Problem Chosen	2023HSB MCM/ICM Summary Sheet	Team Number
B		MI01370

Abstract

It is crucial to accurately measure social stability and judge social risks in order to avoid social unrest. In this paper, based on the selection of indicators that can comprehensively reflect various aspects of social stability, a social stability indicator system and a social stability early warning model based on PSO-C4.5 decision tree were established to avoid social unrest.

For problem 1, based on the four principles of social stability indicator system construction, this paper formed a social stability evaluation indicator system including 11 primary indicators and 50 secondary indicators. For all countries in the world, through various authoritative database websites, collect and collate the indicator data, and conduct qualitative analysis. Considering the large number of indicators, PCA dimension reduction was used to reduce the dimensions of indicators, and Spearman correlation coefficient and hierarchical regression were used to analyze the correlation and causality among indicators. The analysis results showed that there was a causal relationship among indicators, indicating that the selected indicators had a synergistic and balancing effect.

For problem 2, on the basis of the establishment of the evaluation indicator system of social stability, this paper uses entropy weight TOPSIS method to conduct a comprehensive score for the social stability of 215 countries, and codes the comprehensive score to form three levels of social stability, namely weak, medium and strong. Then, a C4.5 decision model was established, and a social stability early warning model with optimal parameters was determined by using PSO for super-parameter optimization. Its evaluation index F1 value was 0.84, indicating excellent performance. At the same time, the decision division threshold of each indicator in the model is displayed to provide strong support for the early warning of social stability indicators.

In view of problem 3 and problem 4, Cambodia and Vietnam are selected as the failed and successful countries of color revolution for analysis. Both of them have some similarities in geographical location and natural resources, so the analysis results are comparable. Based on the early warning model of social stability and through visual machine learning SHAP model, it is found that the main factors affecting the color revolution of the two are investment environment, economy, monetary stability, education and hygiene. From the data view, it is the difference of these indicators that leads to the different outcomes of the two kinds of color revolutions.

In view of problem 5, based on the SHAP model, the importance of each indicator is analyzed. Among them, economy, investment environment, finance, government finance and education contribute the most to social stability. However, due to the mutual checks and balances of each evaluation index and the complex relationship, it is necessary to consider the social stability index and combine the potential social and political factors causing the color revolution. Suggestions for maintaining social stability are given.

Keywords: Social stability indicator system; entropy weight TOPSIS method; PSO-C4.5 decision tree model; SHAP model

Contents

1 Introduction

1.1 Problem Background

Human beings, like all animals, have the instinct to seek advantage and avoid harm. The key to humans being the lord of creation is that they are better at avoiding harm than other animals. Crises are always lurking in the future. "Thinking of danger in times of peace", measuring and warning social stability, has always been highly valued by far-sighted rulers and politicians.

However, it is not easy to overcome the crisis in the increasingly competitive and crisis-prone modern society. We must rely on modern social measurement science and modern social early warning science. Social measurement and social early warning are closely related. In order to accurately conduct early warning of social crisis, it is necessary to find out several factors affecting social stability. There are many factors that affect social stability, such as survival security, economic support, social distribution, social control, social psychology, external environment, etc. These factors intertwine and influence each other, and together support all aspects of society. Once there is a large fluctuation in one factor, it will directly or indirectly affect other factors in society, which will lead to unrest or collapse of society.

The multiple factors are originally an organic whole, but once there is an imbalance in the ratio of the forces acting on the factors, society become deformed and social stability falters. Then, before the emergence of social stability problems, if you can detect some "traces" in some of the precursors, timely intervention, may be timely to avoid social risks, to avoid unnecessary social unrest.

1.2 Restatement of the Problem

(1) The indicator system of social stability is an important prerequisite for social stability early warning. Selecting indicators that are representative of each can reflect all aspects of social stability comprehensively. Please establish a system of indicators that are likely to affect social stability from both qualitative and quantitative perspectives, and discuss the correlation and causality between them.

(2) A stable society requires synergy, checks, and balances among indicators. For example, a society in economic hardship may use the human spirit to compensate for economic deficiencies, which can also make society desperately in need of stability, but only if it can survive. Please consider similar ideas, develop an early warning model of social stability, and discuss.

(3) Select a country or region where a color revolution attempted to overthrow the regime, and evaluate its social stability by using the established social stability early warning model. This paper points out the main reasons for the failed color revolution, judges the trend of social stability in the future, and puts forward some suggestions.

(4) Please select a country or region where color revolution leads to regime change and point out the main reasons for regime change by using the established social stability early warning model.

(5) In order to prevent color revolution and maintain social stability, please put forward relevant suggestions.

2 Problem Analysis

2.1 Analysis of problem 1

The purpose of problem 1 is to establish an indicator system for affecting social stability, discussing the correlation and causality among indicators. On the basis of reading relevant literature, according to the principle of index selection, select indicators that can reflect all aspects of the society stability. In order to fully reflect all aspects of social stability, a large number of indicators are selected, and in consideration of the establishment of problem 2 model, these indicators should also be used. If only the data of a certain country is used for modeling, the applicability of the model is inadequate. Therefore, in countries around the world, data collection and sorting of indicators is carried out, and all indicators are qualitatively analyzed. Due to the large number of indicators, the analysis is too complicated, so the PCA dimension reduction method is utilized to reduce the dimension of indicators. Then, Spearman correlation coefficient and hierarchical regression were utilized to analyze the correlation and causality among all indica-tors.

2.2 Analysis of problem 2

Problem 2 proposes that a stable society requires synergy and checks and balances among indicators, and requires the establishment of an early warning model of social stability. In problem 1, the selected indicators are logically correlated with each other. Based on the selection of indicators in problem 1, the entropy weight method and TOPSIS method are combined to obtain the comprehensive score of social stability of each country, and the data coding is used to rank the social situation of each country in the world. After that, C4.5 decision tree model is constructed and PSO hyperparameter optimization is carried out to establish an early warning model of social stability based on PSO-C4.5, and the decision process of each indicator is plotted, which is conducive to the early warning of each indicator of social stability index system.

2.3 Analysis of problem 3 and problem 4

Problem 3 requires to select a country with failed color revolution to point out the main reasons for the failure of color revolution and judge the trend of social stability in the future, while problem 4 requires to select a country with successful color revolution to carry out the reasons for regime change. The third and fourth problems are combined for consideration, and Cambodia and Vietnam are selected respectively. Both of them are located in Southeast Asia and have similarities in geographical location and natural resources. The analysis results are comparable, which is helpful to understand the indicators affecting social stability and the change process of social stability. On the basis of PSO-C4.5 social stability model, through visualization and machine learning SHAP model, the main reasons for the failure and success of color revolution are analyzed. On the basis of data feedback, political and popular factors are combined to make predictions and put forward suggestions

2.4 Analysis of problem 5

Problem 5 requires relevant suggestions to prevent color revolutions and maintain social stability. On the basis of model building and example analysis, the most important indicators affecting social stability are further explored, and the logical correlation among the indicators is considered, and the impact of increasing or decreasing the value of each indicator on social stability is analyzed. Finally, suggestions for avoiding color revolutions are put forward by combining potential popular and political factors in society.

Figure 1 is a diagram of the overall idea of our problem solving.

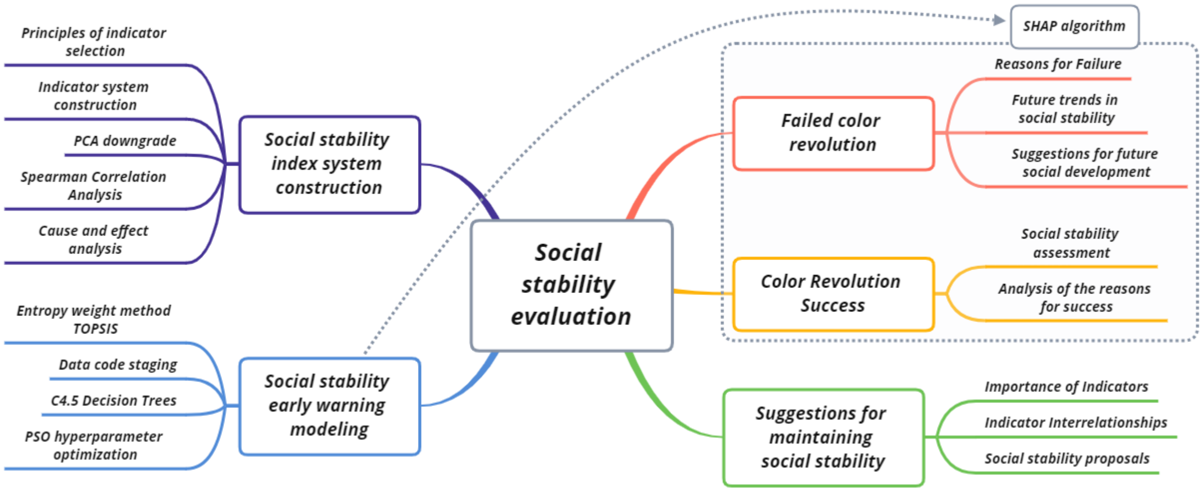

Fig. 1 Overall problem solving idea diagram

3 Assumptions and Justifications

(1) Assume that all the data used in this paper are true. The data used in this paper are all from various authoritative websites, which ensures the authenticity of the data to the greatest extent.

