Problem Chosen	2023HSB MCM/ICM Summary Sheet	Team Control Number
B		MI00302

Research on Social Stability Risk Assessment based on Decision Trees

Summary

"Thinking of danger in times of peace", measuring and warning social stability, has always been highly valued by far-sighted rulers and politicians. Social unrest or collapse through peaceful, nonviolent regime change campaigns. The purpose of this report is to build a social stability risk assessment model to evaluate the reasons for the failure of color revolutions, hoping to make a scientific analysis in terms of the laws of social crisis early warning and provide some suggestions to avoid color revolutions.

For task 1, Construct social stability indicator system. First, the data of social stability indicators from 1980 to 2020 were collected to qualitatively analyze the causal relationship of social stability, and the composition relationship of social stability indicators is shown in Figure 2. Then, the collected quantitative data were data cleaned and the six indicators for qualitative analysis were assigned by AHP. The importance indicators were selected for cluster analysis, and the preliminary index system is shown in Figure 5. Next, correlations between social stability factors were analyzed using correlation coefficients. The correlation heat map is shown in Figure 7. Finally, Kendall's W test was used to test the correlation of social stability factors, and the results are shown in Table 6, which indicates that the significance value of the overall data is 0.000, the degree of correlation is a high degree of consistency.

For task 2, Based on the social indicator system of Task 1, a social stability early warning evaluation model was established. A decision tree police level classification model is used to estimate the parameters, and a comprehensive social stability evaluation model determines the thresholds of the parameters. The threshold value is $V=[0,0.15,0.35,0.5,1]^T$, The determination of the social stability warning level is shown in Table 8.3.

For task 3, Identify the reasons for the failure of the color revolution. First, consider Myanmar. Then, the features were selected based on the decision tree police-level classification model for the causes of failure, and the results are shown in Figure 8, This study shows that the natural growth rate and the characteristic degree of trade export share over 50% contribute to the resistance to the color revolution in Myanmar. Next, the trends of Myanmar's future development over 50 years are estimated based on the Social Stability Early Warning Assessment Model, and the before-and-after comparison is shown in Figure 9, The results show that the giant alert level disappears, the warning alert weakens to about 8%, the no-alert level exceeds 75%, and the society moves toward harmonization. Finally, suggestions were made for the future development of Myanmar.

For task 4, Identification of reasons for regime change. Georgia, Ukraine and Belarus are selected for this paper. Then, the causes of regime change are analyzed based on the social stability early warning assessment model, and the comparison of the causes of regime change in the three countries is shown in Figure 11, Both the natural growth rate and trade exports have an effect on regime change of more than 10%, while the other indicators have an effect close to zero and have a moderating effect on the regime change process.

For Task 5, Suggests 12 recommendations to avoid color revolutions, including three parts: limit, encourage, and monitor.

Finally, The sensitivity analysis shows that the decision tree model is more stable when the splitting criterion is Deviation and the maximum number of splits is 80. Introduction of geometric and arithmetic solution weights to AHP, and it was found that the relative error of the weights did not exceed 5% for $CI=0.0125$, which further improved the robustness of the model.

Keywords: Social Stability ; AHP; Cluster Analysis ; Correlation Coefficient; Decision Trees

Contents

1 Introduction

1.1 Problem Background

From the 1990s to the present, due to the deepening of globalization and the increasingly close economic and political relations among countries, it is not easy to overcome crises in a modern society with increasingly fierce competition and frequent crises. Once the power ratio of the acting factors is out of balance, the society will be deformed and unstable, which will lead to social unrest or collapse [1]. The peaceful and non-violent regime change movements named after colors that occurred in many countries and regions at the end of the last century and the beginning of this century all achieved their goals by attacking the weakest link that affects social stability. To avoid the occurrence of color revolutions, rely on modern With social measurement science and modern social early warning science, accurate early warning of social crisis has become particularly important.

1.2 Restatement of the Problem

Considering the background information and constraints identified in the problem statement, the following questions need to be addressed:

- Task 1 , select representative social stability indicators, establish an indicator system that may affect social stability from both qualitative and quantitative perspectives , and discuss the correlation and causal relationship between them;
- Task 2 , develop an early warning model of social stability and discuss it;
- Task 3 , select a country or region where a color revolution attempts to overthrow the regime , and use the established social stability early warning model to evaluate its social stability. Compare and demonstrate its evaluation indicators, and judge which indicators are the reasons for the failure of the color revolution;
- Task 4 , select the country or region where the color revolution led to regime change, and use the established social stability early warning model to find out the main reasons for regime change;
- Task 5, put forward relevant suggestions to avoid color revolutions and maintain social stability.

1.3 Literature Review

The construction of the social early warning system and its development are based on the measurement and grasp of the system of social early warning indicators. The research and application of social early warning indicator systems originated in the United States in the 1960s and 1970s, and after the American sociologist R. A. Bauer published Social Indicators (1966), a wave of research on social development indicators was formed worldwide, especially in developed capitalist countries and countries with rapid economic development [2].

After the 1980s, scholars in China began to pay extensive attention to social warning indicators. Among them, Zhu Qingfang conceived a comprehensive social warning indicator system composed of more than 40 indicators in four aspects [2], Song Linfei conceived a social risk monitoring and warning indicator system composed of 40 indicators in seven categories [3], Zhang Chunshu aimed at Shanghai’ s For the early warning of social development in large cities, a social early warning index system consisting of 18 warning indicators in 8 categories was proposed, etc. [4].

1.4 Overview of Our work

The topic requires us to establish a social stability early warning model, judge the main reasons for the failure and success of the color revolution, and put forward relevant suggestions to avoid the occurrence of the color revolution. Our work mainly includes the following:

- Qualitative and quantitative analysis of the social system , determine the important indicators that affect social stability ;
- Based on the primary selected social stability indicators, establish a social stability early warning evaluation model;
- Based on the social stability early warning evaluation model, this paper evaluates the reasons for the failure of countries attempting color revolutions, and gives reasonable suggestions for improvement;
- Based on the established social stability early warning evaluation model, analyze the reasons for the failure of regime change in the color revolution, and put forward relevant suggestions for avoiding the color revolution.

First, the social system is analyzed from qualitative and quantitative perspectives, and the clustering algorithm selects important indicators that affect social stability. In this paper, the correlation coefficient is used to analyze the correlation between indicators, and the significance of the correlation coefficient is tested. Second, set up a social stability early warning assessment model. In this paper, the parameters are estimated based on the decision tree police classification model, and the social stability comprehensive evaluation model determines the threshold of the parameters and determines the social stability early warning level . Finally, evaluate the country or region of the Color Revolution. Based on the established social stability early warning evaluation model, this paper evaluates the reasons for the failure of revolution and regime change, and puts forward guiding suggestions in order to avoid color revolutions.

In summary, the whole modeling process can be shown as follows:

