i}^{(1)} = \frac{1}{n^{(1)}}\sum_{y_j=1} x_{ij}, \quad (7)$$

$$\overline{x_i} = \frac{1}{n}\sum_{j=1}^{n} x_{ij}. \quad (8)$$

F_i reflects the magnitude of the influence of the i factor on rural information demand; the greater the F_i is, the greater the impact is; on the contrary, the smaller the impact on rural information demand is.

In the multivariate analysis of rural information demand factors, pure correlation analysis cannot fully reflect the correlation between the factors, because other factors interfere with these factors, so partial correlation analysis is an effective way to solve this problem [20].

(3) Set the deletion criterion based on partial correlation analysis

If the absolute value of the partial correlation coefficient of two related factors $|r_{ik}| > 0.7$, it is considered that the two factors are highly correlated, and the information of the two factors response is highly repeatable, so one of them should be deleted. If the partial correlation coefficient is greater than 0.7, the factor whose F value is less than 0.7 should be deleted.

2.2. Analysis of Influencing Factors Based on Probit Regression

2.2.1. Discrete Probit Regression Model. The probit model is a generalized linear model that follows a normal distribution [20]. The simplest probit model is that the explanatory variable Y is a 0, 1 variable, and the probability of an event occurring depends on the explanatory variable $(Y = 1) = f(x)$, that is, the probability of $Y = 1$ is a function of X, where $f\ ()$ obeys the standard normal distribution. This paper will use the probit model to screen out the factors affecting the information demand in rural areas. When the value of dependent variable y_j is 1, it shows that independent variable has an impact on rural information demand, and when the value of dependent variable y_j is 0, it shows that independent variable has no effect on rural information demand.

(1) Introducing intermediate variables y_j^*

Because the probit model is a linear model, and the dependent variable y_j is 0 and 1, it is a discrete variable, so it cannot be directly calculated by linear regression equation. Therefore, it can be solved by introducing intermediate

variable y_j^* and fitting linear regression equation with influencing factors. y_j^* can represent a state of rural information demand; when $y_j^* > 0$ and the value of y_j is 1, think that this factor has an impact on rural information demand; when $y_j^* < 0$, think that the value of y_j is 0, and this factor has no impact on rural information demand. The linear regression equation is given below.

$$y_j^* = \sum_{i=1}^{m} \beta_i x_{ij} + \alpha + \varepsilon_j = X_j\beta + \alpha + \varepsilon_j. \tag{9}$$

y_j^* is an intermediate variable, representing the rural information demand state of the j village; β_i represents the regression coefficient of the i influencing factor; x_{ij} represents the observed value of the i influencing factor of the j village; α is a constant term; ε_j is a random variable and obeys normal distribution $\varepsilon_j \sim N(0, \sigma^2)$; $\beta = (\beta_0, \beta_1, \cdots, \beta_m)$ is a regression coefficient vector, and $X_j = (x_{1j}, x_{2j}, \cdots, x_{mj})$ is a vector composed of the influencing factors of the j village.

(2) Calculate the probability of rural information demand in each village

The intermediate variable y_j^* of Equation (10) is used to calculate the probability of rural information demand in each village. Because of $\varepsilon_j \sim N(0, \sigma^2)$, it is concluded that

$$P\left(y_j = 1 \mid X_j\right) = P\left(y_j^* > 0 \mid X_j\right) = \Phi\left(\alpha + X_j\beta\right). \tag{10}$$

Similarly, it is possible to calculate the probability of unaffected information demand in rural areas:

$$P\left(y_j = 0 \mid X_j\right) = P\left(y_j^* < 0 \mid X_j\right) = 1 - \Phi\left(\alpha + X_j\beta\right). \tag{11}$$

Where f is a normal distribution function, it can be solved by Equation (12) through maximum likelihood estimation.

$$\text{MAX} \ln L = \sum_{j=1}^{n} \left[y_j \ln\left(\Phi\left(\alpha + X_j\beta\right)\right) + \left(1 - y_j\right) \ln\left(1 - \Phi\left(\alpha + X_j\beta\right)\right) \right]. \tag{12}$$

2.2.2. Testing Based on the Probit Model. Construct a probit model, establish the Wald statistic of the influencing factors, and use the chi-square test [21, 22]. When the corresponding significance probability is greater than 0.01, the factors with the greatest significance probability are deleted. The specific steps are as follows:

(1) Calculate the regression coefficient of the probit model. The probit regression model was constructed according to Equations (9) and (12) of m factors affecting rural information demand and the corresponding observed values of rural information demand state y_i. The corresponding coefficients α, β and corresponding standard errors SE_{β_k} are solved, where $\beta = (\beta_1, \beta_2, \cdots, \beta_m)$

(2) Calculate the significance probability of each factor s, construct the Wald statistics of each factor, and test the hypothesis of the significance of each factor

Suppose H_0: $\beta_k = 0$. If H_0, the k factor has no significant impact on rural information demand.

Suppose H_1: $\beta_k \neq 0$. If H_1, then the k factor has a significant impact on the rural information demand.

Let W_k be the Wald statistical variable corresponding to the k influencing factor of rural information demand, β_k be the parameter estimation value of the k influencing factor, and SE_{β_k} be the standard error of β_k, then.

$$W_k = \left(\frac{\beta_k}{\text{SE}_{\beta_k}}\right)^2. \tag{13}$$

By constructing the Wald statistic W_k, it is possible to test whether the parameter estimation β_k of the influence factors is significantly 0. If $\beta_k = 0$, H_0 is true. W_k obeys the chi-square distribution with degree of freedom 1, that is $W_k \sim \chi^2(1)$; the corresponding significance probability value s is obtained according to the chi-square distribution table.