(2) Assuming there are no extreme global catastrophes in the next few years. Extreme catastrophes can be devastating to a country, rendering predictions and analysis invalidated.

(3) Assume that there are only 215 countries used in this article. The number of countries in the world is affected by the degree of recognition in different regions. In this paper, 215 countries according to the statistics of the National Bureau of Statistics of China are adopted.

4 Notations

The key mathematical notations used in this paper are listed in Table 1.

Table 1. Notations used in this paper

Symbol	Description
R	Covariance matrix
ρ	Correlation coefficient
$D+$	Positive ideal solution distance
$D-$	Negative ideal solution distance
$C1$、$C2$	Acceleration constants
$val\ (s)$	Model output value under feature combination S
R^2	Model fitting degree
P	Accuracy rate

5 Modeling and solving of Problem 1

5.1 Selection of social stability indicators

5.1.1 Principles of indicator selection

An indicator is a specific concept and a specific value reflecting the overall phenomenon, through a specific indicator, it is possible to quantify a certain characteristic of the phenomenon under study. The indicator system is a scientific and complete collection of a limited number of indicators, through which a comprehensive understanding of the assessment object and its development law can be grasped[1,2]. As shown in Figure 2, in order to make an objective and realistic assessment effect, four principles are satisfied when establishing the social stability indicator system.

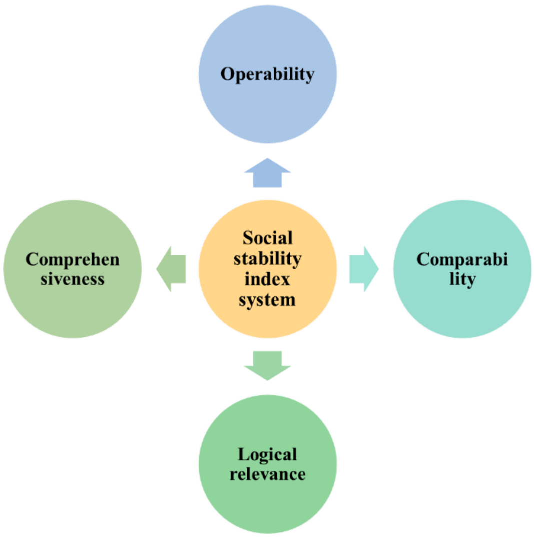

Fig. 2 Principles of social stability index system selection

5.1.2 Results of indicator selection and qualitative analysis

Based on the above principles of index selection, a total of 11 primary indicators and 50 secondary indicators were selected from the aspects of survival security, economic support, social distribution, social control, social psychology, and external environment, as shown in Table 2, to initially establish an evaluation index system that can reflect social stability.

Table 2 Evaluation index system

Primary indicators	Secondary indicators
F1 Population	S1 Growth rate of population
	S2 Crude birth rate
	S3 Average life expectancy
	S4 Total fertility rate
	S5 Crude death rate
	S6 Urban population accounts for the total population
	S7 Urban population
	S8 Rural population
F2 Finance	S9 Shares traded as a share of GDP
F3 Investment climate	S10 Banking sector domestic credit as a share of GDP
	S11 Air passenger volume
	S12 Volume of container shipping

Primary indicators	Secondary indicators
F4 Currency stabilization	S13 deposit interest rate
	S14 loan interest rate
	S15 Currency exchange rate (Annual average price)
F5 Poverty and income	S16 Gini coefficient
	S17 The top 20% of the population as a share of total income
	S18 The second 20% of the population as a share of total income
	S19 The third 20% of the population as a share of total income
	S20 The fourth 20% of the population as a share of total income
	S21 The share of the bottom 20% of the population in total income
F6 Education	S22 Adult literacy rate 15 years and older
	S23 Gross enrolment rate of primary education
	S24 Gross enrolment rate of secondary education
	S25 Gross enrollment rate of higher education
F7 Hygiene	S26 Prevalence of malnutrition in the entire population
	S27 Number of beds per 1000 people
	S28 Number of doctors per 1000
F8 Energy production and use	S29 Energy consumption per capita
	S30 Flammable renewable energy and waste consumption account for the proportion of total energy consumption
	S31 Share of net energy import in energy consumption
F9 Trade	S32 The proportion of agricultural raw material exports in goods exports
	S33 The proportion of food exports in goods exports
	S34 The proportion of fuel exports to goods exports
	S35 The proportion of manufactured exports in goods exports
	S36 Mineral and metal exports as a share of goods exports
	S37 The proportion of imports of agricultural raw materials in imports of goods
	S38 The proportion of food imports in goods imports
	S39 The proportion of fuel imports in goods imports
	S40 The proportion of imports of manufactured goods in imports of goods
	S41 Mineral and metal imports as a proportion of goods imports
F10 Government finance	S42 Share of central government revenue in GDP
	S43 Goods and services tax as a share of revenue
	S44 The proportion of social contributions in government revenue
	S45 Central government public debt as a share of GDP
F11 Economy	S46 GDP growth rate
	S47 Per capita gross national income
	S48 Gross domestic product per capita
	S49 Gross national income
	S50 Gross domestic product

The target level refers to the general goal to be achieved by the social stability risk assessment of various countries. The purpose is to accurately measure the level of social stability, and the result can provide a basis for the government to make policies or social governance in the future.

Due to space limitation, this paper takes population, resources and labor force as an example to make a brief qualitative analysis of the relationship between the indicators. Population has an impact on social stability, as a rapidly growing population puts pressure on resources and infrastructure, leading to problems such as unemployment, poverty and inequality. On the other hand. The ageing of the population will lead to a reduction in the Labour force, which will also affect economic stability, indicating a good logical relationship between the selected indicators.

(1) Data sources

The data are mainly from the following databases and websites, as shown in Table 3.

Table 3 Data sources

Name	Websites
The World Bank	https://www.worldbank.org/
Organization for Economic Cooperation and Development	http://www.oecdchina.org/
Wikipedia	https://www.wikipedia.org/

(2) Missing value padding

The searched data has partial missing, which is random and belongs to random missing, so the mean replacement method is used to fill the missing value. Mean replacement method is also a simple and fast method to deal with missing data. Using the mean replacement method to interpolate the missing data has no effect on the estimation of the mean value of the variable.

Table 4 Data of indicators in each country

Country	**S1**	**S2**	**S3**	……	**S50**
Algeria	1.93	23.58	4.72	……	1
Angola	3.24	40.23	7.98	……	-0.7
Benin	2.72	35.83	8.72	……	6.87
Botswana	2.18	24.23	5.71	……	3.35
……	……	……	……	……	……
Former southern Macedonia	1.26	19.62	7.51	……	2.73

5.2 Analysis of social stability indicators

5.2.1 PCA descending

A total of 50 secondary indicators were selected, which can comprehensively reflect social stability, but too many indicators are not conducive to further analysis, so PCA dimensionality reduction is used to fuse the indicators, fusing secondary indicators into primary indicators, reducing the number of indicators that need to be analyzed while minimizing the loss of information contained in the original indicators, in order to achieve a comprehensive analysis of the collected data. the steps of PCA dimensionality reduction are as follows.