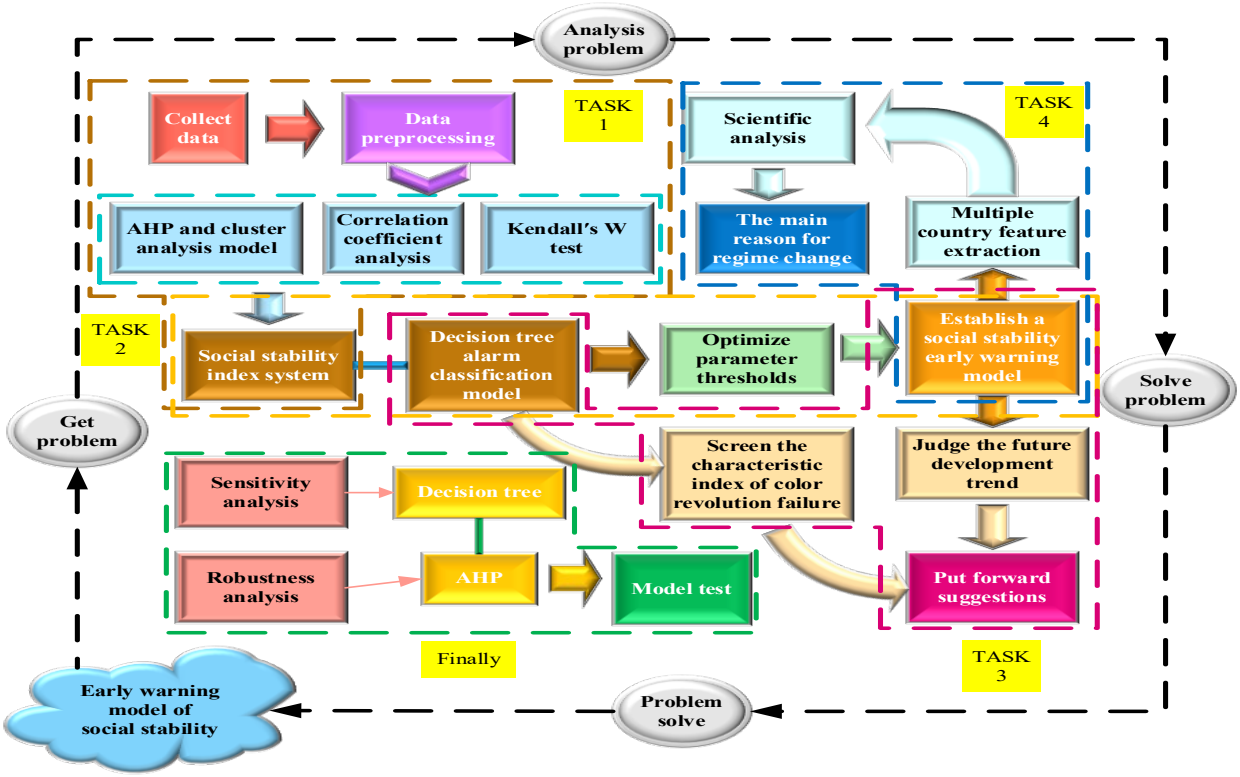


Figure 1: Model Overview

2 Assumptions and Justifications

To simplify matters, we make the following basic assumptions, each of which is reasonable.

- **Hypothesis 1: The social stability indicator system is quantifiable .**

Reason: Although the stable and stable operation of society is intricate , according to the theory of six degrees of separation , the behavior of social stability is limited under certain unpredictable factors , so we can give weight to the indicators .

- **Hypothesis 2: Judging the classification of social stability does not lose generality .**

Reason: We assume that changes can be ignored in the process of selecting social stability indicators, and consider the behavioral laws of social physics and economic operation laws [5] to clarify the feasibility of dividing boundaries..

- **Hypothesis 3: Countries or regions with color revolutions are independent of each other.**

Reason: Because the model takes into account the impact of the country or region where the color revolution occurred on the indicator system, which in turn affects the decision-making on the reasons for the failure of the color revolution. Such a comparison is meaningful only if the external conditions are independent .

- **Assumption 4: Assume that the research data are accurate.**

Reason: We assume that historical population structure data, socio-economic data, medical health data, and social stability index data of some natural disaster accidents have no obvious measurement deviation, which is considered to be false, so we can build a more reasonable model based on this Quantitative model.

3 Notations

Table 1 lists the key mathematical symbols used in this paper .

Table 1: Symbols used in This paper

Symbol	Describe	Symbol	Describe
A	Judgment matrix	d_{ij}	Euclidean distance between i and j
p_{ij}	Scale value	D_{ij}	Distance between class G_i and G_j
CI	Consistency index	D_{kr}	The distance of the new class from other classes
CR	Consistency ratio	x_{norm}	The scaled value of the feature
CR'	Consistency ratio	$D_{(k)}$	k The distance matrix after iterations
w	Weights for Eigenvector Solving	$\hat{\rho}_s$	Spearman rank correlation coefficient
c_i	i total weight of factors	τ	Kendall rank correlation coefficient
$\lambda_{\max}$	Maximum Eigenvalue	KW	Kendall's W test statistic
$w_i^{'}$	Geometric mean solution weights	$Entropy\ (S)$	The information entropy contained in the set S
$w_i^{''}$	Arithmetic average solution weights	$EntropyA\ (S)$	Conditional entropy of attribute to sample set A
$\overline{w_i}$	Average weight	$Gain\ (S)$	Information gain
V	Judging the warning level	RC	Integrated assessment value of phase i

4 Task 1: Study of the Index System affecting Social Stability

4.1 The Data model preparation

In the process of establishing the social stability indicator system, due to the cumbersome collection of quantitative data and the large lack of data, the value of the data itself is not intuitively presented. This study combines qualitative analysis with quantitative analysis, not only making full use of the authoritative knowledge and experience of experts, but also avoiding its subjective arbitrariness as much as possible through certain technical means.

4.1.1 Qualitative Analysis of Social Stability Index

As a complex system, society is an open system that is not absolutely balanced, and its constituent elements will form a specific operating order under the social integration mechanism . Proposed to maintain the social stability of a country or region six indicators support, as follows:

(1) Survival guarantee. The survival security system is the logical starting point of the entire social stability system, and it is at the core of the entire social stability system. Other systems are related to the survival security system at different levels and angles;

(2) Economic support. The economic support system is the material basis for the stable operation of society.

(3) Social distribution. The social distribution system is an intermediate link connecting the economic support system and the survival guarantee system.

(4) Social control. The social control system is a coordination mechanism to maintain social operation.

(5) Social psychology. The state of the psychosocial system is a complete reflection of the state of every other system.

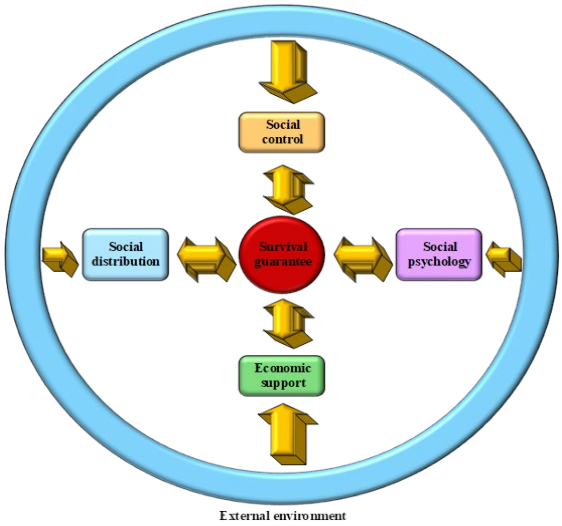


Figure 2: Composition of social stability indicators

(6) External environment. External social factors and natural factors constitute the external environment system for social stability in a specific region [6].

According to the logical self-consistent relationship between the six major support systems, and the nature and status of each of their effects on social stability and security, determine the contribution of each support system, and finally judge the social stability and security of the entire country or region at the overall level The macro situation of the society, and forecast, forecast and early warning of social stability and security situation and development trend.

4.1.2 Ualitative Data Empowerment

Consult the relevant literature [7], this paper mainly adopts the analytic hierarchy process. This method is mainly aimed at those situations with complex structures and lack of necessary data. Based on expert inspection, it combines qualitative analysis with quantitative analysis, and uses methods such as cluster analysis, judgment matrix to find eigenvectors, consistency inspection, and similarity coefficient weighting. Calculate the weight of each indicator.

Step 1 Defines the mathematical expression of the judgment matrix as:

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \tag{1}$$

The numbers 1-9 and their reciprocals are used as the scale, as shown in the table below:

Table 2: Nine-point scale and its definitions

Scaling d_{ij}	Definition
1	i with j important as
3	i with j slightly more important
5	i with j obviously important
7	i much j more important than
9	i with j extremely important
2,4,6,8	i The scale value of the importance of and between two adjacent ranks j
reciprocal of the scale value	i Inverse comparison with $p_{ji}=1/p_{ij}:j$

After hierarchical single sorting, it is necessary to test its consistency and republic, which $\lambda_{\max}$ is the largest eigenvalue of the judgment matrix. Consistency index CI, the calculation formula is as follows.

$$CI=\frac{\lambda_{\max}-n}{n-1} \tag{2}$$

The average random consistency index is RI shown in the table below :

Table 3: Average Stochastic Consistency Metrics RI

	1	2	3	4	5	6	7	8	9	10	11	12
RI	0.00	0.00	0.52	0.89	1.12	1.26	1.36	1.41	1.46	1.49	1.52	1.54

Step 2 To calculate the consistency ratio. CR only at that time , the consistency of the matrix of $CR<0.10$ judging the impact factors of medical care is acceptable, otherwise, make appropriate modifications to the judgment matrix. CR The formula is as follows.