(i) If $s < 0.01$, the original hypothesis H_0 is rejected, which shows that this factor has a significant impact on the rural information demand

(ii) If $s > 0.01$, then accept the original hypothesis H_0, indicating that although $\beta_k = 0$, but this factor has no significant impact on rural information needs

(3) For all the influencing factors of significant probability $s > 0.01$, the maximum s value is removed. $s > 0.01$ shows that accepting the hypothesis H_0, this factor has no significant impact on rural information demand. Among all the factors that have no significant impact, the factors corresponding to the maximum s value can be removed. It should be noted that all factors affecting $s > 0.01$ cannot be deleted at one time, because each factor may be affected by multiple variables, deleting a variable; the original non-significant factors may become significant factors

(4) Repeat Steps (1)–(3) until the coefficients of all variables in the model meet $s < 0.01$

By solving the state variable y of rural information demand and the coefficient of probit regression equation β between influencing factors and its standard error SE_β, construct Wald statistics of influencing factors to test the significance probability of regression equation coefficient β and eliminate the factors that have little impact on rural information demand, and the regression coefficient β is not significant.

Table 1: Classification confusion matrix.

Real situation	Prediction results	
	Positive example	Counter example
Positive example	True example (TP)	False counter example (FN)
Counter example	False positive cases (FP)	True counter example (TN)

Table 2: Classification results of the probit regression model for influencing factors of rural information demand.

Actual impact	Model classification results		
	1 (influential)	0 (no impact)	Total
1 (influential)	The actual influence is determined by the number of factors that are affected by the model TP	The number of factors that actually affect but is misjudged by the model is not affected by FN	TP + FN
0 (no impact)	The number of factors that are misjudged by the model is FP	The actual number of factors that were correctly judged by the model was not affected by TN	FP + TN
Total	TP + FP	FN + TN	

2.3. Validation of Influencing Factors Based on ROC Curve

2.3.1. ROC Curve. The ROC curve refers to the receiver operating characteristic. Each point on the ROC curve reflects the sensitivity to the same signal stimulus [23, 24]. In view of the relationship between the predicted value and the true value, we can divide the sample into four parts: true positive (TP): the predicted value and the true value are all 1; false positive (FP): the predicted value is 1, and the true value is 0; true negative (TN): the predicted value and the true value are both 0; and false negative (FN): the predicted value is 0, and the true value is 1. The classification confusion matrix is shown in Table 1.

The vertical axis of the ROC curve represents true positive rate (TPR), and the horizontal axis represents false positive rate (FPR).

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{14}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{TN} + \mathrm{FP}}. \tag{15}$$

ROC curve is actually a dot plot of TPR and FPR under different thresholds. Given a threshold, we can get the corresponding TPR and FPR values. By detecting a large number of thresholds, a TPR-FPR correlation map can be obtained. In AUC (area under the curve), that is, the larger the area under the ROC curve is, the better the classifier is, the maximum value is 1.

2.3.2. Inspection of Influencing Factors of Rural Information Demand Based on ROC Curve. The ACU value of ROC curve is used to determine whether the factors affecting rural information demand selected by the probit regression model are correct [25]. According to the confusion classification matrix, the number of influential factors is recorded as TP, the number of factors misjudged as influential factors is recorded as FN, the number of factors judged as unaffected factors is recorded as FP, and the number of factors misjudged as unaffected factors is recorded as TN. The specific analysis results are shown in Table 2.

According to Equation (14), the correct discriminant rate is calculated, and the number TP which is discriminated as the influential factor is divided by the number TP + FN which is the actual number of all the influential factors. It indicates that the factors that actually affect the rural information demand are discriminated as the probability of influencing factors by the abovementioned probit model [26].

According to Equation (15), the misjudgment rate is calculated, and the number of factors which are misjudged as influential factors is divided by the number of factors that are not actually affected by the number of FP + TN. It is indicated that the factors that have no influence on rural information demand are identified as influential factors by the abovementioned probit model.

The ROC curve is plotted on the longitudinal axis and the horizontal axis, respectively, by the correct discriminant rate and false discrimination rate [27]. When the abscissa is constant, the larger the ordinate is, the greater the impact of this factor on rural information demand is, and the corresponding AUC value is also larger. Therefore, the larger the AUC value is, the better the classifier is, which means that the greater the impact of this factor on rural information needs is, the maximum value is 1. When AUC = 1, it is a ideal classifier, and with this prediction model, ideal prediction can be achieved no matter what threshold is set. When $0.9 < \mathrm{AUC} < 1$, the influence factor is better. If the threshold is set properly, the model has better predictions. When $0.7 < \mathrm{AUC} < 0.9$, the influence factors are moderate, and the model has a certain predictive value. When $0.5 < \mathrm{AUC} < 0.7$, the discriminant effect is poor, and there is basically no predictive value. Where $\mathrm{AUC} < 0.5$, the discriminant effect of the model is very poor, but it is better than random guess as long as it always goes against prediction.

Therefore, according to all the factors identified by the above probit regression model, if the AUC value is greater than 0.9, it is concluded that this factor has a significant impact on rural information demand. The research shows that the area under the ROC curve constructed by all the factors in this paper is higher than 0.9, which ensures the ability

Table 3: Influencing factors of rural information demand.

Research object	First level influencing factors	Two level influence factors
Influencing factors of rural information demand	Environmental factors	Fixed coverage of administrative villages X1
		Number of cable TV per 1000 people X2
		Optical fiber length per 100 square kilometers X3
		TV coverage rate X4
		Number of information talents per 10000 people X5
		Number of students per 1000 students X6
		The number of computers per 100 households in rural areas X7
		The number of TV sets per 100 households in rural areas X8
		The number of mobile phones per 100 households in rural areas X9
		Number of Internet users per 10000 X10
		Rural per capita postal volume X11
		Fixed investment in telecom industry accounts for the proportion of total social investment X22
		Fixed investment in the information industry accounts for the proportion of fixed asset investment in the whole society X13
	Main factors	Sex X14
		Age X15
		Marital status X16
		Health X17
		Cultural level X18
		Occupation X19
		Personal income X20
		Experience of going out for work X21
	Family factors	Number of family members X22
		Number of family labor force X23
		Number of male family members X24
		Number of family members X25
		Source of family income X26
		Family happiness index X27
	Economic factors	Per capita income of farmers X28
		Source of farmers' income X29
		Per capita disposable income of farmers X30
	Geographical factors	Distance from county highway X31
		Distance from provincial highway X32
		Distance from town center X33
		Distance from county center X34
	Cognitive factors	Knowledge of information X35
		Awareness of information acceptance X36
		The ability to receive information X37
	Policy factors	National informatization policy X38

to distinguish the influence of various factors on rural information demand.