(1) Data Standardization

Since the range interval of each data is different and the span of the data interval is relatively large, the data normalization is used to narrow down to the interval [0, 1], which unifies the data change interval and also enables different independent variables to have a better presentation on a graph, which is more intuitive to analyze the change trend and brings a great convenience to do visibility prediction subsequently; the data normalization uses the maximum-minimum method, and the formula of the Min-max method is as follows.

$$x_{ij}^{*} = \frac{x_{ij} - m_j}{M_j - m_j} \tag{1}$$

Where

$$M_j = \max_i \{x_{ij}\}, \cdots\cdots m_j = \min_i \{x_{ij}\} \quad (2)$$

For extremely small indicators, the above equation becomes:

$$x_{ij}^* = \frac{M_j - x_{ij}}{M_j - m_j} \quad (3)$$

After normalized processing, the data is sorted into Excel for the convenience of subsequent analysis[3].

(2) Calculate the covariance matrix R

$$\mathbf{R} = (r_{ij})_{n\times n} = \begin{bmatrix} r_{11} & r_{12} & \therefore 1 & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & 3 & & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix} \quad (4)$$

Calculate the eigenvalue $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n \geqslant 0$ of matrix R and the corresponding eigenvector $u_n, u_2, \cdots, u_n$, where $u_j = (u_{1j}, u_{2j}, \cdots, u_{nj})$ and u_{nj} represent the n-th component of the j-th eigenvector; n new index variables are composed of feature vectors:

$$\left.\begin{aligned} y_1 &= u_{11}\overline{x}_1 + u_{21}\overline{x}_2 + \cdots + u_{n1}\overline{x}_n \\ y_2 &= u_{12}\overline{x}_1 + u_{22}\overline{x}_2 + \cdots + u_{n2}\overline{x}_n \\ &\vdots \\ y_n &= u_{1w}\overline{x}_1 + u_{2n}\overline{x}_2 + \cdots + u_{nn}\overline{x}_{n^*} \end{aligned}\right\} \quad (5)$$

Where, y_1 is the first principal component, y_2 is the second principal component... y_n is the *n*-th principal component. Calculate the cumulative contribution rate α_p of contribution rate $b_j (j = 1, 2, \cdots, n)$ and $y_1, y_2, \cdots, y_n (p \leqslant n)$ of each principal component y_j.

$$b_j = \frac{-\lambda_j}{\sum_{k=1}^{n} \lambda_k} \quad (j = 1, 2, \cdots, n) \quad (6)$$

$$\alpha_p = \frac{\sum_{k=1}^{k} \lambda_k}{\sum_{k=1}^{n} \lambda_k} \quad (p \leqslant n) \quad (7)$$

Due to space constraints, the intermediate process is not explained in detail, and only part of the results are shown, as shown in Table 5:

Table 5 PCA analysis result

Country	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11
Algeria	0.33	0.08	0.15	0.40	0.39	0.38	0.002	0.07	0.48	0.25	0.11
Angola	0.74	0.10	0.15	0.40	0.39	0.41	0.296	0.02	0.83	0.25	0.06
Benin	0.78	0.08	0.15	0.40	0.39	0.11	0.102	0.02	0.74	0.25	0.03
Botswana	0.48	0.13	0.15	0.40	0.39	0.41	0.376	0.07	0	0.04	0.16
……	…	…	…	…	…	…	…	…	…	…	…
Former southern Macedonia	0.33	0.19	0.15	0.28	0.4	0.41	0.15	0.147	0.48	0.25	0.21

5.2.2 Correlation analysis

Correlation analysis is to analyze the degree of correlation between two variables. Since not all indicator data meet normal distribution, Spearman correlation coefficient is used to analyze the correlation among indicators. Spearman correlation coefficient is defined as Pearson correlation coefficient between rank variables[4]. For the sample size of n, n original data are converted into grade data, and the correlation coefficient ρ is:

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \tag{8}$$

In practice, the connection between variables is insignificant, so ρ can be calculated in a simple step. The difference of the grades of the two variables being observed, ρ is:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)} \tag{9}$$

Spearman analysis results are shown in Table 6 (only part of them are shown due to space limitation), including correlation coefficient and significance P value:

Table 6 Phase relation table

	F1	F2	……	F10	F11
F1	1(0.000***)	-0.606(0.000***)	……	-0.27(0.000***)	-0.792(0.000***)
F2	-0.606(0.000***)	1(0.000***)	……	0.224(0.001***)	0.626(0.000***)
……	……	……	……	……	……
F10	-0.27(0.000***)	0.224(0.001***)	……	1(0.000***)	0.297(0.000***)
F11	-0.792(0.000***)	0.626(0.000***)	……	0.297(0.000***)	1(0.000***)
Note: ***, ** and * represent the significance level of 1%, 5% and 10% respectively					

Fig. 3 Heat map of correlation analysis

Due to space limitation, only part of the data is listed in Table 6, and the specific data has been shown in Fig. 3. Through the table and figure, it can be found that population-related indicators show significant correlation with financial, economic poverty and income, education, health, energy production and use, investment environment, trade, government fiscal and monetary stability. Meanwhile, the correlation degree with finance and health is the largest, while for finance, it shows non-significant correlation with investment environment and monetary stability. There is also a negative correlation between poverty and income, health, trade and currency stability. More correlation among indicators can be seen directly from the table. On the whole, there is a high degree of correlation among these indicators.

5.2.3 Causal analysis

Hierarchical regression is an effective method to solve causality analysis. In essence, it is built on the basis of regression analysis. The difference lies in how hierarchical regression is divided into multiple layers, which is used to study the differences between two or more regression models[5]. If the variable still has a significant contribution, then it can be concluded that the variable does have a unique role that cannot be replaced by other variables. This method is mainly used when there is a high correlation between the independent variables and the unique contribution of one of the independent variables is difficult to determine.

Table 7 Regression of financial and demographic stratification

Hierarchical regression								
	Control layer				Layer 1			
	B	Standard deviation	t	P	B	Standard deviation	t	P
Constant	0.436	0.06	7.2	0.000***	0.446	0.059	7.538	0.000***
F2					-0.275	0.082	-3.333	0.001***
F3	0.135	0.072	1.892	0.060*	0.185	0.071	2.593	0.010**
F4	0.029	0.119	0.245	0.807	0.071	0.117	0.611	0.542
F5	-0.124	0.073	-1.708	0.089*	-0.133	0.071	-1.871	0.063*
F6	-0.19	0.078	-2.442	0.015**	-0.14	0.077	-1.81	0.072*
F7	0.362	0.053	6.904	0.000***	0.329	0.052	6.308	0.000***
F8	-0.171	0.119	-1.443	0.15	-0.181	0.116	-1.558	0.121
F9	0.075	0.044	1.698	0.091*	0.063	0.044	1.438	0.152
F10	-0.064	0.084	-0.762	0.447	-0.042	0.082	-0.514	0.608
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			
△F value	F(9, 215) =23.04，P=0.000***				F(1, 214) =11.106，P=0.000***			
Dependent variable (Y): population								
Note: ***, **, * represent 1%, 5%, 10% significance levels, respectively								

Table 7 shows the test results of this model and the parameters of the model, including the coefficients and P values of the model, which can be used to analyze the formula of the model.

The F test is used to determine whether the model is meaningful ($P<0.05$, indicating that the model is meaningful), and the F values wants to calculate to get the P value, which needs to provide two values of degrees of freedom df1 and df2. In general, df1 equals the number of independent variables; df2 equals the sample size (number of independent variables + 1). R^2 represents the degree of fit of the model, and the closer to 1 the better the effect (for example, R^2 is 0.045, which means that all X can explain 4.5% of the variation in Y).

Layer 1, including fields constants, economy, poverty and income, education, health, energy production and use, investment environment, trade government finance, monetary stability, and finance, have a significance P value of 0.00***, which presents significance at the level, rejecting the original hypothesis, therefore the model is valid, while the model's goodness of fit R^2 is 0.529, and the model performs well.

Thus we can see that with population as the independent variable and finance as the dependent variable, controlling for economic poverty and income education health, energy production and use, investment climate, trade, government finance, and money as stable variables, we can find that finance presents a significant level in the hierarchical 1 model, which indicates that there is a significant correlation between finance and population after excluding the interference of other variables, i.e., population is the cause of finance and finance is the effect of population. Because of the complicated process, limited by space, we do not analyze them one by one here, and Table 8 shows the results after conducting two t levels of regression.