$$CR=\frac{CI}{RI} \tag{3}$$

Step 3 To finally obtain the ranking weight of each factor to the system, so as to select the scheme. Particular attention should be paid to the weighting of indicators during the selection process . Consistency checks are also required for the total sorting of the layers. In the upper level , D there n are factors $P_1,P_2,\cdots,P_n$ whose level weight values are respectively $w_1,w_2,\cdots,w_n$. If the consistency index of the D hierarchical model factors for P_i single sorting is CI , and the corresponding average consistency index is RI_i ,then D the total sorting consistency ratio of the level is :

$$CR' = \frac{\sum_{i=1}^{n} b_i CI_i}{\sum_{i=1}^{n} b_i RI_i} \quad (i = 1, 2, \cdots, n) \tag{4}$$

The weights are solved using the eigenvector method. The weight vector W is multiplied by the weight comparison matrix to the right A, and W the components are all positive components. Finally, the obtained weight vector is normalized to obtain the position, and the solution model is as follows:

$$AW = \lambda_{\max} w \tag{5}$$

Step 4 To calculate the matrix under the condition of the social stability of the focus of the problem , and then continuously modify the calculated matrix to meet the consistency of AHP. The weight coefficient of each factor can be obtained, and the weight coefficient of each factor can be obtained. The total weight of the aspect factors. Its mathematical expression is as follows:

$$c_i = w_1 p_1 + w_2 p_2 + \cdots + w_n p_n$$

Among them, w refers to the weight of the p indicator , and refers to the relevant index of social stability .

Step 5 The weighting results of the social stability index system are as follows.

Table 4: Judgment Matrix of Social Stability Index System

Index system	d11	d12	d13	d14	d15	d16	w_i
Survival Guarantee	1	2	5	7	3	4	0.3571
economic support	1/2	1	6	9	4	5	0.3317
social distribution	1/5	1/6	1	3	1/4	1/3	0.0554
social control	1/7	1/9	1/3	1	1/7	1/5	0.0287
social psychology	1/3	1/4	4	7	1	1/5	0.1272
external environment	1/4	1/5	3	5	1/3	1	0.0999
$\lambda_{\max} = 6.1623$, CI=0.0325, CR<0.10, passed the consistency test.							

It can be seen that $CR < 0.10$, through the consistency test , it is believed that the empowerment of the social stability index system is feasible, and the survival guarantee and economic support indicators play a key role in social stability.

4.1.3 Quantitative Data Collection

The data we use mainly includes historical demographic data, socioeconomic data, medical and health data, and some social stability indicators of natural disasters . Table 5 summarizes the sources of quantitative data .

Table 5: Data source collation

Name database	Database website data	Type
Office for National Statistics	http://www.stats.gov.cn/	statistical database
OECD Data	https://data.oecd.org/	Industry database
China Statistical Yearbook	http://www.stats.gov.cn/tjsj/ndsj/	Industry Reports
Google Scholar	https://scholar.google.com/	academic paper
Cnki.net	https://www.cnki.net/	academic journals

4.1.4 Data cleaning

data from 1980 to 2020, group by year , and calculate the average value of key data in each group. For missing values in the data, we try to skip and only find the effective mean. For complete missing groups, values are recorded as missing values. Missing values are then linearly interpolated along the time axis. If there are four or more missing values in a column, the data in that column is considered invalid. Finally set the position of the invalid data column as unreachable, ignored in the model calculation.

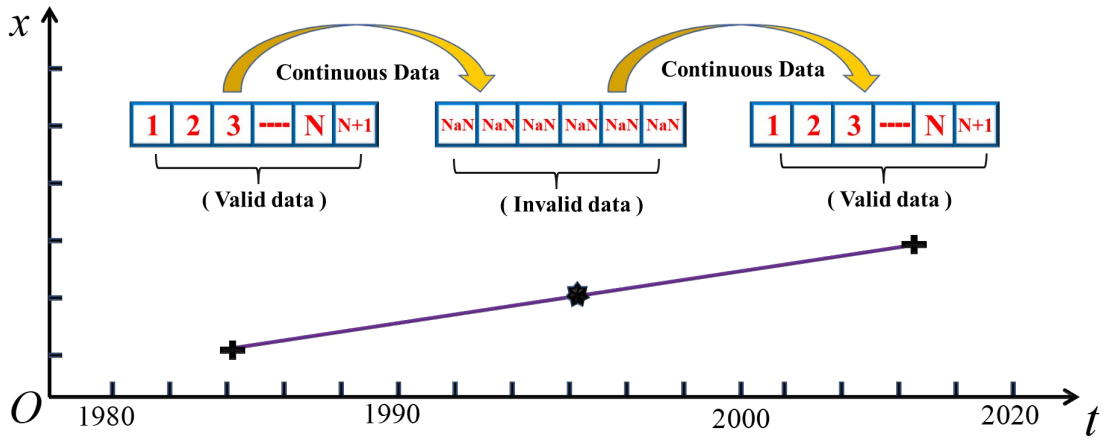


Figure 3: Data Cleaning

4.2 Selection of Social Stability Indicators

According to the relevant results of social stability theory research, combined with the characteristics of politics, economy, culture, society, etc., this paper initially selects six first-level indicators, including subsistence security, economic support, social distribution, social control, social psychology, and external environment. 72 secondary indicators. However, in the early warning analysis of social stability, it is not that the more indicators selected, the better. Since the correlation between various indicators and the contribution to social stability are difficult to judge subjectively, it is impossible to determine the structure and main characteristics of the entire social stability system. In order to reduce the amount of data collected and accurately reflect the social stability According to the characteristics and development trend, this paper adopts the cluster analysis method to further screen the index system [8].

4.2.1 Principle of Cluster Analysis

Cluster analysis , also known as group analysis, is a multivariate statistical method to study the classification of indicators (or samples). Its core idea is to regard the indicator (or sample) as P a point in the dimensional space, and define the distance in the space , the indicators (or samples) with a closer distance are classified into one category6. In this paper, square Euclidean distance is used as index cluster analysis:

$$d_{ij} = \sum_{a=1}^{p} \left| X_{ia} - X_{ja} \right|^2 \tag{6}$$

step 1 To calculate the distance between indicators d_{ij}, and get a distance matrix $D_{(0)}$.

At the beginning, each indicator forms a class by itself, that is, at this time, $G_i = \{X_i\}$ the distance between the $D_{ij} = d_{ij}$ class G_i and G_j the

$$D_{(0)} = (D_{ij}) = (d_{ij}) = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{n1} \\ d_{21} & d_{22} & \cdots & d_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ d_{n1} & d_{n2} & \cdots & d_{nn} \end{bmatrix} \tag{7}$$

In, $d_{ii} = 0, d_{ij} = d_{ji}, i \neq j.$

Step 2 Minimum off-diagonal element D_{pq} found in $D_{(0)}$ is $G_r = \{G_P, G_q\}$.set G_p as G_q

Step 3 Defines the distance between classes as the distance between the two closest indicators, namely.

$$D_{ij} = \min d_{ij}, \quad x_i \in G_i, x_j \in G_j \tag{8}$$

And define the distance of the new class from other classes:

$$D_{kr} = \min\{D_{kp}, D_{kq}\} \tag{9}$$

After calculation, the new matrix obtained is denoted as $D_{(1)}$;

$$D_{(1)} = D_{(ij)} = \begin{bmatrix} D_{11} & D_{12} & \cdots & D_{1n_1} \\ D_{21} & D_{22} & \cdots & D_{2n_1} \\ \cdots & \cdots & \cdots & \cdots \\ D_{m1} & D_{m2} & \cdots & D_{nn_1} \end{bmatrix} \tag{10}$$

Among them, n_1 is the number of classes after the first merging (iteration) (obviously n_1 smaller than the initial number of classes n) $D_{ij} = 0$, $D_{ij} = D_{ji}$, at that time $.i \neq j$

Step 4 Repeat the $D_{(0)}$ first and second steps of the $D_{(2)}$ above pair, and the $D_{(1)}$ number of classes (that is, the number of iterations) is recorded as n_1 (less than n_1); and so on, until all elements are combined into one class, and the number of classes at this time is recorded as n_s. (Actually $n_s = 1$). It s is called the final number of iterations. If $D_{(k)}$ there is more than one off-diagonal minimum element in a certain step, the classes corresponding to these minimum elements can be merged at the same time.