3. Empirical Analysis of Rural Information Demand

3.1. Analysis of Influencing Factors of Rural Information Demand. Through the combing and research of domestic and foreign literatures, the factors affecting rural information demand are summarized into seven aspects: environmental factors, subject factors, family factors, economic factors, geographical factors, cognitive factors, and political factors [28, 29].

3.1.1. Environmental Factors. At the micro level, the popularity of the Internet, the number of computers, and the number of mobile phones, television, and radio coverage have become important factors affecting rural information needs.

Table 4: Distribution of selected respondents.

Village name	Sample size	Village name	Sample size	Village name	Sample size
New village	78	Lijiazhuang	55	Xichatou	85
Xituo	66	Wanli	61	Lijiagou	54
Ximufu	48	Nanyanchuan	44	Majiazhuang	53
Beijicheng	65	Sijiazhuang	46	Liatong	44
Zhushi	39	Nanbaishi	54	Zhangjiatai	35
Xiaohanlou	67	Xiqingtong	39	Xiwan	29
Niucheng	69	Ciyu	70	Niuzhuang	34
Dongchengnan	43	Dongjiazhuang	46	Zhaitou	48
Sunzhuang	57	Beitanzhuang	80	Nanying	32
Nangoutai	34	Shanmenkou	42	Manshan	26

First, rural information infrastructure and technology are the basic resources of rural information environment and an important premise of rural information environment optimization. Its construction level is an important part of rural information environment. The second is rural information talents. The optimization of rural information environment needs high-quality and professional talent team to achieve, in order to continuously promote the improvement of rural informatization level. Rural scientists and technicians are an important force in the construction of rural information environment and an important guarantee for the continuous advancement of rural informatization. Rural college students have higher professional quality and professional ability, which is an important force in the future construction and optimization of rural information environment. The third is the rural information network coverage. It reflects the application of rural information infrastructure. The four is the input and output of rural informatization.

3.1.2. Subject Factors. Individual characteristics mainly include gender, age, marital status, health status, educational level, occupation, personal income, and migrant work experience. Gender is an important factor affecting rural information need. Generally speaking, men's demand for information is more intense than that of woman. From the perspective of information economics, the subjective desire of different age structures for rural information needs is quite different. Young people are more likely than the elderly to accept new information technology and information products. The impact of marital status on rural information needs research results that are rare, and it is unclear whether there is a correlation. This paper will explore this issue through follow-up models. Health status is also a major impact on rural information needs. The cultural level affects the information quality of rural subjects to a great extent. The traditional theory of rural informatization holds that the farmers' information quality has a positive correlation with the demand and acceptance of informatization. People are engaged in agricultural and nonagricultural occupations in rural areas; the dual nature of occupation may also have an impact on rural information needs. In general, the higher the personal income is, the stronger the demand for information is. Farmers with migrant experience have a wider horizon and a stronger sense of information needs.

3.1.3. Family Factors. Family factors mainly include the number of family population, the number of family labor force, the number of male family, the number of female family, family happiness index, and family income sources. The theory of network externalities believes that as the number of users increases, utility gained by each user from the network increases. Therefore, the number of family members may also be an important factor affecting the information needs in rural areas. Statistical studies have shown that gender is an important factor affecting Internet demand. For rural households, the more males there are, the stronger the rural information demand there is. Similarly, the number of women in the family may also affect the family's demand for rural information. The quantity of household labor force is proportional to household income to a certain extent. The more the labor is, the higher the household income is, the stronger the demand for information is. On the contrary, the less the labor is, the lower the household income is, the lower the desire for information demand is. Family happiness index in a sense reflects the level of family income and indirectly affects the farmers' demand for information. At present, the relationship between happiness index and information access demand has not been found in academic and theoretical circles. However, we can see that the higher the family happiness index is, the higher the income is, so it will indirectly affect the farmers' demand for information.

3.1.4. Economic Factors. Economic factors mainly include the per capita income of farmers, the source of farmers' income, and the level of regional economy.

3.1.5. Geographical Factors. The geographical characteristics of rural information demand have great influence. The geographical features are mainly reflected in the geographical location of rural areas, including county-level roads, provincial highways, distance from township centers, and distance from county centers.

Table 5: Part village related raw data.