Table 8 Results of causal analysis

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11
F1	1.0	0.001***	0.001***	0.063*	0.072*	0.000***	0.1	0.010**	0.2	0.6	0.5
F2		1.0	0.000***	0.354	0.023**	0.3	0.5	0.000***	0.4	0.3	0.099*
F3			1.0	0.2	0.200	0.6	0.000***	0.4	0.066*	0.163	0.090*
F4				1.0	0.305	0.7	0.429	0.015**	0.000***	0.6	0.5
F5					1.0	0.8	0.2	0.4	0.6	0.149	0.706
F6						1.0	1.0	0.9	0.6	0.937	0.098*
F7							1.0	0.003***	0.2	0.7	0.7
F8								1.0	0.057*	0.3	0.001***
F9									1.0	0.5	0.142
F10										1.0	0.3
F11											1.0

The above table shows that population and energy production and use, trade, and monetary stability of government finance do not show significant correlation after removing the disturbance. Finance and poverty are not significantly correlated with income, health, energy production and use, trade, and government finance. Finance and poverty are not significantly correlated with income education health investment environment, and government finance. Due to space limitations and the fact that the table clearly shows the relationship between the indicators, it is not repeated here.

In summary, it can be found that after removing the interference of variables, causal relationships can exist between many variables, further - proving that the selected indicators have a synergistic checking balance effect.

6 Modeling and solving of Problem 2

6.1 TOPSIS social stability binning based on entropy power method

TOPSIS method is a commonly used intra-group comprehensive evaluation method, which can make full use of the information of the original data, and its results can accurately reflect the gap between the evaluation schemes. The combination of entropy weight method and TOPSIS can evaluate the weight of indicators, eliminate the influence of default the same weight of indicators on the evaluation results, and make the evaluation results more reliable. The specific steps to establish TOPSIS model based on entropy weight method are as follows:

(1) Determine whether the input matrix has negative numbers

The maximum number of evaluation objects is represented by n, and the positive matrix formed by m evaluation indexes is as follows[6,7]:

$$X=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \tag{10}$$

Then, its normalized matrix is denoted as Z, and each element in Z:

$$z_{ij}=x_{ij}\Big/\sqrt{\sum_{i=1}^{n}x_{ij}^2} \tag{11}$$

It is necessary to determine whether there is a negative number in Z matrix. If so, another normalization method should be used for X to standardize matrix X once to obtain matrix. The standardized formula is as follows [8] :

$$\tilde{z}_{ij}=\frac{x_{ij}-\min\left\{x_{1j},x_{2j},\cdots,x_{nj}\right\}}{\max\left\{x_{1j},x_{2j},\cdots,x_{nj}\right\}-\min\left\{x_{1j},x_{2j},\cdots,x_{nj}\right\}} \tag{12}$$

(2) Calculate the proportion of the i-th sample of the j-th index

Suppose there are n objects to be evaluated and m evaluation indexes, and the non-negative matrix obtained after processing in the previous step is:

$$\tilde{Z}=\begin{bmatrix} \tilde{z}_{11} & \tilde{z}_{12} & \cdots & \tilde{z}_{1m} \\ \tilde{z}_{21} & \tilde{z}_{22} & \cdots & \tilde{z}_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{z}_{n1} & \tilde{z}_{n2} & \cdots & \tilde{z}_{nm} \end{bmatrix} \tag{13}$$

Calculate the probability matrix P, where the formula for each element P_{ij} in P is as follows:

$$p_{ij}=\frac{\tilde{z}_{ij}}{\sum_{i=1}^{n}\tilde{z}_{ij}} \tag{14}$$

Easy to verify: $\sum_{i=1}^{n}P_{ij}=1$, that is, the corresponding probability sum of each indicator is guaranteed to be 1.

(3) Calculate the information entropy and normalize the entropy weight

For the *j*-th index, its information entropy calculation formula is:

$$e_j = -\frac{1}{Inn}\sum_{i=1}^{n} P_{ij} In(P_{ij})(j=1,2\cdots,m) \tag{15}$$

(4) The score is calculated and normalized

The above steps of entropy weight method have completed the original matrix positive normalization and the normalization of the matrix, so it is no longer processed here, but only used TOPSIS to calculate the score and normalization.

Define *i* (i=1,2... ,n) distance between evaluation objects and the maximum value：

$$D_i^+ = \sqrt{\sum_{j=1}^{m} \omega_j \left(Z_j^+ - z_{ij}\right)^2} \tag{16}$$

Define *i* (i=1,2... ,n) distance between evaluation objects and the minimum value:

$$D_i^- = \sqrt{\sum_{j=1}^{m} \omega_j \left(Z_j^- - z_{ij}\right)^2} \tag{17}$$

Then, the ith (i=1,2... ,n) unnormalized scores of evaluation objects can be calculated:

$$S_i = \frac{D_i^-}{D_i^+ + D_i^-} \tag{18}$$

Obviously, $0 \leqslant S_i \leqslant 1$, and the larger S_i is, the smaller D_i^+ is, that is, approximately close to the maximum value.

Table 9 shows the weight calculation results of the entropy weight method. According to the results, the weight of the reserve index is analyzed.

Table 9 Weight calculation results of entropy weight method

Item	Information entropy value	Information utility valued	Weight (%)
F1	0.973	0.022	4.40
F2	0.978	0.027	9.17
F3	0.993	0.077	19.09
F4	0.886	0.005	2.10
F5	0.944	0.009	2.80
F6	0.991	0.056	18.57
F7	0.985	0.038	11.49
F8	0.962	0.015	3.85
F9	0.976	0.024	5.53
F10	0.923	0.007	2.47
F11	0.995	0.114	20.54

Table 10 shows the comprehensive score calculated by entropy weight TOPSIS method. The values of D+ and D- respectively represent the distance (Euclidean distance) between the evaluation object and the optimal or the worst solution (i.e. A+ or A-). The practical significance of these two values is that the distance between the evaluation object and the optimal or the worst solution, the larger the value, the farther the distance; the larger the value of D+ of the research object, the farther the distance from the optimal solution; The greater the D- value, the further away from the worst solution. The ideal object of study is the smaller the D+ value and the larger the D- value. The greater the relative value of D-, the further the research object is from the worst solution, the better the research object is. The higher the C value, the better the study object.

Table 10 Comprehensive score

Country	Positive ideal solution distance (D+)	Negative ideal solution distance (D-)	Composite score index	Rank
Algeria	0.912	0.158	0.148	195
Angola	0.852	0.276	0.245	39
Botswana	0.864	0.222	0.204	77
……	……	……	……	……
Former southern Macedonia	0.860	0.189	0.181	127

In order to facilitate the model solution of the social early warning model later, the comprehensive score is divided into three grades by the three-digit number. The higher the value, the higher the comprehensive evaluation score of social stability. Due to space limitation, only part of the data is shown, as shown in Table 11:

Table 11 Social stability grades

Country	Population	Finance	...	Composite score index	File classification
Central Africa	1.00	0.04	...	0.36	3
Chile	0.24	0.32	...	0.20	2
Gibraltar	0.32	0.20	...	0.18	2
Chad	0.56	0.03	...	0.27	3
Vietnam	0.29	0.45	...	0.17	2
The West Bank and Gaza	0.38	0.03	...	0.14	1
Jordan	0.23	0.28	...	0.16	1

6.2 The warning model of social stability based on PSO-C4.5 decision tree

6.2.1 Establishment of early warning model of social stability

(1) Define the ability of indicators to classify models

In the process of model building, it is necessary to constantly adjust the training results of the model so that it gradually approaches the real situation. Therefore, some indicators are needed to measure the execution ability of the model[8]. The two most basic indicators in the field of data mining classification are Recall Rate and accuracy rate.

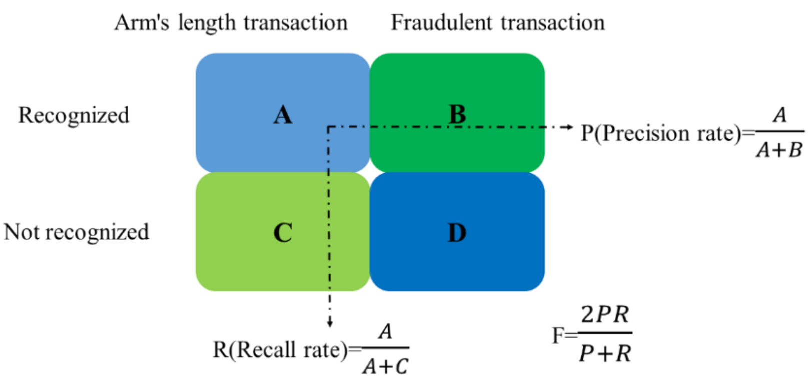

Fig. 4 Relationship between accuracy and recall rate

Accuracy and recall rate are mutually affecting, ideally, it is sure to achieve both high, but in general, high accuracy, low recall rate, on the contrary, low recall rate, high accuracy. In order to weigh the influence of the two, F-Measure is introduced to synthesize the model. F-Measur 'e is a weighted harmonic average of Precision and Recall:

When parameter $\alpha=1$:

$$F=\frac{2*\mathrm{P}*\mathrm{R}}{\mathrm{P}+\mathrm{R}} \tag{19}$$

Among the criteria for evaluating the classification effect, misjudgment rate *M* is also an important indicator, which can be expressed as follows:

$$M=\frac{FP+FN}{\text{Allcount}} \tag{20}$$

Where *FP* is the count that is actually 0 when the prediction is 1, *FN* is the count that is actually 1 when the prediction is 0, and *Allcount* is the total sample number.

