Journal of Sociology and Ethnology (2023)
Clausius Scientific Press, Canada

DOI: 10.23977/jsoce.2023.050506
ISSN 2616-2318 Vol. 5 Num. 5

Construction of Social Stability Index System Based on Spearman Analysis

Hengfeng Shen[1], Henghua Shen[2], Ning Wang[1,*]

[1]College of Data Science, Guangzhou Huashang College, Guangzhou, 522000, China
[2]Capital University of Economics and Business, Jieyang, 522000, China
[*]Corresponding author

Keywords: PCA, Spearman analysis, Causal relationship, Effect relationship

Abstract: Peaceful and non-violent regime change movement has been happening since ancient times. Even in today's society, there are still many countries facing such problems. Social stability has always been one of the important issues that rulers need to consider. By reading relevant literature and checking relevant national government reports, this paper established a social stability risk assessment index system with 11 categories of first-level indicators and 45 second-level indicators. At the same time, by collecting index data from the websites of national statistical offices around the world, PCA dimension reduction analysis, Spearman correlation analysis and causation analysis were carried out on the data. The results show that there is a causal relationship between most of the variables, at the same time, there is a certain effect relationship, indicating that the selected indicators have the function of cooperative balance check. Through this index system, the risk of social stability can be well measured, and it can also provide a reference for avoiding color revolutions and maintaining social stability.

1. Construction Principles of Social Stability Risk Assessment Index System

1.1. Principle of Legality and Rationality

When assessing the risk of social stability, it is necessary to consider whether to follow the legal procedures, reflect all aspects of the policy implementation as far as possible, try to analyze and study the risks brought to social stability from more aspects, and consider whether it conforms to the law of economic and social development.

1.2. Principle of Objectivity and Subjectivity

In social stability risk assessment, it is necessary to collect relevant data and conduct quantitative analysis. Starting from the specific situation, understanding the differences and changes of risks, and combining with the results of objective analysis, can not only reflect the tolerance and authenticity of risk assessment indicators, but also accurately reflect the actual situation. Therefore, when selecting indicators, we must not have a personal subjective bias, and always maintain a neutral attitude.

1.3. Feasibility and Controllability Principles

Social stability risk assessment, in the approval stage of most social matters, social stability risk assessment should mainly focus on the feasibility of assessment, in the implementation stage of assessment, social stability risk assessment should mainly focus on the controllability of social issues. Only by combining controllability and feasibility can the introduction of the index system be feasible and operable.

1.4. Principle of Logical Relevance

To establish a social stability risk indicator system, it is necessary to consider the logical correlation between indicators. Some indicators may not have clear relationships, so further calculations are needed to display their internal connections. By calculating the connections between indicators, we can reflect specific social stability risks through indicators, and evaluate and summarize them accordingly.

2. Risk Assessment Indicators of Social Stability

2.1. Establishment of the Qualitative Perspective of the Index System

Social stability index system is an important premise of social stability early warning, fully measure all aspects of social stability index system helps to accurate prediction and maintain social stability, in addition, it also can timely help the government find the threat of social stability, take effective intervention measures, so as to realize the sustainable development of social stability.

The index system that affects social stability should include population, finance, investment environment, economy, poverty and income, education, health, energy production and use, trade, government finance, monetary stability, etc.

2.2. Establishment of the Index System of the Quantitative Angle

Selection of indicators.

As the table 1 shows, Level 1 risk assessment index layer: refers to a number of basic categories selected to assess the overall level and changing trend of social risks, including population, finance, investment environment, economy, poverty and income, education, health, energy production and use, trade, government finance and monetary stability.

Secondary risk assessment index layer: refers to the set of basic assessment indicators to assess the overall level and change trend of social risks, with a total of 45 single assessment indicators including deposit interest rate and other factors.