4.2.2 Program Implementation Process

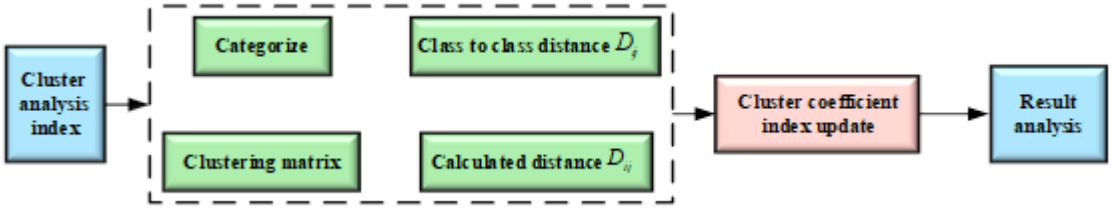


Figure 4: Process of primary selection indicators for cluster analysis

4.2.3 Result Analysis

Through the cluster analysis of the primary index system , this paper finally selects 11 indicators to form the social stability index system as shown below.

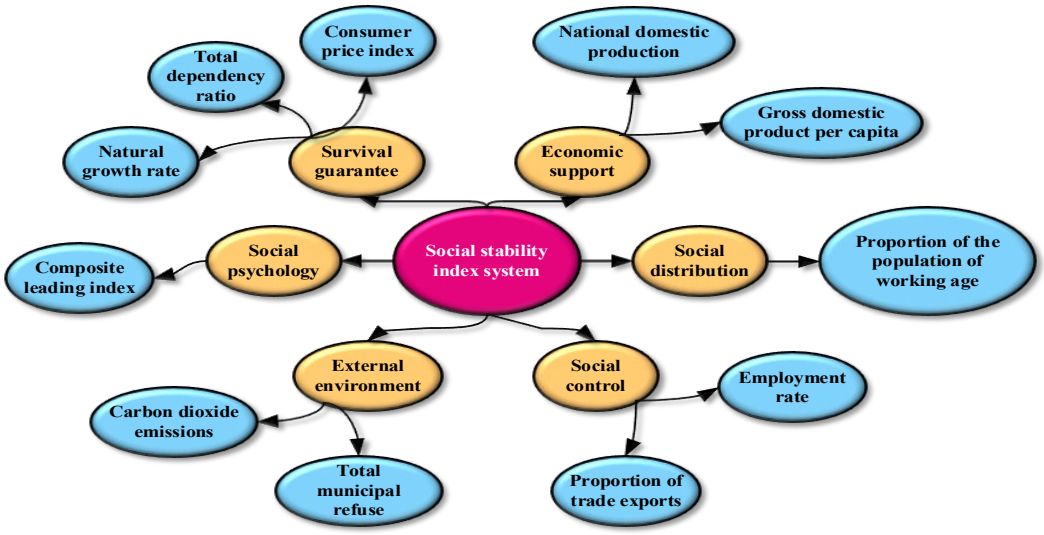


Figure 5: Social Stability Index System

4.3 Correlation Analysis of Social Stability Indicators

In order to eliminate the influence of dimensions, this paper needs to perform normalization processing before using the algorithm to model, that is, feature scaling, which can not only improve the speed of algorithm solution, but also improve the accuracy of model recognition. The formula is as follows:

$$x_{norm} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \times 10 \tag{11}$$

Among them, x is the original eigenvalue, x_{norm} is the value after feature scaling, $x_{\max}$ is the maximum value of $x_{\min}$ the original eigenvalue, and is the minimum value of the original eigenvalue.

Considering the degree of influence of a single factor on social stability indicators and the quantitative analysis of the degree of influence on characteristic indicators, the correlation analysis in probability theory and statistics is used to carry out weighted screening of the significance indicators of the degree of influence [8]. The idea and model flow chart are as follows:

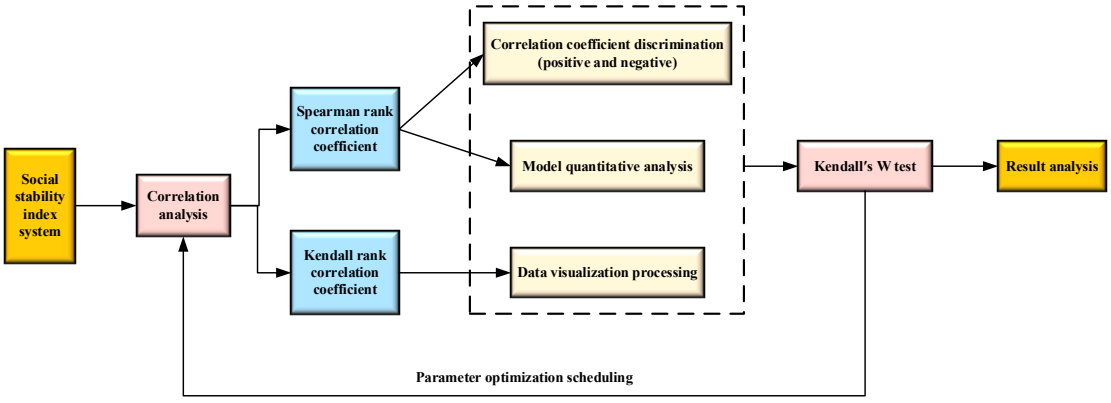


Figure 6: Idea and model flow chart

4.3.1 Correlation Coefficient

Correlation shows the strength and direction of a linear relationship between two or several random variables. Two-team correlation analysis refers to the analysis of two or more correlated

variable elements, so as to measure the closeness of the two variable factors. Correlation coefficients are usually used between the elements of correlation to measure the degree of co-variation of these random variables. When the random variables show a trend of change in the same direction, it is called positive correlation, otherwise it is called negative correlation.

◆**Spearman correlation coefficient** It is a non-parametric measure of the correlation between two variables. It uses a monotonic function to evaluate the correlation of two statistical variables. If there are no repeated values in the data, and when the two variables are completely monotonously correlated, the Spearman correlation coefficient is +1 or −1 . It is a non-parametric method, and it can also be seen from its name that it is a rank-based method [9] . which is defined as.

$$\hat{\rho}_s = \frac{\sum_{i=1}^{n}(R_i - \bar{R})(Q_i - \bar{Q})}{\sqrt{(R_i - \bar{R})^2}\sqrt{(Q_i - \bar{Q})^2}} \tag{12}$$

In

$$\bar{R} = \frac{1}{n}\sum_{i=1}^{n} R_i \qquad \bar{Q} = \frac{1}{n}\sum_{i=1}^{n} Q_i$$

Because

$$\sum_{i=1}^{n} R_i = \sum_{i=1}^{n} Q_i = \frac{n(n+1)}{2} \qquad \sum_{i=1}^{n} R_i^2 = \sum_{i=1}^{n} Q_i^2 = \frac{n(n+1)(2n+1)}{6}$$

◆ **Kendall correlation coefficient** It is the method proposed by Maurice Kendall in 1938. There are two random variables x ; y , which x ; y can be (ordered) categorical variables or continuous variables, and the sample size is n . Defined $\left(x_i\ ;\ y_j\right)$ as a sample pair, the calculation formula for Kendall is as follows:

$$\tau = \frac{G - H}{\sqrt{(n_3 - n_1)(n_3 - n_2)}} \tag{13}$$

In

$$n_3 = \frac{1}{2}n(n+1) \qquad n_1 = \sum_{i=1}^{s}\frac{1}{2}u_i(u_i - 1) \qquad n_2 = \sum_{i=1}^{t}\frac{1}{2}v_i(v_i - 1)$$

4.3.2 Kendall's W Test

In order to improve the recognition degree of the model to the sample data, this study uses Kendall's W test to test the correlation of social stability factors and optimize the model [9]. Kendall concordance analysis, the KW statistic is called concordance coefficient or consistency coefficient, which is used to measure the consistency. The closer the coefficient is to 1. The higher the consistency. Calculated as follows:

$$KW = \frac{12\mathrm{S}}{\mathrm{K}^2\left(\mathrm{N}^3 - \mathrm{N}\right)} \tag{14}$$

In the formula, S is the sum of squares of the difference between the rank sum and its average value, the K number of groups evaluated for rank, and N the number of objects evaluated for rank. Using the chi-square test, construct the statistic $\chi^2 = K(N-1)_{rw}$, where the degrees of freedom are $N-1$.