(2) Establishment of decision tree model

Decision tree is a working tree-like structure. It starts from the root node, tests the data samples, divides the data samples into different data sample subsets according to different results, and each data sample subset constitutes a sub-node. The purpose of constructing decision tree is to find out the relationship between attributes and categories and use it to predict the future unknown categories. Gain rate formula:

$$\text{GainRatio}(D|\ A)=\frac{\text{infoGain}(D|\ A)}{\text{IV}(A)} \tag{21}$$

$$IV(A)=-\sum_{n=1}^{N}\frac{|D_n|}{|D|}*\log_2\frac{|D_n|}{|D|} \tag{22}$$

Where $A=[a_1,a_2,\ldots,a_n]$ has n values. If *A* is used to divide the sample set *D*, *n* branch nodes will be generated.

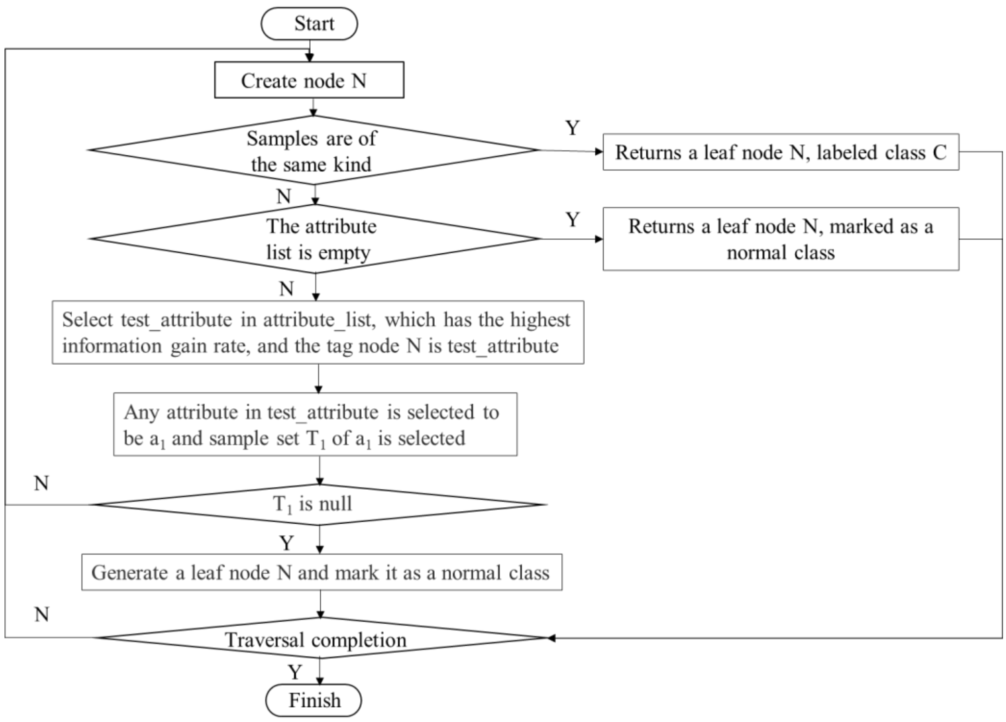

Fig. 5 Decision-making process

6.2.2 PSO hyperparameter model optimization

Hyperparameters refer to parameters that cannot be learned from data in the training process. Usually, there are a large number of hyperparameters in decision tree models. The setting of super parameters will directly affect the performance and performance of the model[9]. When determining model hyperparameters, search is usually carried out according to Fig. 6

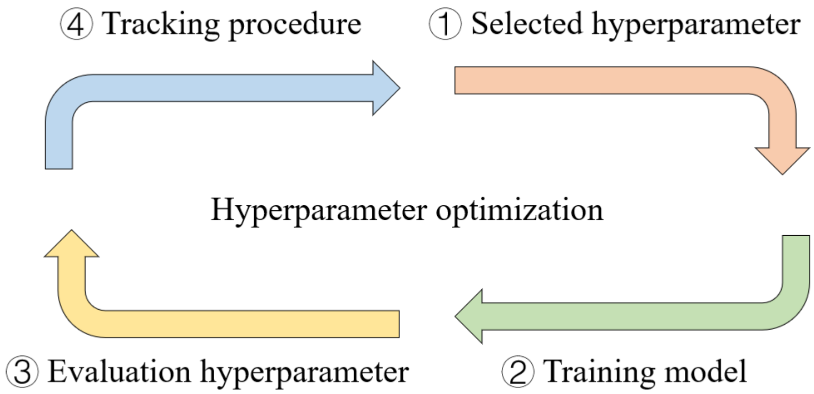

Fig. 6 Hyperparameter optimization

The efficient hyperparameter optimization method is of great practical significance for the improvement of model effect. The PSO algorithm is selected for hyperparameter optimization. The standard PSO algorithm is a global optimization algorithm, which integrates the principles of "swarm" and "optimization", and optimizes by the fit value of the particle. PSO algorithm retains the global search strategy based on population and regards each individual as a particle without weight or volume in the N-dimensional search space. In each iteration, each particle adjusts its flight speed and position according to the following formula.

$$v_{ij}(t+1) = wv_{ij}(t) + c_1 r_{1j}\left[p_{vj}(t) - x_{ij}(t)\right] + c_2 r_{2j}\left[p_{gj}(t) - x_{ij}(t)\right] \tag{23}$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1) \tag{24}$$

Where, j represents the j-th dimension of the particle; i is the i-th particle; t is the t generation; C_1 and C_2 are acceleration constants; $r_1 \sim U(0,1)$ and $r_2 \sim U(0,1)$ are two independent random functions. The tuning process of PSO algorithm is shown in Fig. 7.

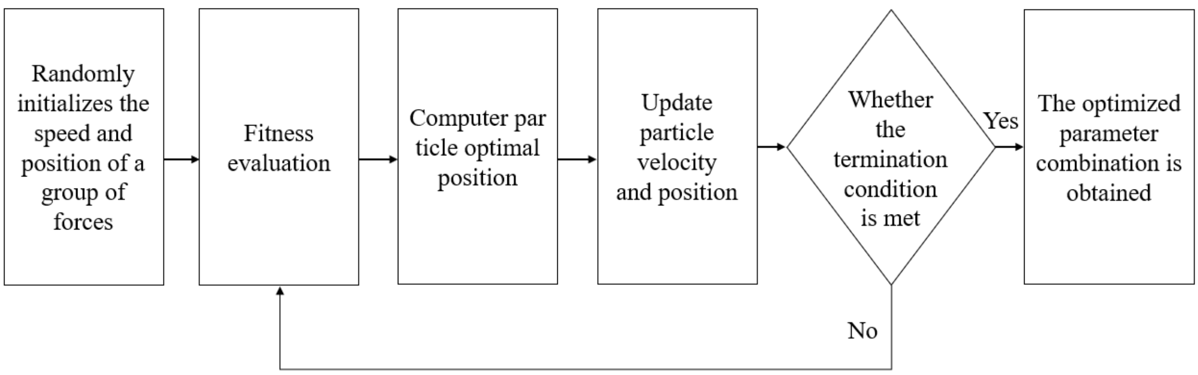

Fig.7 PSO algorithm tuning process

6.2.3 Model solving

The comprehensive score grade was taken as Y, and the data was divided into training set and test set, in which the test set accounted for 20%. Table 12 shows the model performance parameters calculated based on the prediction results of the classification model, and the overall effect of the model is excellent.