First level influencing factors	Two level influence factors	Related raw data					
		C1	C2	C3	C28	C29	C30
Environmental factors	Fixed coverage of administrative villages X1	80	91	68	86	69	74
	Number of cable TV per 1000 people X2	306	203	156	489	543	345
	Optical fiber length per 100 square kilometers X3	5	23	45	21	32	37
	TV coverage rate X4	90	97	88	96	76	85
	Number of information talents per 10000 people X5	123	178	45	36	234	132
	Number of students per 1000 students X6	45	67	39	44	29	56
	The number of computers per 100 households in rural areas X7	39	43	28	50	45	77
	The number of TV sets per 100 households in rural areas X8	90	93	96	97	99	99
	The number of mobile phones per 100 households in rural areas X9	156	137	269	211	169	304
	Number of Internet users per 10000 X10	1267	3304	890	1278	908	4512
	Rural per capita postal volume X11	4	7	11	6	4	2
	Fixed investment in telecom industry accounts for the proportion of total social investment X22	10	13	21	8	25	14
	Fixed investment in the information industry accounts for the proportion of fixed asset investment in the whole society X13	12	13	16	21	15	17
Main factors	Sex X14	1	1	2	2	1	1
	Age X15	3	3	4	3	4	5
	Marital status X16	1	1	1	1	1	1
	Health X17	3	4	3	4	5	4
	Cultural level X18	2	1	2	1	2	3
	Occupation X19	1	1	1	1	1	1
	Personal income X20	1	2	2	2	2	3
	Experience of going out for work X21	2	2	2	2	2	2
Family factors	Number of family members X22	2	3	3	2	3	3
	Number of family labor force X23	2	3	2	3	3	2
	Number of male family members X24	1	2	1	2	2	2
	Number of family members X25	2	1	2	1	1	1
	Source of family income X26	1	1	1	1	1	1
	Family happiness index X27	3	2	3	3	3	4
Economic factors	Per capita income of farmers X28	3247	4563	3349	5640	5530	6742
	Source of farmers' income X29	1	1	1	1	1	1
	Per capita disposable income of farmers X30	3111	4290	3150	4890	4369	5548
Geographical factors	Distance from county highway X31	12	34	18	22	9	15
	Distance from provincial highway X32	2	5	7	4	6	13
	Distance from town center X33	6	8	9	4	7	15
	Distance from county center X34	15	18	14	26	33	18
Cognitive factors	Knowledge of information X35	2	3	2	2	3	3
	Awareness of information acceptance X36	1	2	2	1	2	2
	The ability to receive information X37	2	3	3	3	2	3
Policy factors	National informatization policy X38	1	1	1	1	1	1

3.1.6. Cognitive Factors. Cognitive factors have an important impact on rural information needs. Cognitive factors mainly include the cognitive level of rural subjects to information, the awareness of information acceptance, and the ability to receive information.

3.1.7. Policy Factors. It mainly refers to the national policy information on rural informatization. Government informatization policies, such as rural revitalization strategies, rural e-commerce, digital rural areas, and smart rural areas, affect farmers' perceptions of rural information needs.

In summary, rural information needs are affected by 38 factors in 7 aspects of the environment. This paper uses partial correlation coefficient, probability model, and ROC curve to screen and identify the factors affecting rural information demand, and finally find out the real key factors affecting rural information demand. Specific factors are shown in Table 3.

Table 6: Quantitative design of qualitative variables and its implications.

Variable	Variable name	Variable value	The meaning of variable value
X14	Age	{1, 2, 3, 4, 5}	1 = 18 years old and below, 2 = 19 to 28 years old, 3 = 29-38 years old, 4 = 39 to 48 years old, and 5 = 49 years old and above
X15	Sex	{1, 2}	1 = male and 2 = female
X18	Educational level	{1, 2, 3, 4, 5}	1 = primary school and below, 2 = junior high school, 3 = high school or technical secondary school, 4 = specialist, and 5 = undergraduate and above
X20	Personal income	{1, 2, 3, 4, 5}	1 = 1000 and below, 2 = 1000-3000, 3 = 3000-5000, 4 = 5000-7000, and 5 = 7000 above
X22	Number of family members	{1, 2, 3, 4, 5}	1 = 1, 2 = 2-3, 3 = 4-5, 4 = 6-7, and 5 = 8 and above
X16	Marital status	{1, 2, 3, 4}	1 = unmarried, 2 = married, 3 = divorced, and 4 = widow
X35	Knowledge of information	{1, 2, 3, 4, 5}	1 = conflict, 2 = is unwilling, 3 = is general, 4 = is willing, and 5 = is very willing
X36	Awareness of information acceptance	{1, 2, 3, 4, 5}	1 = is very confused, 2 = does not understand, 3 = is general, 4 = understands, and 5 = knows very well
X37	The ability to receive information	{1, 2, 3, 4, 5}	1 = is very bad, 2 = is bad, 3 = is general, 4 = is strong, and 5 = is very strong

The original data are standardized according to different data types.

Table 7: Partial factors and partial correlation coefficient calculation results.

Factor name	Fixed coverage of administrative villages X1	Number of cable TV per 1000 people X2	The ability to receive information X37	National informatization policy X38
Fixed coverage of administrative villages X1	-1.00	***	***	***
Number of cable TV per 1000 people X2	-0.07	-1.00	***	***
The ability to receive information X37	0.13	0.09	-1.00	***
National informatization policy X38	0.22	0.16	0.31	-1.00

3.2. Sample Selection and Data Sources

3.2.1. Sample Selection. Because this paper studies the rural information needs, so from the regional survey object selected as the villagers of natural villages. Considering the convenience of data acquisition and the homogeneity of sample division, and covering the plain, hilly, and mountainous terrain in the regional space, this study selected 30 natural villages of 15 townships in Lingshou County, Hebei Province, as the sample. The specific distribution is shown in Table 4.

3.2.2. Data Source. The empirical data mainly come from two aspects: the first is the statistical yearbook data of Lingshou County. Second is the survey data; this part of the data mainly includes interviews with relevant personnel data and sample survey data. The specific data is shown in Table 5.

3.3. Data Standardization

3.3.1. Standardization of Data Indicators. For the data indicators including positive, negative, and interval three categories, respectively, the above formulas are used to calculate the standardized 0-1 interval data [30–32].

3.3.2. Quantitative Processing of Qualitative Data. The qualitative data are quantified by using the Likert scale principle. The specific variable design and its meaning are shown in Table 6.

3.4. Evaluation of Influencing Factors of Rural Information Demand Based on Partial Correlation Analysis. Partial correlation analysis of standardized data is carried out to avoid the correlation of indicators only existing in data and the lack of correlation of economic significance [33]. Using the data in Table 7 and according to Equations (4)–(7), the partial correlation coefficients of each factor can be calculated by SPSS software. The results are shown in Table 8. According to the calculation results, the partial correlation coefficients of six pairs of factors are greater than 0.7, so the six pairs of factors are highly correlated and there is information redundancy. Therefore, it is necessary to further calculate the F value of six pairs of related factors. The six related factors are the number of cable TV per 1000 people and the number of TV sets per 100 households in rural areas, the number of information talents per 10000 people and the number of college students per 1000 people, the number of computers per 100 households in rural areas and the number of Internet

Table 8: Factors to be deleted based on F value.