Table 1: Index system table

Level 1 indicators	Secondary indicators
Population	growth rate of population (%)
	crude birth rate (‰)
	Average life expectancy (years)
	total fertility (-)
	crude actual (‰)
	The urban population accounts for the total population
	infant actual (‰)
Finance	The share of stock transactions in GDP is (%)
Investment climate	The proportion of domestic credit in the banking sector to GDP is (%)
Economy	The GDP growth rate is (%)
	Per capita gross national income (current price international yuan) (international yuan / person)
	Per capita GDP (current price: international yuan) (international yuan / person)
Poverty and income	Gink coefficient (%)
	The maximum 20% of the population income in the proportion of the total income (%)
	The second 20% of the population income as a proportion of the total income (%)
	The third 20% of the population's share of total income is (%)
	The fourth 20% of the population income as a proportion of total income (%)
Education	The literacy rate of adults aged 15 years and above is (%)
	Gross enrollment rate of primary education is (%)
	Gross enrollment rate of secondary education is (%)
	Gross enrollment rate of higher education is (%)
Hygienism	The incidence of malnutrition in the whole population is (%)
	Beds per thousand people (one / thousand people)
	Doctors per thousand people (person)
Energy production and use	Per capita energy consumption (kg standard oil equivalent / person)
	Flammable renewable energy and waste consumption accounted for the total energy consumption proportion of (%)
	Net energy imports are a share of energy in energy consumption (%)
Trade	The export of agricultural raw materials accounted for the export of goods (%)
	Food exports accounted for the proportion of goods exports of (%)
	The proportion of fuel exports in the export of goods is (%)
	Exports of finished goods accounted for (%) of goods
	Exports of minerals and metals accounted for (%) in exports of goods
	Imports of agricultural raw materials accounted for the import of goods (%)
	Food imports accounted for the proportion of goods imports of (%)
	Fuel imports account for the proportion of goods imports of (%)
	Imports of manufactured goods accounted for imports of goods (%)
	Imports of minerals and metals account for (%) proportion of imports of goods
Government finance	The proportion of central government fiscal revenue in GDP is (%)
	The proportion of goods and services tax in fiscal revenue is (%)
	The proportion of social contributions in fiscal revenue is (%)
	The central government public debt to GDP (%)
Monetary stability	deposit rate (%)
	lending rate (%)
	Currency exchange rate (annual average price) (local currency / USD)

3. Correlation and Causal Analysis of the Index System

3.1. Data Collection

Table 2: Data Summary Table (in part)

country	growth rate of population (%)	crude birth rate (‰)	Average life expectancy (years)	total fertility (-)	crude actual (‰)	infant actual (‰)	The share of stock transactions in GDP is (%)	The proportion of domestic credit in the banking sector to GDP is (%)	The GDP growth rate is (%)	Per capita gross national income (current price international yuan) (international yuan
Algeria	1.93	23.5	76.8	2.99	4.72	24.3	19.191	21.8	1	11710
Angola	3.24	40.2	61.1	5.44	7.98	49.9	19.191	27.2	-0.7	6360
Benin	2.72	35.8	61.7	4.77	8.72	57.9	19.191	21.4	6.87	3390
Botswana	2.18	24.2	69.5	2.84	5.71	36.9	19.191	33.8	3.35	16200
Burundi	3.13	38.3	61.58	5.32	7.77	24.32	19.191	14	1.81	790
Cameroon	2.58	34.8	59.2	4.51	9.06	49.9	19.191	16.3	3.48	3820
Cape Verde	1.13	19.0	72.9	2.24	5.58	24.32	19.191	61.1	5.67	7320
Central African	1.67	35.0	53.2	4.65	11.9	79.1	19.191	12.3	3.1	21006.941
Chad	2.99	41.6	54.2	5.65	11.9	24.32	19.191	9.9	3.25	21006.941
Comoros	2.21	31.4	64.3	4.14	7.14	48.7	19.191	26.6	1.76	3120
Congolese (cloth)	2.56	32.4	64.5	4.37	6.61	35.9	19.191	22	-0.09	3130

Data source: https: www.stats.gov.cn/tjsj/, Data collection is the most important step in this process. Incomplete or incorrect data collection will directly affect the accuracy of the following information. In the process of data collection, we can clearly feel that there are still many shortcomings in the current data statistical work:

(1) Although China has established social statistics, including health, environment and population, the statistical work in other countries is not complete, and sometimes different data will appear for the same content.

(2) The most perfect data is still in the economic field, and the data in other fields is not complete enough.

(3) As far as the current situation is concerned, it is difficult to collect the data of objective indicators, but it can still be found. The most difficult thing is some subjective index data. Because the database is not comprehensive and systematic, the subjective data is very lacking, which greatly affects the design of the index system.