Table 12 Performance parameters of the model

Category	precision	recall	fl-score
1	0.43	1	0.6
2	1	0.78	0.88
3	0.51	1	0.67
weighted avg	0.91	0.81	0.84

Table 13 is the parameter of the PSO-C4.5 decision tree model. The optimal solution can be obtained by setting the classification model with this parameter.

Table 13 Parameters of PSO-C4.5 decision tree model

Parameter	Optimum value	Parameter	Optimum value
max-depth	5	min-impurity-decrease	0.1
min-samples-split	2	ccp-alpha	0.2
min-samples-leaf	1	leaning-rate	0.14
min-weight-fraction-leaf	1	max-leaf-nodes	0

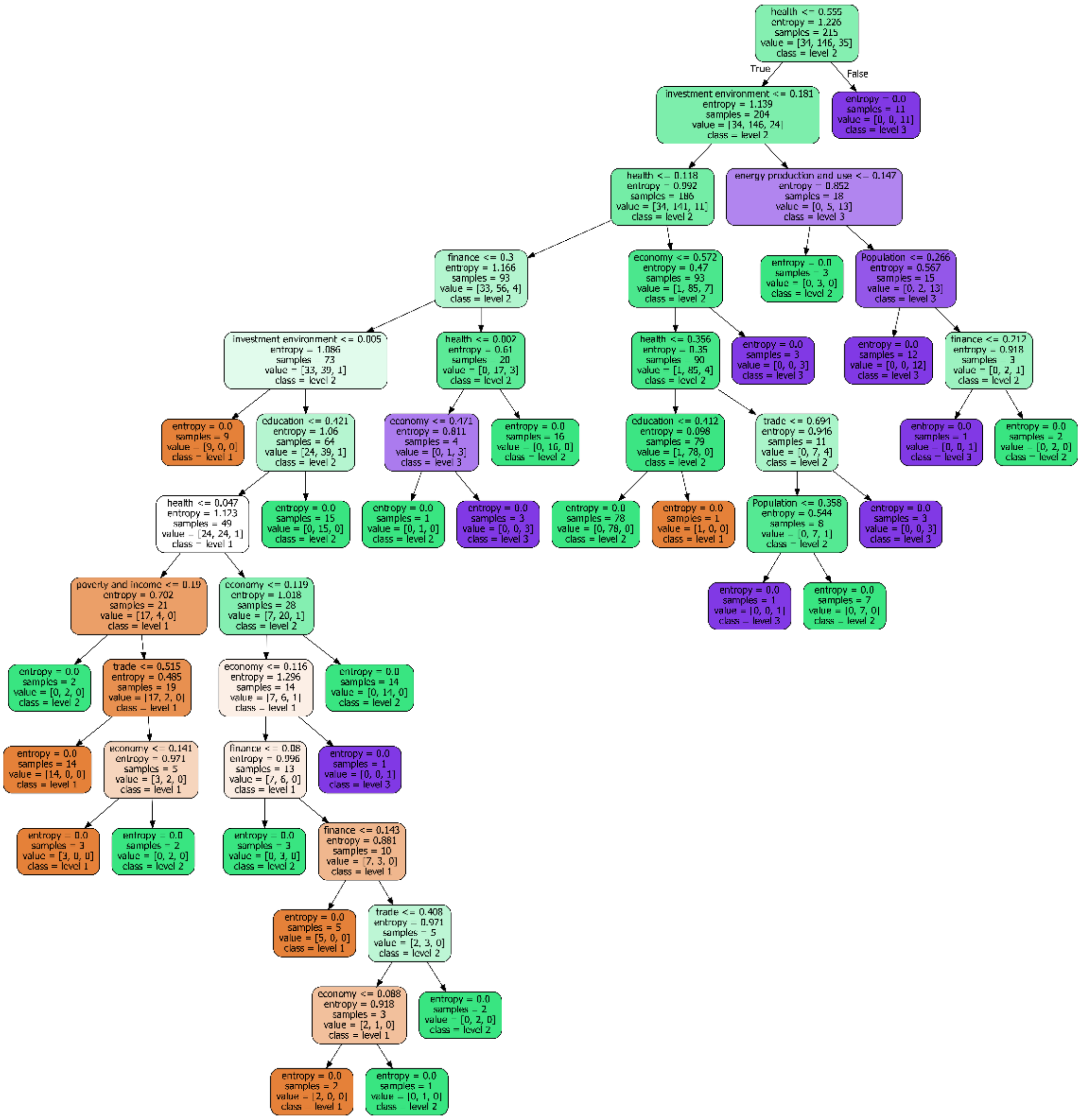

Fig. 8 Decision tree partitioning path

Fig. 8 shows the two-node division path of the decision tree model. Based on the indicator threshold, the early warning value of the social stability indicator of each country can be obtained.

7 Modeling and solving of Problem 3

7.1 Assessment and prediction of Social stability in Cambodia

Cambodia is a major Southeast Asian country that maintains good international relations with China. Cambodia had a color revolution, which ended in failure, and its supported opposition leaders such as Gen Sok Ha were caught and sentenced for treason. Based on the PSO-C4.5 decision model, the SHAP model was used to analyze Cambodian data, assess its social stability and make predictions.

The SHAP model is an emerging method for learning explanatory machine models, as shown in Fig. 9, which infers the model's decisions by computing the effective value of each feature to produce meaningful explanatory results. The method is able to generate a graph based on the input data provided by the user, through which the behavior of the model is graphically displayed[10]. The advantage of SHAP over other explanatory model methods is that it can provide the effective contribution to each feature and the decision process of the explanatory model.

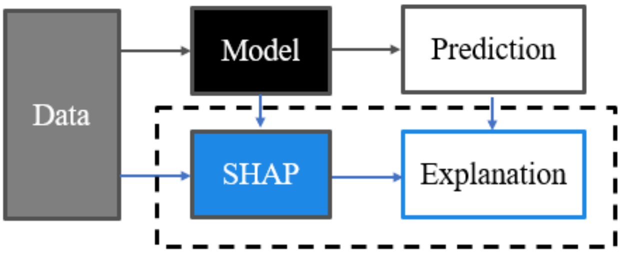

Fig. 9 SHAP method framework

The calculation formula of SHAP is as follows:

$$\phi_j(val) = \sum_{S \subseteq \{x_1,\ldots x_p\} \backslash \{x_j\}} \frac{|S|!(p-|S|-1)!}{p!} \left(\mathrm{val}\left(S \cup \{x_j\}\right) - \mathrm{val}(S) \right) \tag{25}$$

Where S is the subset of features used in the model, x is the vector of eigenvalues of samples to be interpreted, p is the number of features, and val(S) refers to the model output value under the feature combination S.

Through SHAP model, combined with the above PSO-C4.5 decision tree model, results as shown in Fig. 10 and Fig. 11 are obtained, with red representing positive influence and blue representing negative influence. In Cambodia, there are five main reasons for the color revolution: investment climate, economic, hygiene, government finance and education. Poor hygiene, backward education and poor government led to the outbreak of color revolutions. Although Cambodia broke out the color revolution due to the above reasons, the overall stability remained stable and the social stability was at the second level. The color revolution failed.