Factors with partial correlation coefficient greater than 0.8					
Influencing factors 1	F value	Influencing factors 2	F value	Partial correlation coefficient	Deleting factors
Number of cable TV per 1000 people X2	0.014	The number of TV sets per 100 households in rural areas X8	0.023	0.84	X2
Number of information talents per 10000 people X5	0.132	Number of students per 1000 students X6	0.095	0.91	X6
The number of computers per 100 households in rural areas X7	0.059	Number of Internet users per 10000 X10	0.018	0.93	X10
Personal income X20	0.004	Per capita income of farmers X28	0.012	0.88	X20
Source of family income X26	0.236	Source of farmers' income X29	0.170	0.85	X29
Distance from county highway X31	0.301	Distance from provincial highway X32	0.332	0.87	X31

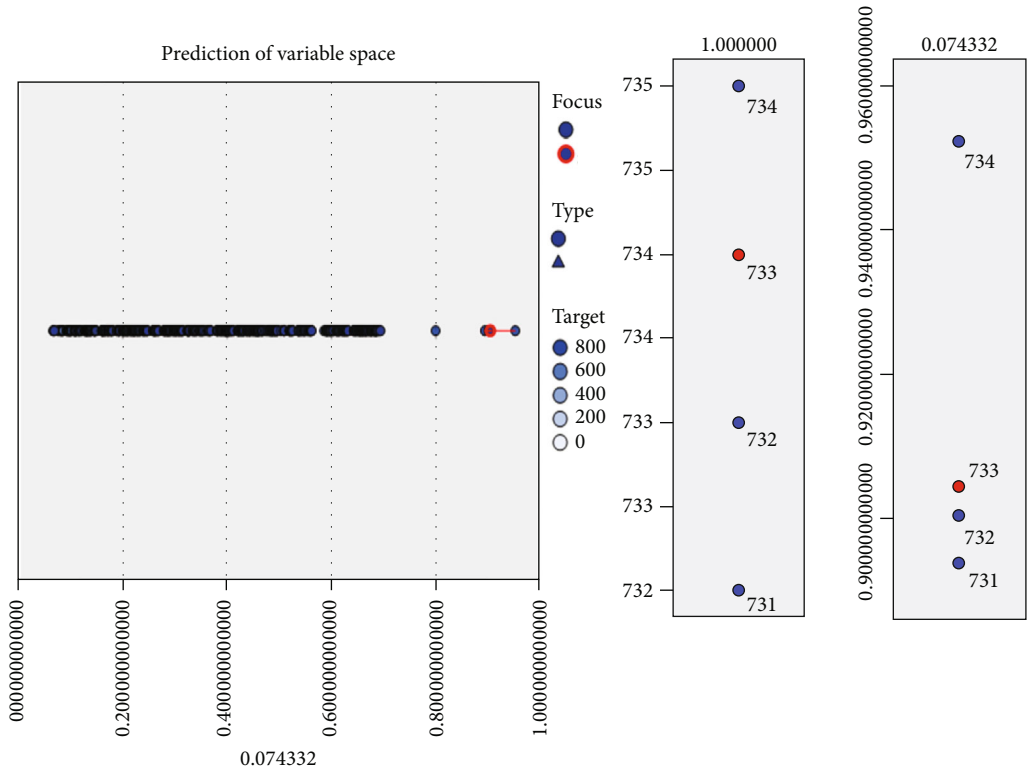

Figure 1: Clustering results of partial correlation coefficients of various factors.

users per 10000 people, personal income and per capita income of farmers, family income sources and farmers' income source, distance from county highway, and distance from provincial highway. The specific results are shown in Table 7.

The F values of six pairs of related factors are calculated, and the results are shown in Figure 1. At the same time, six pairs of factors with F values were compared, and 6 factors with smaller F value were deleted. From the data in Table 9, we can see that the F value of the number of cable TV per 1000 people is less than the F value of the number of television per 100 households in rural areas, the F value of the number of college students per 10000 people is less than the F value of the number of information personnel per 10000 people, the F value of the number of Internet users per 10000 people is less than the F value of the number of computers per 100 households in rural areas. The F value is smaller than the per capita income of farmers, and the F value of the distance between county highway and provincial highway is smaller than that of provincial highway. Therefore, six factors such as X2, X6, X10, X20, X29, and X31 with smaller F value are deleted. The specific results are shown in Table 8.

3.5. Analysis of Influencing Factors of Rural Information Demand Based on Probit Regression. On the basis of partial correlation analysis, the remaining factors are screened by using the probit regression model to find out the factors that have a greater impact on rural information demand [34]. After regression analysis of the remaining 32 factors, the relevant regression parameters were calculated. The specific results are shown in Table 9.

Table 9: Probit regression results.