Table 6: Kendall's W test results

Indicator name	Rank mean	Median	Kendall's W coefficient	P
Total dependency ratio	3.122	7.58		
Percentage of working age population	6.098	45.9		
National Domestic Production	6.878	68.545		
Gross Domestic Product per capita	10.39	89366.5		
GDP per capita	9	7229		
Employment rate	1.024	0.563	0.969	0.000 ***
Consumer Price Index	4.073	18		
Emissions of CO2	2.39	2.45		
Municipal waste	10.61	158048		
Trade export share	4.415	18.541		
Overall Leading Index	8	100.919		

Note: ***, **, * represent the significance levels of 1%, 5%, and 10% respectively

It can be seen from the table that the significance P value of the Kendall W coefficient consistency test for the overall data is 0.000*** , which is significant at the level and rejects the null hypothesis, so the data is consistent, and the Kendall coordination coefficient W of the model is 0.969, so the degree of correlation is high agreement .

4.4 Result Analysis

Kendall rank correlation coefficient were analyzed for the social stability index system , and the results are shown in the correlation coefficient heat map below.

(a) Spearman coefficient (b) Kendall coefficient

Figure 7: Correlation coefficient heat map

It can be seen from the figure that the correlation coefficients between Percentage of working age population and National Domestic Production , Gross Domestic Product per capita , and GDP per capita are all greater than 0.9, and the positive correlation is the highest. It is considered that the degree of influence on them is relatively high.

5 Task 2 : Early warning Assessment of Social Stability

The purpose of constructing a social stability early warning and evaluation model is to evaluate and monitor social risks, and to do this, scientific evaluation methods must be adopted, and the comprehensive evaluation method based on decision trees is a widely used method at present [10]. This model proposes a regional social stability risk assessment method in order to provide a reference for early warning assessment of social stability.

5.1 Decision Tree Police Classification Model

5.1.1 Principle of Decision Trees

Decision tree algorithm is a commonly used classification prediction method in data mining technology. Its decision-making process is to start from the root node, test the corresponding feature attributes in the item to be classified, and select the output branch according to its value until reaching the leaf node, and use the category stored in the leaf node as the decision result. [11].

Three common algorithms for constructing decision trees , namely ID3, C4.5 and CART algorithms. This paper uses the ID3 algorithm, and selects the attribute with the largest information gain value in the sample set as the test attribute, that is, as each non-leaf node [14] .

S set as the sample collection, $s_i \in S(i=1,2,3,\cdots,s)$, and the category attribute is $C_i(i=1,2,3,\cdots,m)$. Assuming that s_i it is the number of samples in the S category , C_i the information entropy contained in the set is:

$$Entropy\ (S) = -\sum_{i=1}^{m} p_i \log_2 p_i \tag{15}$$

Where is $p=s_i/S$ the probability that any one data object belongs to. C_i Assuming A that the samples in the set S are divided by attributes, and there are k different values in the attribute A, then A the conditional entropy [12]is:

$$EntropyA(S) = \sum_{i=1}^{k} \frac{|s_i|}{|S|} Entropy(s_i) \tag{16}$$

Among them, where |s| and |S| are s_i the number of samples contained in and respectively. S Using attributes to A divide the sample set S , the information gain $Gain(A)$ is:

$$Gain(A) = Entropy(S) - EntropyA(S) \tag{17}$$

This paper uses the above formula to calculate the information gain of each feature sample data set of voice service and Internet service.

5.1.2 Program Implementation

Decision tree model is one of the commonly used algorithms in the field of data mining and machine learning. This paper is classifying and predicting the early warning level of social stability. The decision tree classification algorithm [12] is as follows:

Algorithm : Parameter Estimation of Decision Tree Police Classification Model

Generate node root node

If _

The samples in D all belong to the same category C_k then

Mark the node as C_k a leaf node according to the formula (15)

return
end if
The samples in D all belong to the same category C_k then
Mark the node as C_k a leaf node according to the formula (16)
return
end if
If A = Empty OR The samples in D take the same value in A then
Mark the node as a leaf node, and its category is marked as D the class with the largest number of samples in
return
end if
Select the optimal partition attribute from A a_*
F or
a_* Each value of a_*^v do
Generate a branch for node D_v according to the formula (17) : Let the sample subset of the a_* upper value in
The a_*^v representation be D
if D_v is empty then
Mark the branch node as a leaf node, and its category is marked as D the class with the most samples in
return
else
Branch node $\mathrm{TreeGenerate}(\mathrm{D_v}, \mathrm{A} - \{\mathrm{a_*}\})$
end
end

5.1.3 Estimated Parameters

Table 7: Decision Tree Model Parameters

Parameter name	Parameter value
Cross-validation	10
Node Split Evaluation Criteria	gini
Feature division point selection criteria	the best
Maximum proportion of features to consider when partitioning	none
Minimum number of samples for internal node splits	2
The minimum number of samples for a leaf node	1
The minimum weight of samples in leaf nodes	0
Maximum number of leaf nodes	50
the maximum depth of the tree	10

5.2 Comprehensive Evaluation Model of Social Stability

5.2.1 Parameter Threshold Calculation

To determine the judgment and early warning level, the social stability risk level is divided into five levels: no police, light police, medium police, heavy police, and major police [13]. The corresponding judgment level is 4 intervals and 5 thresholds. Let the evaluation warning level be V, then it can be defined as:

$$V = \left[V_1, V_2, V_3, V_4, V_5\right]^T \tag{18}$$

In the formula, $V_1 \sim V_5$ from low to high, it represents the risk level of a single risk factor,

which is no alarm, light alarm, medium alarm, heavy alarm, and severe alarm, and the assignment value is $V=[0,0.15,0.35,0.5,1]^T$ [10].

After obtaining the social judgment warning level threshold, the comprehensive value of each indicator can be multiplied by its corresponding weight, and the total can be added to obtain a comprehensive measure of the overall social stability assessment. The formula for calculating the final comprehensive evaluation value of the first period is: i

$$RC=\sum_{i=1}^{5}A_iV_i \tag{19}$$

Among them, A_i is i the prediction score of the period.

5.2.2 Social Stability Early Warning Level

After calculating the social risk subsystems and the overall evaluation value, the risk level of social stability can be judged according to the table, and marked with corresponding early warning signals [11].

Table 8: Social Stability Police Level Evaluation Form

Judging grade	0-2	2-4	4-6	6-8	8-10
police rank	no police	light police	Central police	heavy police	giant police
Signal	green light	blue light	yellow light	orange light	red light

In the table, the green light indicates that the possibility of social risk occurrence is very small, and the society is very safe; the blue light indicates that the society is relatively safe, but it is in a turbulent turning point; the yellow light indicates that social risks may occur at any time and aggravate social unrest; the orange light indicates that the society In a state of chaos, social risks have affected social stability; red lights indicate that society is in a state of extreme chaos, and large-scale social risk events may occur at any time.

5.3 Results Analysis and Discussion

Table 9: Fitting parameters of the model

	Accuracy	recall	Accuracy rate	F1
Training set	1	1	1	1
Cross-validation sets	0. 6 83	0. 6 83	0. 6 69	0. 6 64
Test set	0. 8 91	0. 8 91	0. 8 55	0. 8 68

It can be seen from the table that the accuracy rate, recall rate, precision rate and F1 of the fitting parameters of the training set are all equal to 1, indicating that the model has a high degree of recognition of the data and the accuracy of the model evaluation is high.

6 Task 3 : Reasons why the Color Revolution Failed

As for the "color revolution", it mainly refers to the revolution that occurred in the Commonwealth of Independent States and Central Asia after the disintegration of the Soviet Union at the beginning of this century. A revolution that overthrows the current government through flowers and other objects of a specific color. For example, Georgia is rich in roses. At that time, the opposition Saakashvili would take a bouquet of roses every time he appeared on the stage, so it was called the "Rose Revolution" and so on [13]. The following article focuses on Myanmar .