Fig. 10 Forth plot of Cambodia during the Color Revolution

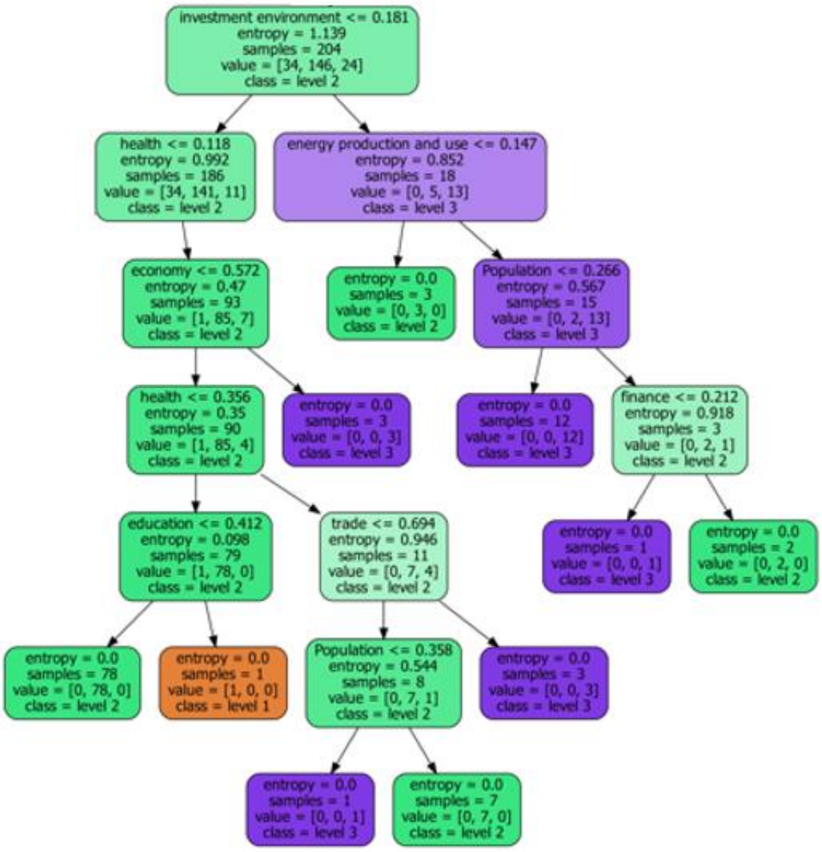

Fig. 11 Decision Tree evaluation of social stability in Cambodia

The social stability of Cambodia in the future is predicted, and the results are shown in Fig. 12. It can be seen that in the future, the economic situation of Cambodia is good, but the sanitary conditions are still not good enough, and the investment environment is crowded to some extent. At present, Cambodia's GDP growth rate has exceeded 7% for seven consecutive years, which is the fastest in the world. Health needs to be improved gradually. As the economic situation improves and investment increases, the investment environment is bound to be affected. Compared with the actual situation and the forecast situation, they tend to be consistent, which reflects the accuracy of the forecast from the side.

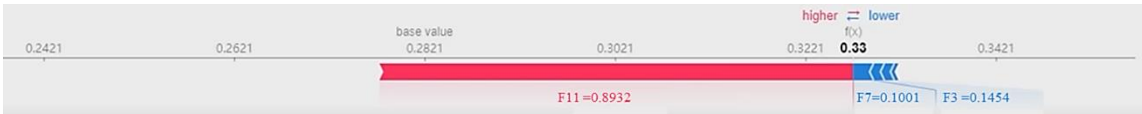

Fig. 12 Future Forth plot of Cambodia

7.2 Analysis and suggestion of failure causes of color revolution

Based on the evaluation and prediction results in 7.1, the reasons for the failure of Cambodia's color revolution are as follows:

(1) Cambodia has a good investment environment, which is a huge advantage of its economic growth, and has prevented the success of the color revolution. Import in Cambodia does not require a license, and the Ministry of Commerce is responsible for export approval procedures. Some products need to obtain the relevant government departments special export authorization or permission before they can be exported. Out of the

(2) Cambodia has a good demographic dividend, resulting in fewer problems of unemployment, poverty and inequality. The average age of Cambodia's population is 25, with young people under 30 accounting for 70% of the population. Cambodia's entire population structure presents a perfect "straight pyramid shape", with high demand and strong demographic dividend. The population is increasing by 4% every year. About 50 percent of that population is under the age of 25.

(3) Cambodia's poor health conditions, relatively backward education, lack of government finance, and the influence of other countries led to the outbreak of color revolution. However, under the favorable investment environment, high-quality demographic dividend, and effective government repression, its national stability was on the rise.

In order to further improve the stability of the country, suggestions on how to avoid color revolution in Cambodia are put forward as follows:

(1) Increase government finance and improve the government's governance and credibility. Although Cambodia has a good investment environment, it is based on the self-sacrifice of the government. A balance should be made between the investment environment and government finance to stabilize the government's rule.

(2) Improve health standards. The health level has a great impact on people's life quality. Although the health level in Cambodia has been improved to a certain extent, it still has deficiencies and needs to be further improved.

(3) Strengthen education. Education is the foundation of a country. Although its direct influence is small, its indirect influence is huge, and it is the most complicated evaluation index.

8 Modeling and solving of Problem 4

8.1 Assessment of social stability in Myanmar

Color revolution broke out in Myanmar and led to regime change. The established PSO-C4.5 decision-making model was used to evaluate the stability of Myanmar society, and the results are shown in Fig. 13 and Fig. 14.

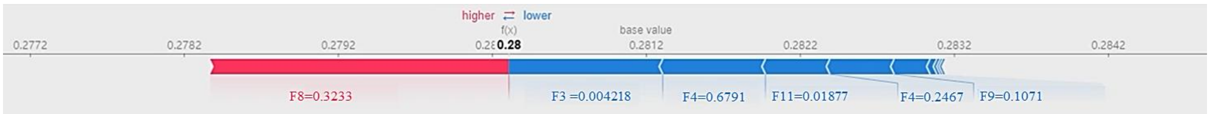

Fig. 13 The Forth plot of Burma during the Color Revolution

According to Fig. 13, although Myanmar has a good amount of resources, its poor investment environment, poor monetary stability and poor economy eventually lead to the outbreak of color revolution and the change of regime. According to the analysis of Fig. 14, it can be seen that the poor investment environment is not the direct cause of regime change, but the poor investment environment leads to the economic downturn, the low level of education and poor health of the people, and the poor trade of Myanmar. A series of factors lead to regime change step by step. The model established in this paper can not only evaluate social stability, but also directly reflect the causes and logical relations leading to social instability and regime change. It also reflects the relevance and logic among various indicators and shows the rationality of the selected indicators.

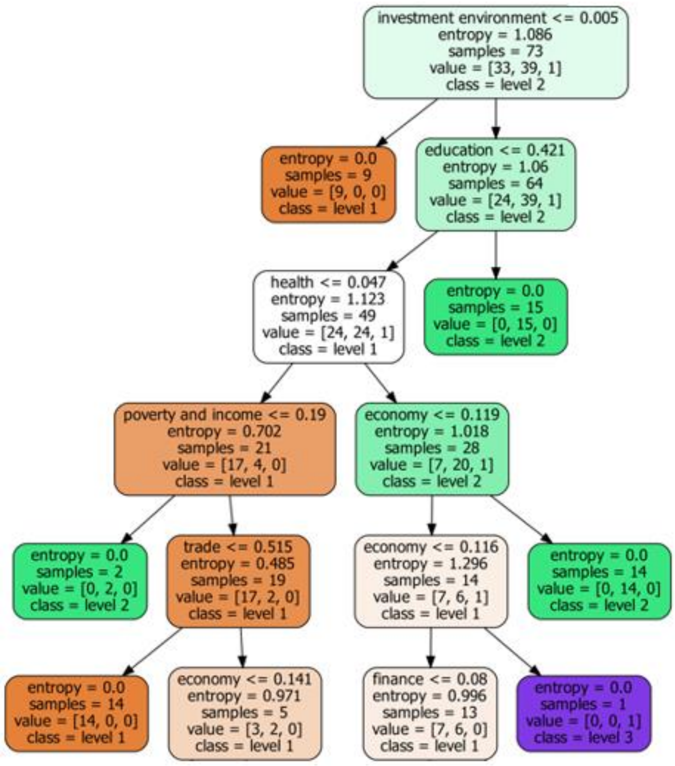

Fig. 14 Social stability decision tree evaluation in Myanmar

8.2 Analysis of reasons for regime change

On the basis of the above analysis and evaluation, the reasons for regime change in Myanmar are analyzed:

(1) Although Myanmar is rich in domestic resources, it is difficult for the former government to maintain its power due to a series of blows such as poor investment environment, poor currency stability and low education level.

(2) The domestic situation in Myanmar has been unstable, leading to huge investment risks. Even though Myanmar has abundant labor force, it is difficult to play its role. Instead, it is used by other countries' capital, further crowding out the domestic industrial environment.

(3) General public dissatisfaction. The widespread unpopularity of the military regime and widespread public anger and frustration with government policies and practices provided a solid foundation for the color revolution to gain momentum.

(4) The peaceful nature of the protests. The largely peaceful protests in Myanmar, where protesters have been able to maintain discipline and avoid violent confrontations with the military, have helped maintain public support and increase their chances of success,

(5) Erosion of the military regime's power: Over time, as key Allies defected and the economy struggled, the military regime's grip on power weakened. This created an opportunity for the success of the color revolution and the transfer of power to the army

All in all, popular discontent with the government due to risky domestic investments, poor currency stability and low education levels, combined with orderly opposition groups, eventually led to regime change in Myanmar.