Factor name	Regression coefficient	Standard error	Wald test value	Saliency probability
Fixed coverage of administrative villages X1	0.176	0.287	0.361	0.073
Optical fiber length per 100 square kilometers X3	0.384	1.236	0.445	0.129
TV coverage rate X4	0.279	0.883	0.750	0.069
Number of information talents per 10000 people X5	0.783	0.176	0.115	0.230
The number of computers per 100 households in rural areas X7	0.212	0.478	1.004	0.176
The number of TV sets per 100 households in rural areas X8	0.334	0.579	0.668	0.097
The number of mobile phones per 100 households in rural areas X9	0.913	0.336	0.654	0.209
Rural per capita postal volume X11	0.592	0.448	0.790	0.075
Fixed investment in the information industry accounts for the proportion of fixed asset investment in the whole society X13	0.398	0.667	0.085	0.148
Sex X14	-1.023	0.369	0.981	0.668
Age X15	0.359	0.783	0.245	0.189
Marital status X16	-1.382	0.450	0.695	0.033
Health X17	-0.659	0.560	0.235	0.439
Cultural level X18	0.785	0.207	0.638	0.091
Occupation X19	-3.772	0.697	0.346	0.087
Experience of going out for work X21	-2.910	0.458	0.442	0.037
Number of family members X22	0.709	0.430	0.675	0.127
Number of family labor force X23	0.559	0.650	0.283	0.076
Number of male family members X24	0.389	0.452	0.109	0.087
Number of family members X25	0.669	0.127	1.245	0.343
Source of family income X26	0.945	0.457	2.331	0.061
Family happiness index X27	0.775	0.451	0.609	0.108
Per capita income of farmers X28	0.707	0.532	0.246	0.079
Per capita disposable income of farmers X30	0.409	0.610	3.026	0.417
Distance from provincial highway X32	-1.731	0.796	0.458	0.065
Distance from town center X33	-1.088	0.569	0.337	0.098
Distance from county center X34	-3.952	0.480	0.639	0.112
Knowledge of information X35	0.515	0.707	3.041	0.046
Awareness of information acceptance X36	0.738	0.449	2.064	0.032
The ability to receive information X37	0.893	0.649	1.372	0.018
National informatization policy X38	0.776	0.985	0.689	0.050

The standard error of each factor reflects to a certain extent of the variation degree of sample average to total average [35]. The difference of standard errors of factors shows that there are certain differences in the selection of samples for each factor. However, the significance of this effect on each factor is acceptable.

In the significant probability $s > 0.01$ factor, delete the biggest factor of s value. According to this principle, we compare the s of all $s > 0.01$ factors in Table 10 to delete the largest one. Probit regression is performed on the remaining 31 factors, and the corresponding regression parameters are calculated until the s value of all the factors is less than 0.01. For example, according to the results of the first regression, all s values are less than 0.1, but the gender factor has the largest s value, so the gender factor is deleted, and then probit regression is performed again until the s value of all factors is less than 0.1. Finally, through the probit regression analysis, 8 factors such as the proportion of administrative village, gender, marital status, health status, number of family members, number of male family members, number of female family members, and family happiness index were deleted, which did not significantly affect rural information demand.

3.6. Validation and Analysis of Factors Affecting Rural Information Demand Based on ROC Curve. The data of 24 selected factors were brought into Equations (9)–(12). The probability $P(y_j = 1|X_j)$ of each village affected by relevant factors was calculated by using the probit model. When $P(y_j = 1|X_j) > 0.5$, the effect was obvious, and when $P(y_j = 1|X_j) < 0.5$, it was not.

First, the AUC value is a probability value. When you randomly select a positive sample and a negative sample, the probability that the current classification algorithm ranks the positive sample before the negative sample according to the calculated score value is the AUC value. The larger the

Table 10: AUC value of all influencing factors.

Serial number	Factor name	AUC value
1	The number of mobile phones per 100 households in rural areas X9	0.975
2	The number of computers per 100 households in rural areas X7	0.972
3	Number of family members X22	0.967
4	The ability to receive information X37	0.966
5	Optical fiber length per 100 square kilometers X3	0.96
6	Distance from county center X34	0.955
7	Number of family labor force X23	0.953
8	Awareness of information acceptance X36	0.953
9	The number of TV sets per 100 households in rural areas X8	0.947
10	Distance from town center X33	0.946
11	Experience of going out for work X21	0.942
12	National informatization policy X38	0.941
13	Knowledge of information X35	0.937
14	Source of family income X26	0.936
15	Per capita income of farmers X28	0.935
16	Fixed investment in the information industry accounts for the proportion of fixed asset investment in the whole society X13	0.933
17	Age X15	0.931
18	Distance from provincial highway X32	0.928
19	Cultural level X18	0.925
20	Occupation X19	0.921
21	Rural per capita postal volume X11	0.919
22	Number of information talents per 10000 people X5	0.917
23	Per capita disposable income of farmers X30	0.917
24	TV coverage rate X4	0.913

AUC value is, the more likely the current classification algorithm will rank the positive sample before the negative sample, so that they can be better classified.

Specifically, it is to count all $M \times N$ (M is the number of positive samples; N is the number of negative samples) positive and negative sample pairs; how many groups of positive samples have a score greater than the negative sample score. When the scores of the positive and negative samples in the binary group are equal, the calculation is performed according to 0.5. Then divide by MN. The formula for calculating the AUC value is as follows:

$$\mathrm{AUC} = \frac{\sum_{i \in \text{positive class}} \mathrm{rank}_i - (M(1+M)/2)}{M \times N}. \tag{16}$$

The ROC curve corresponding to 24 factors and the area under the curve (AUC) value were obtained by calculation. The results show that all AUC values are greater than 0.9, indicating that all factors are significantly related to rural information demand. At the same time, according to the rule that the greater the AUC value is, the more significant the demand relationship is, the order of 24 factors is ranked. The impact of every 100 households in rural areas that have mobile phones is most significant. The AUC values for the specific 24 factors are shown in Table 10.

The ROC curve is composed of dot plots of TPR and FPR corresponding to multiple critical values. Therefore, different threshold values can be used to obtain points above the multiple ROC curves, and the TPR and FPR values are used as the horizontal and vertical axes, respectively. The SPSS software draws the most significant factor. The ROC curve of the number of mobile phones per 100 households in rural areas is shown in Figure 2.

The area below the ROC curve indicates that the AUC value reflects the significant impact of the number of mobile phones per 100 households in rural areas on rural information demand. In Figure 2, AUC = 0.975 is greater than 0.9, so there are 100 rural households screened by the probit model. The number of mobile phones has a significant impact on rural information needs.