6.1 Feature Selection for Failure Reasons

Based on the social stability early warning evaluation model to fit the data of Myanmar, the model is optimized, and the selected characteristic indicators are as follows:

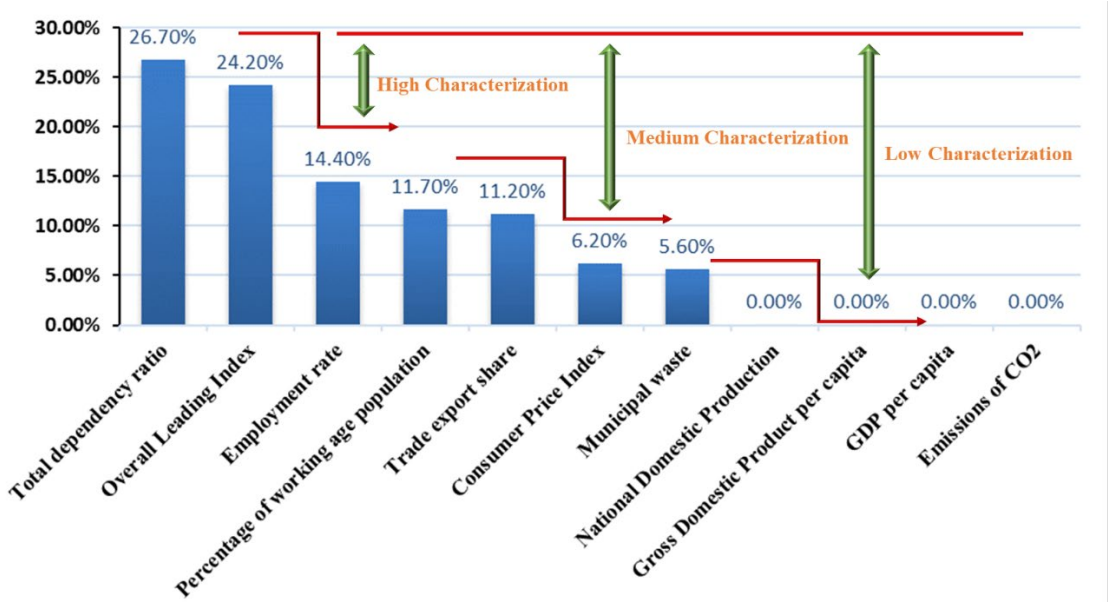
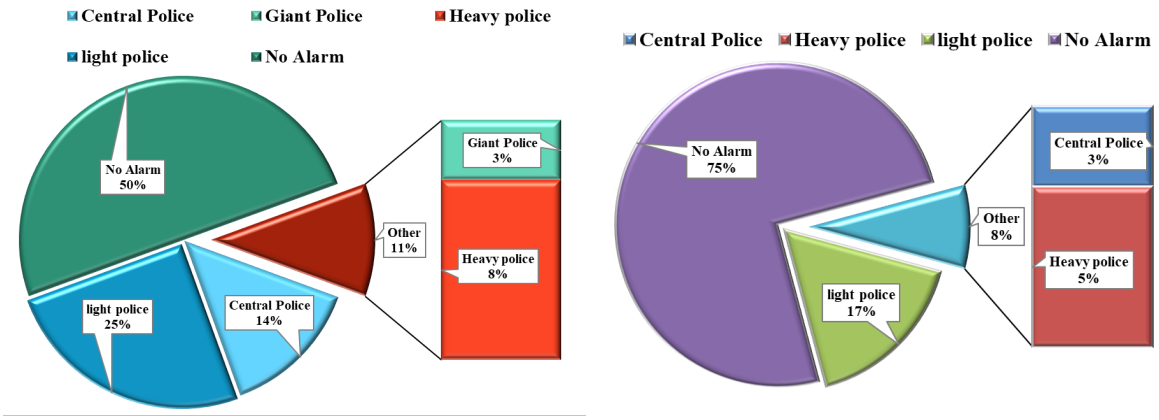


Figure 8: Important features of a decision tree model

It can be seen from the figure that in the color revolution countries or regions represented by Myanmar, the proportion of natural growth rate and trade export ratio to the characteristics of failure reasons exceeds 50% of the total, showing a resistance to it , which is beneficial to the country or region harmonious operation.

6.2 Future Trend Forecast

Based on the selection of important features, this paper uses the natural growth rate and trade exports to predict social stability. The accuracy of the model drops by 1%, and the impact can be ignored. To this end, the social status of Myanmar in the next 50 years is evaluated, and the comparison before and after is as follows:



(a) Social stability before 2020 **(b) The state of society in the past 50 years**

Figure 9: Forecast of future trends in Myanmar

It can be seen from the figure that in the color revolution countries or regions represented by Myanmar, before 2020, due to the blocking effect of the color revolution, the giant police

level appeared in the society and accounted for more than 10%, while the non-police level was only 50% Around %, social unrest led to a decline in social productivity and increased poverty in people's lives.

In the 50 years after the passage, this paper studies the disappearance of the giant police level in the social state, the warning reminder has weakened to about 8%, while the non-police level exceeds 75%, the society is becoming more harmonious, the social productivity has been greatly improved, and the people's life is prosperous.

6.3 Results and Recommendations

The countries touched by the "color revolution" will always lead to social differentiation. Social differentiation will lead to the gradual widening of the income gap and living standards between different groups in society. Mechanism, so as to build a harmonious society [13]. we suggest:

- Maintain a stable economic environment, strive to improve people's living standards, and develop harmoniously in society.
- Promote international exchanges, strengthen the implementation of education and develop foreign trade.
- Improve the people's willingness to bear children and provide preferential living security for the people.

7 Task 4 : Regime Change in Color Revolutions

7.1 National Distribution of Color Revolutions

The color revolution affected many countries. The countries involved in the color revolution were in Europe, Central Asia, and South Asia. This region included both the countries of the Soviet Union at that time and the sphere of influence of the Soviet Union. The overall shape is "semi-circular", which is caused by the US strategy.

The color revolutions in European countries are mainly Georgia, Ukraine, and Belarus. These three countries were once republics of the Soviet Union. Although Georgia is located in Asia, it has always been counted within the scope of Europe. Belarus, Ukraine, and Georgia have deep ties to Russia [14].

The countries affected by Central Asia are mainly Kyrgyzstan, Kazakhstan and other countries. The five countries in Central Asia are also former republics of the Soviet Union . It is normal for them to be affected by the color revolution, but the most severely affected by the color revolution is Kyrgyzstan.

The West Asian countries are mainly affected by Iran, but the Iranian government is very strong, and the revolution did not succeed. Although Iraq, Lebanon and other countries in the Middle East were affected by the color revolution, such as the "Purple Revolution" in Iraq, they used purple ink to print their fingerprints and got their name. The revolution in Lebanon was not big. It was named after the country's national tree "Cedar", and it was called the "Cedar Revolution [14].

South Asia was affected by Burma, where the "saffron revolution" against the military government took place, also known as the "cassock red revolution".

Countries or regions affected by color revolutions since 2020.

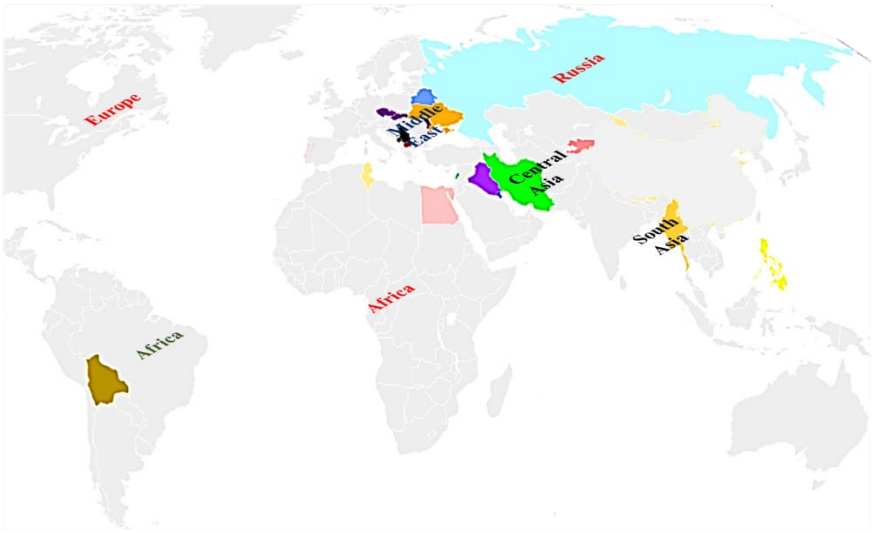


Figure 10: National distribution of color revolutions

7.2 Reasons for Regime Change

Considering the countries or regions where the color revolution occurred, this article selects the main representative countries, such as Georgia, Ukraine and Belarus , for discussion. According to the distribution of social stability early warning assessment model, the important features of Georgia, Ukraine and Belarus are selected.