According to the table 2, this paper selects from the overseas database of the National Bureau of Statistics to the yearbook statistical data that are similar or directly calculated. Several of them can be found for the original data in the yearbook center. The 2019 Yearbook statistics are shown in the table (due to space constraints, only some data will be displayed):

3.2. Spearman Correlation Coefficient Analysis

Spearman's correlation[1,2] coefficient was defined as between rank variables, and for samples with a sample size of n, n raw data were converted into rank coefficients, with the correlation coefficient ρ being.

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \tag{1}$$

The raw data is assigned a corresponding rank based on its average descending position in the overall data, as shown in the table 3:

Table 3: Raw Data Level Allocation Table

variable Xi	Decline position	grade xi
0.8	5	5
1.2	4	4
1.2	3	3
2.3	2	2
18	1	1

In practice, the connection between variables is insignificant, so ρ can be calculated in simple steps. For the difference between the grades of the two observed variables, ρ is:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)} \tag{2}$$

The heat map is shown in Figure 1:

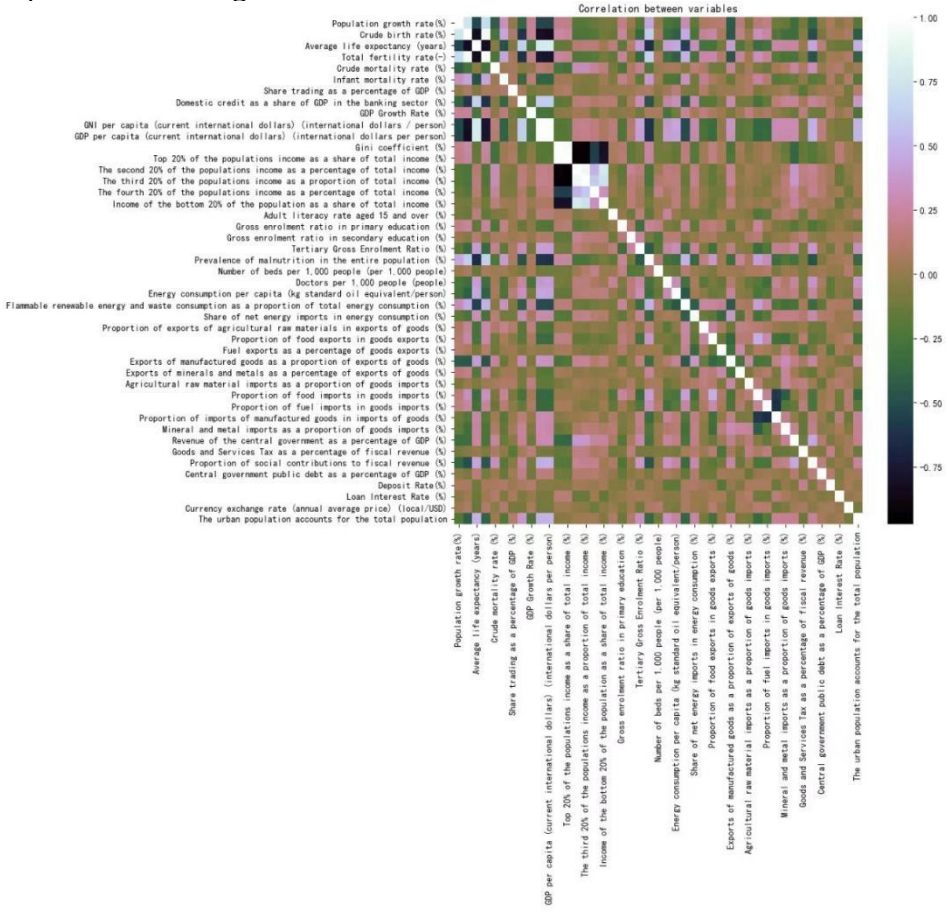

Figure 1: Heat map of the raw data

Above for secondary index spieman's correlation coefficient analysis, due to figure 1, here through the heat map, through the heat map, we can find that most of the index correlation degree is below, most concentrated in between 0.5~0.5, the closeness between variables is not high, and because the index data, in order to be able to better analyze the variable correlation, we will use the principal component analysis of secondary index.

3.3. Dimensional Reduction Analysis of Principal Components

Principal component analysis method[3-6] is a multivariate statistical method that cites the idea of "dimension reduction" and transforms multiple indicators into a few comprehensive indicators. The comprehensive index here is the principal component. Each principal component is a linear combination of the original variables, independent of each other, and retains the vast majority of the information of the original variables. Its essence is to seek the comprehensive replacement object of the related variables through the correlation of the original variables, and to ensure the minimum information loss in the transformation process.