4. Conclusion

This chapter mainly analyzes and studies the information demand problem caused by the lack of rural information supply as a whole, and obtains the following conclusions:

(1) The traditional factor analysis of rural information demand does not consider the correlation between factors, so the factors themselves carry a lot of

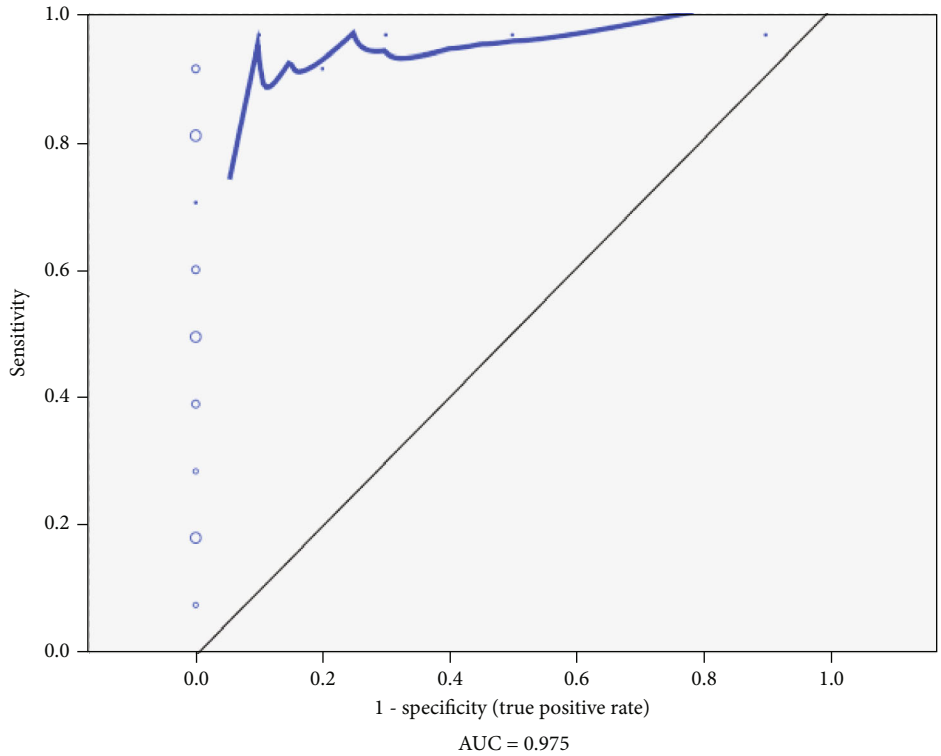

Figure 2: ROC curve of the number of mobile phones per 100 households.

redundant information, which is a certain interference to the judgment of the impact degree. Taking Lingshou County as an example, using the method of partial correlation analysis, by calculating F value, the influencing factors with highly repetitive information are eliminated, and the complexity of calculation is reduced. The probit regression model is constructed to test the influencing factors of rural information demand. Through the comparison of regression coefficient and test probability, the nonsignificant correlation of rural information demand is deleted, and ROC curve is introduced to test the above results twice, which improves the reliability of factor correlation

(2) The 24 influencing factors of rural information demand directly or indirectly affect the supply of rural information services. They provide the basis for the supply of rural information services from the seven aspects of objective environment, subject characteristics, family, economy, geography, cognition, and policy, such as improving infrastructure construction, training information service talents, and providing differentiation, and at the same time, the research results also show that the supply of rural information is related to farmers' information cognitive ability, acceptance awareness, and acceptance ability

The innovation of this paper lies in the partial correlation analysis of influencing factors of rural information demand and the ROC secondary test. It provides a new idea and method to solve the related problems.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research is supported by the "Three Three Three Talent Project" funded by Hebei Province (Project No.: A202001064).

References

[1] A. Antunes, D. Bonfim, N. Monteiro, and P. M. M. Rodrigues, "Forecasting banking crises with dynamic panel probit models," *International Journal of Forecasting*, vol. 34, no. 2, pp. 249–275, 2018.

[2] A. E. Elshaikh, X. Jiao, and S.-h. Yang, "Performance evaluation of irrigation projects: theories, methods, and techniques," *Agricultural Water Management*, vol. 203, pp. 87–96, 2018.

[3] A. M. Valente, H. Binantel, D. Villanua, and P. Acevedo, "Evaluation of methods to monitor wild mammals on

Mediterranean farmland," *Mammalian Biology*, vol. 91, pp. 23–29, 2018.

[4] A. I. Bandos, B. Guo, and D. Gur, "Estimating the area under ROC curve when the fitted binormal curves demonstrate improper shape," *Academic Radiology*, vol. 24, no. 2, pp. 209–219, 2017.

[5] U. Benjamin and U. CLN, "Libraries and information in Nigerian rural development," *International Journal of Information Management*, vol. 34, no. 1, pp. 14–16, 2014.

[6] C. A. Damalas and M. Khan, "RETRACTED: Pesticide use in vegetable crops in Pakistan: insights through an ordered probit model," *Crop Protection*, vol. 99, pp. 59–64, 2017.

[7] G. Msoffe and P. Ngulube, "Farmers access to poultry management information in selected rural areas of Tanzania," *Library & Information Science Research*, vol. 38, no. 3, pp. 265–271, 2016.

[8] H. Hu, B. Tang, X. Gong, W. Wei, and H. Wang, "Intelligent fault diagnosis of the high-speed train with big data based on deep neural networks," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 2106–2116, 2017.

[9] G. Fountas and P. C. Anastasopoulos, "A random thresholds random parameters hierarchical ordered probit analysis of highway accident injury-severities," *Analytic Methods in Accident Research*, vol. 15, pp. 1–16, 2017.

[10] G. Zhang, C. Zhang, and H. Zhang, "Improved K-means algorithm based on density canopy," *Knowledge-Based Systems*, vol. 145, pp. 289–297, 2018.