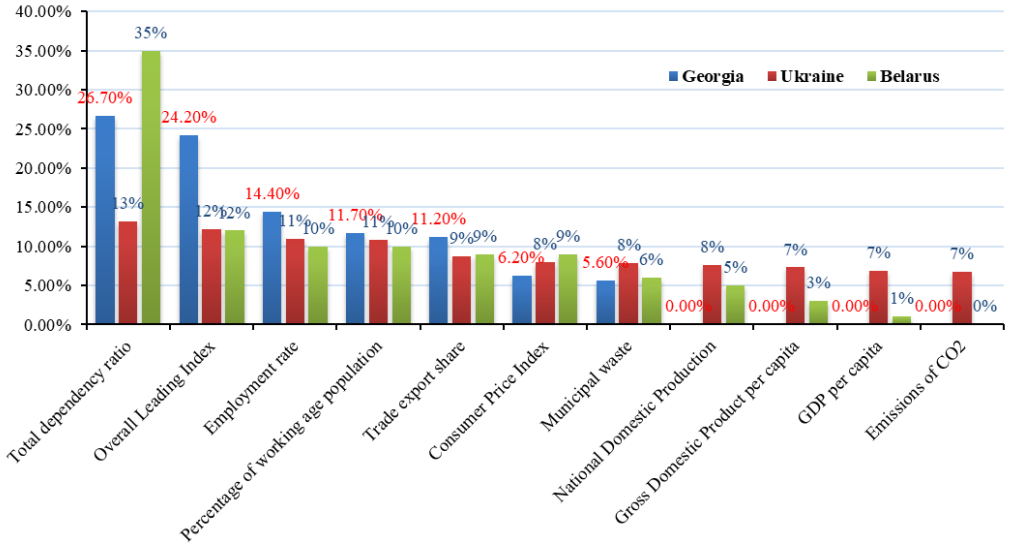


Figure 11: Comparison of reasons for regime change

It can be seen from the figure that the influence of natural growth rate and trade export on the regime change of each country is more than 10%, while the influence of other characteristic indicators is close to 0, indicating that it plays a slowing role in the process of regime change ; while in Belarus The characteristic indicators are all over 7%, which is basically the same.

8 Task 5: Avoid Color Revolutions and Maintain Social Stability

The Color Revolution, also known as the Flower Revolution, refers to the peaceful and non-violent regime change movement named after the color that occurred in the Commonwealth of

Independent States and Central Asia at the end of the 20th century. Participants usually adopt a special color or flower as their symbol, resist what they believe to be a dictatorial government through non-violent means, and support democracy, freedom and national independence [15]. At present, color revolutions have overthrown the original pro-Russian governments in Georgia, Ukraine, and Kyrgyzstan, and established pro-American democratically elected governments.[16]

Factors that produce a color revolution:

1. Economic field: economic development is slow, and people's lives have not been improved;
2. Political field: Verification of corruption within the ruling party, public disapproval and dissupport of the ruling party;
3. In the field of culture: multiple values cause the rupture of thought, leading to the confusion of value and thought;
4. Public opinion field: public dissatisfaction. If social governance cannot keep up, people's livelihood issues cannot be effectively resolved, and public goods are in short supply, public dissatisfaction will be aroused.
5. The opposition of regions and ethnic groups is the historical and cultural background that cannot be ignored in the color revolution;
6. The key factor: a profound contradiction between production relations and productivity

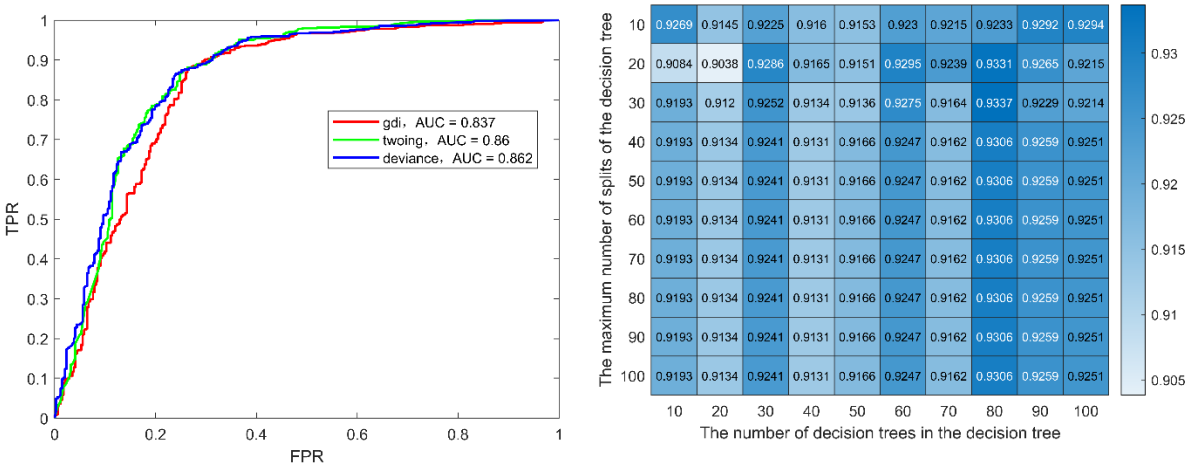
Tips to prevent color revolutions from happening:

1. Vigorously develop regional economic construction;
2. Strengthen the construction of political culture and gain the recognition and support of the people;
3. Spread the correct core values to guide people's thinking;
4. Pay attention to public opinion, adjust the negative emotions of the people, and attach importance to national political and cultural education. Since young people are immature and easily exploited by revolutionary organizations, it is necessary to strengthen the cultural education of young people.
5. Coordinating antagonistic relations between regions and ethnic groups;
6. The study of this paper shows that the key indicators of the contradiction between production relations and productivity are "population growth rate and trade export rate". An increase in population growth rate and a decrease in trade export rate will directly increase the economic burden on the people. The country should formulate a family planning policy to control the population growth rate; at the same time control customs duties, strengthen international trade with foreign countries, and ensure trade export rates.

9 Test model

9.1 Sensitivity Analysis

In Section 5.1, the splitting criteria of the decision tree police classification model are introduced, including gdi, twoing, and deviance . At the same time, the maximum number of splits is iterated from 0 to 100, and the appropriate parameters are searched to make the ACU value of the de-modeling the highest. Therefore, the calculation results are shown in Figure 10 .



(a) ROC curves of different splitting criteria **(b) Iteration of the maximum number of splits**

Figure 12: Sensitivity Analysis Split Criteria and Maximum Number of Splits

Shows that the split criterion is deviance , and the model fitting accuracy is the highest when ACU=0.862. At the same time, when the iteration of the maximum number of splits is 80, ACU=0.9295, the fitting accuracy is the highest, which reflects the feedback effect of parameter selection on model accuracy . The model trend obtained by the sensitivity test is consistent with the actual situation, which also proves the rationality and robustness of the social stability early warning assessment model .

9.2 Robustness Analysis

In the AHP, one of the methods is generally used to find the weight vector, and the results are obtained. According to the cited literature [7], we found that the problem can be solved by comparing the weights calculated by the four conventional calculation methods. But there are also subtle differences. Considering that the subtle differences may lead to different results in solving this practical problem, this paper adopts the geometric mean method, arithmetic mean method , and eigenvector method to solve the weight vector, compare each other and take the mean value as its weight , improve the robustness of the model, and take comprehensive considerations to make more scientific decisions. The following is a brief description of the calculation method.