The covariance matrix R was calculated from the normalized dataset:

$$R = (r_{ij})_{n*n} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix} \tag{3}$$

First solve the eigenequation | λ I-R | =0 and find the eigenvalue λ i (i=1,2,..., p), and arrange it in size order, namely λ 1 λ 2..., λ p 0; then find the eigenvector ei (i=1,2, λ i,..., p).

$$r_i / \sum_{k=1}^{p} \gamma_k (i = 1, 2, \cdots, p) \tag{4}$$

Principal component contribution rate:

$$\sum_{k=1}^{m} \gamma_k / \sum_{k=1}^{p} \gamma_k \tag{5}$$

Generally, the characteristic values λ1, λ2, λm, λm correspond to the first, second,..., m (mp).

Calculate the principal component load

$$p(z_k, x_i) = \sqrt{\gamma_k} e_{ki} (i, k = 1, 2, \cdots, p) \tag{6}$$

The summary results are shown in the figure below (only some data are displayed due to space constraints):

Table 4: Data after PCA dimension reduction (partial)

population	finance	economy	Poverty and income	education	hygienism	Energy production and use	investment climate	trade	government finance	Monetary stability
0.330	0.078	0.113	0.395	0.377	0.002	0.067	0.145	0.479	0.253	0.282
0.738	0.100	0.060	0.395	0.411	0.296	0.022	0.145	0.826	0.247	0.282
0.777	0.077	0.026	0.395	0.110	0.102	0.015	0.145	0.741	0.253	0.282
0.480	0.12	0.162	0.395	0.411	0.376	0.066	0.145	0	0.044	0.282
0.499	0.04	0	0.395	0.064	0.146	0.147	0.145	0.741	0.253	0.282

The table 4 shows Data after PCA dimension reduction, after another Spearman correlation

coefficient analysis, the results are as follows:

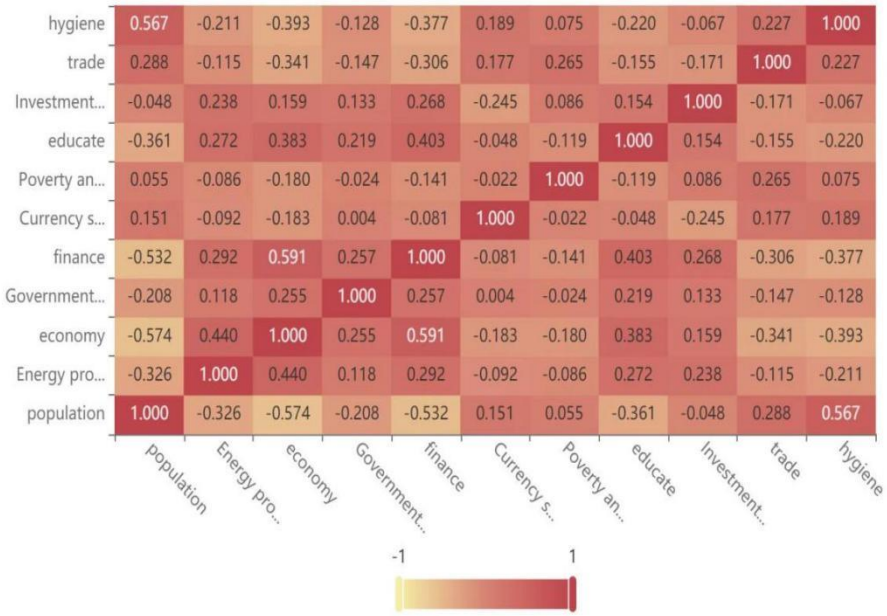

Figure 2: Heat map of the correlation coefficient

Through figure 2 we can find that population related indicators of financial economy, poverty and income, education, health, energy production and use, investment environment, trade, government fiscal and monetary stability have shown significant correlation, and the financial, health, and for financial, its investment environment, monetary stability presents the significant correlation, at the same time to the poor and income, health, trade, monetary stability is a negative impact.

More correlations between the indicators can be intuitively seen from Figure 2, so overall, we can see a higher degree of correlation between these indicators. In order to better analyze the correlation between these indicators, that is, the significant analysis between the indicators while excluding the interference of other variables, we then conducted a causal analysis on the first-level indicators.