[11] H. S. Loh, Q. Zhou, V. V. Thai, Y. D. Wong, and K. F. Yuen, "Fuzzy comprehensive evaluation of port-centric supply chain disruption threats," *Ocean & Coastal Management*, vol. 148, pp. 53–62, 2017.

[12] J. A. Cook, "ROC curves and nonrandom data," *Pattern Recognition Letters*, vol. 85, pp. 35–41, 2017.

[13] J.-K. Park, S.-K. Lee, and J.-H. Kim, "Development of an evaluation method for nuclear fuel debris–filtering performance," *Nuclear Engineering and Technology*, vol. 50, no. 5, pp. 738–744, 2018.

[14] J.-F. Chen, H.-N. Hsieh, and Q. H. Do, "Evaluating teaching performance based on fuzzy AHP and comprehensive evaluation approach," *Applied Soft Computing*, vol. 28, pp. 100–108, 2015.

[15] J. J. C. Tambotoh, A. D. Manuputty, and F. E. Banunaek, "Socio-economics factors and information technology adoption in rural area," *Procedia Computer Science*, vol. 72, pp. 178–185, 2015.

[16] Y. Jin, G. Li, and H. Zhang, "Evaluation of regional rural information environment based on fuzzy method in the era of the Internet of things," *IEEE Access*, vol. 6, pp. 78530–78541, 2018.

[17] K. Kwon, J. W. Shin, and N. S. Kim, "Incremental basis estimation adopting global k-means algorithm for NMF-based noise reduction," *Applied Acoustics*, vol. 129, pp. 277–283, 2018.

[18] L. Zhang, Y. Feng, P. Shen et al., "Efficient finer-grained incremental processing with MapReduce for big data," *Future Generation Computer Systems*, vol. 80, pp. 102–111, 2018.

[19] K. Papangelis, N. R. Velaga, F. Ashmore, S. Sripada, J. D. Nelson, and M. Beecroft, "Exploring the rural passenger experience, information needs and decision making during public transport disruption," *Research in Transportation Business & Management*, vol. 18, pp. 57–69, 2016.

[20] M. de Figueiredo, C. B. Y. Cordella, D. J.-R. Bouveresse, X. Archer, J.-M. Bégué, and D. N. Rutledge, "A variable selection method for multiclass classification problems using two-class ROC analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 177, pp. 35–46, 2018.

[21] M. F. M. Firdhous and P. M. Karuratane, "A model for enhancing the role of information and communication technologies for improving the resilience of rural communities to disasters," *Procedia Engineering*, vol. 212, pp. 707–714, 2018.

[22] M. Filippini, W. H. Greene, N. Kumar, and A. L. Martinez-Cruz, "A note on the different interpretation of the correlation parameters in the bivariate probit and the recursive bivariate probit," *Economics Letters*, vol. 167, pp. 104–107, 2018.

[23] P. Mozharovskyi and J. Vogler, "Composite marginal likelihood estimation of spatial autoregressive probit models feasible in very large samples," *Economics Letters*, vol. 148, pp. 87–90, 2016.

[24] P. Matous, "Complementarity and substitution between physical and virtual travel for instrumental information sharing in remote rural regions: a social network approach," *Transportation Research Part A: Policy and Practice*, vol. 99, pp. 61–79, 2017.

[25] R. Khajouei, S. H. Gohari, and M. Mirzaee, "Comparison of two heuristic evaluation methods for evaluating the usability of health information systems," *Journal of Biomedical Informatics*, vol. 80, pp. 37–42, 2018.

[26] R. Fattahi and M. Khalilzadeh, "Risk evaluation using a novel hybrid method based on FMEA, extended MULTIMOORA, and AHP methods under fuzzy environment," *Safety Science*, vol. 102, pp. 290–300, 2018.

[27] R. H. Lange, "The predictive content of the term premium for GDP growth in Canada: evidence from linear, Markov-switching and probit estimations," *The North American Journal of Economics and Finance*, vol. 44, pp. 80–91, 2018.

[28] S. T. Yen and E. M. Zampelli, "Religiosity, political conservatism, and support for legalized abortion: a bivariate ordered probit model with endogenous regressors," *The Social Science Journal*, vol. 54, no. 1, pp. 39–50, 2017.

[29] S. Han and E. J. Vytlacil, "Identification in a generalization of bivariate probit models with dummy endogenous regressors," *Journal of Econometrics*, vol. 199, no. 1, pp. 63–73, 2017.

[30] T.-t. Gao and S.-m. Wang, "Fuzzy integrated evaluation based on HAZOP," *Procedia Engineering*, vol. 211, pp. 176–182, 2018.

[31] W. Yang, K. Xu, J. Lian, L. Bin, and C. Ma, "Multiple flood vulnerability assessment approach based on fuzzy comprehensive evaluation method and coordinated development degree model," *Journal of Environmental Management*, vol. 213, no. 1, pp. 440–450, 2018.

[32] W. Li, W. Liang, L. Zhang, and Q. Tang, "Performance assessment system of health, safety and environment based on experts' weights and fuzzy comprehensive evaluation," *Journal of Loss Prevention in the Process Industries*, vol. 35, pp. 95–103, 2015.

[33] W. Cai, L. Dou, M. Zhang, W. Cao, J.-Q. Shi, and L. Feng, "A fuzzy comprehensive evaluation methodology for rock burst forecasting using microseismic monitoring," *Tunnelling and Underground Space Technology*, vol. 80, pp. 232–245, 2018.

[34] X. Yu, W. Meng, and L. Xiang, "Comprehensive evaluation chronic pelvic pain based on fuzzy matrix calculation," *Neurocomputing*, vol. 173, Part 3, pp. 2097–2101, 2016.

[35] Y. Jin and G. Li, "Application of improved K-means algorithm in evaluation of network resource allocation," *Boletín Técnico*, vol. 55, no. 5, pp. 284–292, 2017.