9 Modeling and solving of Problem 5

On the basis of solving the above problems, further analysis is carried out. Based on the model, the importance of various indicators can be seen. As shown in Fig. 15, the main indicators affecting social stability are economy, investment environment, finance, government finance and education. Although the contribution of other indicators is close to 0, it does not mean that they have no contribution to social stability. Based on the previous analysis, it can be seen that social stability indicators are correlated with each other and there is a causal relationship.

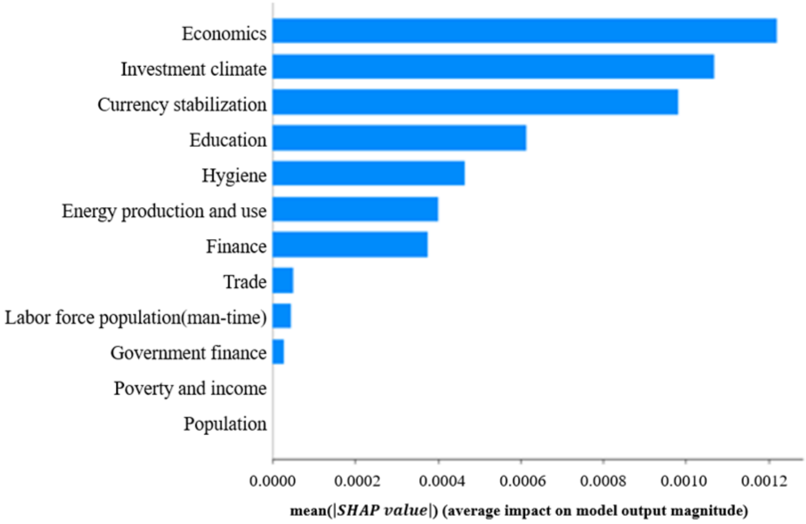

Fig. 15 Importance of each indicator

Fig. 16 shows the difference in results caused by the increase or decrease of each index value. It can be seen that the economic and investment environment is more complicated. Some samples also have the effect of reducing social stability when the economic level is low, especially the investment environment basically combines red and blue. The same is true of the following variables. This means, therefore, that the stability of a country's society cannot be evaluated only from the temporary value of some indicators on the surface.

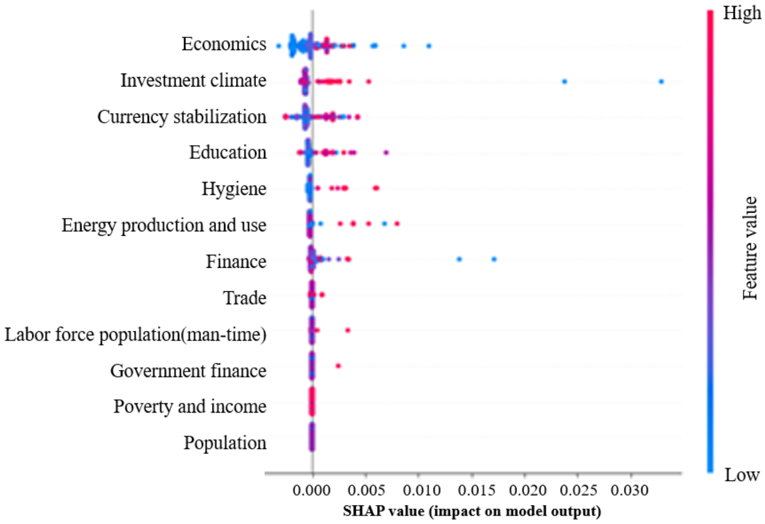

Fig. 16 Influence of indicator changes on social stability

Combined with the establishment, solution and analysis of the model in this paper, the potential social and political factors leading to color revolution are considered based on the complex dispute forms of the influence of social stability indicators on social stability. In order to prevent color revolution and maintain social stability, the following suggestions are put forward:

(1) Promoting economic growth and development: The Government should focus on promoting economic growth and development, striving to create employment opportunities, raise income levels, and reduce poverty and inequality.

(2) Promote education and civic participation: Governments should promote education and civic participation, encourage citizens to participate in public life, and develop the skills and knowledge needed to participate actively in political and social debate.

(3) Address social and economic inequalities: Governments should address social and economic inequalities, ensure access to opportunities and resources for all citizens, and work to reduce poverty and inequality.

(4) Improve transparency and accountability: Governments should strive to increase transparency and accountability, for example by providing regular and open access to information, strengthening the rule of law, and ensuring that those in power are held accountable for their actions.

(5) Address public grievances: The government should take measures to address public grievances and concerns, such as corruption, lack of political freedom, and economic inequality.

(6) Promote freedom of expression and the press: Governments should protect and promote freedom of expression and the press and allow open and free debate and discussion on political and social issues.

(7) Strengthen the rule of law: Governments should strengthen the rule of law to ensure that laws and regulations are observed and enforced fairly and equally, and that all citizens have equal access to justice.

(8) Promote inclusive governance: Governments should strive to promote inclusive governance by engaging with diverse groups and communities and ensuring that all voices are heard and represented in decision-making.

10 Model Evaluation and Further Discussion

10.1 Strengths

(1) In the selection of indicators, the selection principle is fully considered, and the selected indicators are representative. All the selected data are from authoritative websites, which ensures the accuracy of the data. On this basis, the model is analyzed and established. The accuracy of the model is high, and the results obtained are reliable.

(2) Complete data collection was conducted, using data from 215 countries worldwide, not just one country. Analysis and classification of data from 215 countries gives a more compelling ranking of social stability.

(3) The TOPSIS evaluation model based on entropy weight method is established. Compared with the general TOPSIS model, the contribution of different indicators to the weight of

evaluation results is considered, and the weight of evaluation indicators is given to eliminate the influence of the default weight of all indicators on the evaluation results, making the results more reliable.

(4) The early warning model of social stability was established based on the PSO-C4.5 decision tree model, and the SHAP model was used for analysis. The correlation and logic among indicators were fully considered, and the evaluation results were more reasonable.

(5) In the selection of specific countries for analysis, countries in Southeast Asia are selected. The resource background of the two countries is close, and the analysis results are comparable, which makes this paper more realistic.

10.2 Weaknesses

(1) Due to limited time, although some indicators are selected as much as possible, which can represent social stability (representativeness is above 95%), more indicators will be better representativeness.

(2) Due to space limitations, part of the process is not shown, and only part of the data is shown, but key formulas and results have been given.

References

[1] Chen Qiuling, Xiao Lu, Cao Qingjin. Design of Social Early Warning Indicator System and Definition of Early Warning Judgment Criteria--Based on Social Stability Perspective[J]. High-level Forum on Public Management,2008,No.7(01):62-76.

[2] Zhong Ning. Research on social stability and its assessment model under the guidance of scientific development concept [D]. Northeast Normal University, 2014.

[3] Mo Yun, Guo Yan, Mo Wo Sheng, Lu Zhongwei, Zhang Shaorong. A hybrid feature selection method based on feature filtering and PCA dimensionality reduction[J]. Journal of Guilin Institute of Aerospace Technology,2022,27(02):145-151.

[4] Jia Ke, Yang Zhe, Wei Chao, Zheng Liming, Li Yanbin, Bi Tianshu. Longitudinal protection of new energy transmission lines based on Spearman's rank correlation coefficient[J]. Power System Automation,2020,44(15):103-111.

[5]Yu Miao,Ma Yumeng. Hierarchical regression analysis of the impact of knowledge sharing in the relationship between innovation climate and team creativity[J]. China Health Statistics,2018,35(05):752-753+757.

[6] Gao Weicheng, Wang Xiaoping, Xiao Xiaolin, Shi Yiwei, Lin Shaofen. Comprehensive evaluation of the quality of loquat leaves from different origins based on entropy weight TOPSIS method[J]. Chinese Pharmacy,2023,34(01):72-77.

[7] Xu Hao, Tian Zhenyu, Yang Jianlan. Design and research of hospital outpatient queuing based on entropy power TOPSIS[J]. Chinese Journal of Health Information Management,2022,19(06):860-866.

[8] Liu Riling. performance analysis of C4.5 decision tree classification algorithm [J]. Information Systems Engineering,2019(01):153-154.

[9]Wang Qian, Zhao J, Wang Yousan. Tailings dam deformation prediction based on PSO-BP combined model[J]. Geospatial Information,2022,20(11):119-122+128.

[10] Song Dandan, Wang Hui, Wen Gang. Credit risk assessment based on SHAP interpretation method[J]. Science and Technology Economic Market,2022(10):58-60.