3.4. Causality and Effect Relationship

In a broad sense: "causal Relation refers to the relationship between an event and another event, in which the previous event is considered to be the cause of the latter event and the latter event is considered to be the result of the previous event".

Stratified regression[7,8] is an effective method to solve causal analysis, which is essentially based on regression analysis. The difference is that hierarchical regression can be divided into multiple layers due to the differences between two or more regression models. The stratified regression puts the variable of the core study in the final step into the model to examine the contribution of this variable to the regression equation, when the contribution of other variables were excluded. If the variable still has a significant contribution, then it can be concluded that the variable indeed has a unique role that cannot be replaced by other variables. This method is mainly applicable to cases where there are high correlation between party variables and the unique contribution of one independent variable is difficult to determine.

Table 5: Results table of stratified regression

Hierarchical regression								
	operating floor				Hierarchical 1			
	B	standard error	t	P	B	standard error	t	P
constant	0.436	0.06	7.2	0.000***	0.446	0.059	7.538	0.000***
economy	-0.28	0.055	-5.111	0.000***	-0.203	0.058	-3.479	0.001***
Poverty and income	-0.124	0.073	-1.708	0.089*	-0.133	0.071	-1.871	0.063*
education	-0.19	0.078	-2.442	0.015**	-0.14	0.077	-1.81	0.072*
hygienism	0.362	0.053	6.904	0.000***	0.329	0.052	6.308	0.000***
Energy production and use	-0.171	0.119	-1.443	0.15	-0.181	0.116	-1.558	0.121
investment climate	0.135	0.072	1.892	0.060*	0.185	0.071	2.593	0.010**
trade	0.075	0.044	1.698	0.091*	0.063	0.044	1.438	0.152
government finance	-0.064	0.084	-0.762	0.447	-0.042	0.082	-0.514	0.608
Monetary stability	0.029	0.119	0.245	0.807	0.071	0.117	0.611	0.542
finance					-0.275	0.082	-3.333	0.001***
R^2	0.503				0.529			
adjust R^2	0.481				0.505			
F	F (9, 215) =23.04, P=0.000***				F (10, 214) =22.869, P=0.000***			
$\triangle R^2$	0.503				0.026			
$\triangle$F price	F (9, 215) =23.04, P=0.000***				F (1, 214) =11.106, P=0.000***			

The above table 5 shows the test results of the model and the parameters of the model, including the coefficient and P-value of the model, which can be used to analyze the formula of the model.

Based on the hierarchical model[9,10]: control layer, including field constant, monetary stability, education, government finance, trade, investment environment, poverty and income, health, economy, energy production and use, significant P value of 0.000 * * *, the level is significant, reject the original hypothesis, so the model is effective, and the degree of goodness of fit model R^2 is 0.503, the model performed well. Based on hierarchical model: layered 1, including field constant, monetary stability, education, government finance, trade, investment environment, poverty and income, health, economy, energy production and use, finance, significant P value is 0.000 * * *, the level is significant, reject the original hypothesis, so the model is effective, and the model of fit R^2 is 0.529, the model performed well.

So we can see the population as the independent variable, financial as the dependent variable, control the investment environment, economy, poverty and income, education, health, energy production and use, trade, government fiscal, monetary stability variables, can be found in financial level a significant model, which illustrates the interference of other variables, there are significant correlation between financial and population, namely the population is the financial, financial is the result of the population. Because the comparison process is redundant, it will not be analyzed here. The following are the results of two-wise regression.

Table 6: Summary table of the results of the pairwise level regression

	population	finance	economy	Poverty and income	education	hygienism	Energy production and use	investment climate	trade	government finance	Monetary stability
population	1	0.001***	0.001***	0.063*	0.072*	0.000***	0.121	0.010**	0.152	0.608	0.542
finance		1	0.000***	0.354	0.023**	0.266	0.495	0.000***	0.369	0.326	0.099*
economy			1	0.218	0.200	0.614	0.000***	0.433	0.066*	0.163	0.090*
Poverty and income				1	0.305	0.695	0.429	0.015**	0.000***	0.617	0.543
education					1	0.829	0.197	0.38	0.628	0.149	0.706
hygienism						1	0.951	0.856	0.646	0.937	0.098*
Energy production and use							1	0.003***	0.221	0.72	0.656
investment climate								1	0.057*	0.278	0.001***
trade									1	0.458	0.142
government finance										1	0.286
Monetary stability											1

From the above table 6, we can find that population and energy production are not significantly associated with use, trade, government finance and monetary stability after removing interference. Finance and poverty are not significantly associated with income, health, energy production and use, trade, and government finance after removing interference. Economy and poverty are not significantly related with income, education, health, investment environment, and government finance after removing interference. Poverty and income are not significantly associated with education, health, energy production and use, government finance, and monetary stability after removing interference. Education has no significant correlation with health, energy production and use, investment environment, trade, government finance, and monetary stability after removing interference. There is no significant correlation between energy production and use, investment environment, trade, and government finance after removing interference. Energy production and use are not significantly related with investment environment, trade and government finance after removing interference. There is no significant correlation between the investment environment and the government finance after removing the interference. There is no significant correlation between government fiscal stability and monetary stability after removing interference.

In conclusion, it can be found that after excluding the interference of variables, there can be a causal relationship between many variables, which further proves that the selected indicators have a coordinated check and balance effect.

4. Conclusion

Based on the four principles of the index system, namely legality and rationality, objectivity and subjectivity, feasibility and controllability, and logical correlation, this paper gives the initial index system through literature research, and then confirms the final index system by reviewing relevant reports of the national government. After obtaining the indicator system, based on the overseas database of the National Bureau of Statistics and the official website data of the statistical bureaus of various countries, we use the Spearman correlation coefficient to analyze the correlation between indicators, due to too many indicators, resulting in the poor correlation of some indicators, at this time we use principal component dimensionality reduction analysis, compress the data into a primary or secondary index, and then carry out correlation analysis, correlation analysis has the

shortcomings of interference variables, so in order to further analyze the causal relationship within the data, We used hierarchical regression for secondary analysis.

The social stability index system established in this paper can provide a good guiding role for avoiding color revolutions and maintaining social stability, and also proves the rationality and accuracy of this index system.

Acknowledgement

This project is supported by the Youth Academic Program of Guangzhou Huashang University. Project funding: 2022HSXS089. Project Name: Research on Smart Community Management.

References

[1] Wang Yuanyuan, Bai Hongkun, Wang Shiqian, Bu Feifei, Wu Xiong, Li Haoyu. Power User Behavior Portrait Based on Information Gain and Spearman Correlation Coefficient. Electric Power Engineering Technology, 2022, 41 (04): 220-228. (in Chinese)

[2] Chen H. Analysis of water quality change characteristics in Shanghai based on Spearman rank correlation coefficient method. Environmental science and technology, 2020 (3): 28 and 33, DOI: 10.19824/j.carol carroll nki cn32-1786/x. 2020.0037.

[3] Li Ying, Ge Xizhen. Construction of cucumber downy mildew early warning system in greenhouse based on PCA and multiple regression algorithm. Journal of Anhui Agricultural Sciences, 2022, 50 (21): 232-234.

[4] Mo Yun, Guo Yan, Mo Hesheng, Lu Zhongwei, Zhang Shaorong. Hybrid feature selection method based on feature filtering and PCA dimension reduction. Journal of Guilin University of Astronautics Technology, 2022, 27 (02): 145-151. (in Chinese)

[5] Hou Meien, Wang Xiangbin. Extraction of biomechanical characteristics of gait in KOA patients by PCA. Medical Biomechanics, 2021, 36 (S1): 224.

[6] Wang Xueli. A study on the Effect of Online celebrities' Live streaming on Consumers' Purchase Intention. Gansu Science and Technology, 2020, 49 (11): 76-80.

[7] Cong Xiaoqi. Research on Entrepreneurial self-efficacy of accounting based on hierarchical regression. Journal of Science and Technology Entrepreneurship, 2020, 33 (08): 129-131.

[8] Liu Xiaojun. Study on Comprehensive Benefit Evaluation of Green construction Land saving Technology based on topsis entropy weight method. Mathematics practice and understanding: 1-10 [2023-02-05].

[9] Wang Zitao, Wang Jianping, Han Guang. Evaluation of groundwater richness based on entropy weight TOPSIS method in Qaidam Basin. Salt Lake Research, 2022, 30 (02): 42-51.

[10] Hu Mingqi, Zhang Senchang. Leaf node weighted random forest algorithm based on PSO optimization. Modern Computer, 2022, 28 (04): 1-4.