Geometric mean method : first A multiply the elements of the new vector by row, then divide each component of the new vector to the n power, and finally normalize the obtained vector to be the position weight vector. The solution formula is as follows:

$$w'_i = \frac{(\prod_{j=1}^{n} a_{ij})^{\frac{1}{n}}}{\sum_{k=1}^{n}(\prod_{j=1}^{n} a_{kj})^{\frac{1}{n}}} (i = 1,2,3,\cdots,n) \tag{20}$$

Arithmetic mean method: first A normalize the elements by column, that is, to calculate $a_{ij} / \sum_{k=1}^{n} a_{kj}$, add the normalized columns , and finally divide the added vector by n the weight vector. The solution formula is as follows :

$$w''_i = \frac{1}{n}\sum_{j=1}^{1}\frac{a_{ij}}{\sum_{k=1}^{n} a_{kj}}(i=1,2,3,\cdots,n) \tag{21}$$

Weight mean value: the geometric mean method, arithmetic mean method and eigenvector method will be used to solve the weight vector, compare each other and take the mean value as its weight, the formula is as follows:

$$\overline{w_i} = \frac{1}{n}\sum_{i}^{n} w_i (i=1,2,3,\cdots,n) \tag{22}$$

Therefore, the calculation results are as follows.

Table 10: AHP weight calculation in robustness test

Index system	w_i	w_i'	w_i''	$\overline{w_i}$
Survival Guarantee	0.3571	0.3709	0.3651	0.3644
economic support	0.3317	0.3446	0.3345	0.3369
social distribution	0.0554	0.0544	0.0533	0.0544
social control	0.0287	0.0279	0.0279	0.0282
social psychology	0.1272	0.1064	0.1226	0.1187
external environment	0.0999	0.0958	0.0966	0.0974
$\lambda_{max} = 6.1623$, CI=0.0 1 25 , CR<0.10, passed the consistency test.				

10 Model Evaluation and Further Discussion

10.1 Advantage

- The decision-making process is closer to the human mind, so the model is easier to explain.
- Does not require any domain knowledge and parameter assumptions;
- It is proposed that when designing social early warning indicators, we should focus on those sensitive indicators with strong warning effects. Therefore, the designed indicator system is composed of sensitive indicators with strong warning function.
- The sensitivity analysis of the model demonstrates the validity of the model under different parameter combinations and proves the robustness of the model.

10.2 Possible Improvements

- For data with unbalanced sample size of each feature, information gain is more biased towards features with more values. If there is more complete data, we can analyze the social stability index system more accurately ;
- The decision tree model is easy to overfit, and the accuracy of the model can be adjusted by cutting the branches and leaves;
- The influencing factors of social stability are complex and can vary from time to time and from region to region. In the process of identifying and analyzing risk factors, this paper mainly uses reference materials and subjective analysis of some survey data feedback phenomena.

References

[1] Li Qingru, Ye Dongqing. The founder of "Political Arithmetic"—William Petty[J]. Chinese Journal of Disease Control, 2020,24(03):369-372.

[2] Wu Yuanjing. On the research paradigm and index system of social early warning theory at home and abroad[J]. Economic Research Guide, 2015(13):252-253.

[3] Song Linfei.Social Risk Index System and Social Fluctuation Mechanism[J].Sociological Research,1995(06):90-95.

[4] Tong Yumei. Research on Social Stability Index and Its Index System in Risk Society [D]. Zhejiang University, 2019. .

[5] Yan Yaojun.The Measurement of Social Stability and the Construction of Early Warning and Pre-control Management System[J].Sociological Research,2004(03):1-10.

[6] Fang Yong. On the Application of Social Combustion Theory to "Major Events"——Theoretical Basis and Method Model of Social Stability Risk Assessment of Major Events[J].Journal of the Party School of Zhejiang Provincial Committee of the Communist Party of China,2012,28(04):106-111 .

[7] Zhu Zhubang, Wu Wanshui, Sun Jinzhong, Xiao Chun.Research on Social Stability Monitoring Index System and Early Warning Methods in Guangdong Province[J].Practice and Understanding of Mathematics,2013,43(11):33-41.

[8] Chen Qiuling, Xiao Lu, Cao Qingjin. Design of Social Early Warning Index System and Definition of Early Warning Judgment Standards——Based on the Perspective of Social Stability[J]. Public Management Forum, 2008(01):62-76.

[9] Yan Yaojun, Lei Ming. Time-series analysis and simulation early warning of social harmony and stability in my country——Explore the possibility of implementing feedforward control using social index theory and methods[J]. Journal of Tianjin Administration College, 2006(02):71-77.

[10] Li Dianwei, Zhao Liming.Research on Social Stability and Risk Early Warning Mechanism[J].Economic System Reform,2006(02):29-32.

[11] Wang Li. Research on the construction of early warning system for mass incidents[D]. Nanjing University of Posts and Telecommunications, 2011.

[12] Zhang Dan, Liu Huan. A Digital Sharing Method of Electronic Archives Resources Based on Decision Tree Feature Selection[J].Automation Technology and Application, 2022,41(10):125-128.DOI:10.20033/j.1003-7241.(2022) 10-0125-04.

[13] Color Revolution in Myanmar[J]. Party and Government Forum , 2007(12): 9.

[14] Liu Shaohua. The Causes of the "Color Revolution" and Its Enlightenment to my country[J]. Journal of the Party School of Shijiazhuang Municipal Committee of the Communist Party of China, 2006(03):39-41.

[15] Zeng Zhisheng. "Color Revolution" and Its Thoughts[J]. Journal of Tianjin Administration Institute, 2008(01):42-46.

[16] Zhang Ji, Zhang Aili. On the Causes of "Color Revolution" in Central Asian Countries and its Enlightenment to my Country - Analysis from the Perspective of Cultural and Ideological Security[J]. Journal of Hebei University of Economics and Business (Com-prehensive Edition), 2007(04): 19-24.

Appendices

Appendix 1: Tools and Software

Paper written and generated via Office 2021.
Graph generated and calculation using MATLAB R2021a.
Modeling calculations via SPSS.26
Data Visualization via Origin.20222

Appendix 2: Problem solving code

Introduce: Social Stability Early Warning Evaluation Model Code

```
clear;clc
filename = 'data.csv' ;
delimiter = ',' ;startRow = 2;
formatSpec = '%f%C%C%q%C%f%f%f%q%f%C%C%[^\n\r]' ;
fileID = fopen(filename, 'r' );
dataArray = textscan(fileID, formatSpec, 'Delimiter' , delimiter, 'TextType' ,
'string' , 'EmptyValue' , NaN, 'HeaderLines' ,startRow-1, 'ReturnOnError' ,
false, 'EndOfLine' , '\r\ n' );
fclose(fileID);
data = table(dataArray{1:end-1}, 'VariableNames' , { 'PassengerId' , 'Sur-
vived' , 'Pclass' , 'Name' , 'Sex' , 'Age' , 'SibSp' , 'Parch' , 'Ticket' ,
'Fare' , 'Cabin' , 'Embarked' });
clearvars filename delimiter startRow formatSpec fileID dataArray ans ;
trainingData = data1; inputTable = trainingData;
predictorNames = { 'VarName1' , 'VarName2' , 'VarName3' , 'VarName4' };
predictors = inputTable(:, predictorNames);
response = inputTable.VarName5;
isCategoricalPredictor = [false, false, false, false];
ClassNames = unique(response); % The name of each class in y
group = response; % The column where y is located (the real category)
SplitCriterion = { 'gdi' , 'twoing' , 'deviance' };
MaxNumSplits = 1:30; MinLeafSize = 1:5;
num_i = length(SplitCriterion); num_j = length(MaxNumSplits);
num_q = length(MinLeafSize); MICRO_F1_SCORE = zeros(num_i,num_j);
mywaitbar = waitbar(0); % set a progress bar
TOTAL_NUM = num_i*num_j*num_q; now_num = 0;
for i = 1:num_i
    for j = 1:num_j
        for q = 1:num_q
rng(520) classificationTree = fitctree(predictors,response, 'SplitCriterion' ,
SplitCriterion{i}, ...
```

```
'MaxNumSplits' , MaxNumSplits(j), 'MinLeafSize' , MinLeafSize(q), 'Surrogate' ,
'off' , ... 'ClassNames' , categorical({ ' x1 ' ; ' x2 ' ; ' x3 ' })) ;
predictorExtractionFcn = @(t) t(:, predictorNames);
treePredictFcn = @(x) predict(classificationTree, x);
trainedClassifier.predictFcn = @(x) treePredict-Fcn(predictorExtractionFcn(x));
trainedClassifier.RequiredVariables = { 'VarName1' , 'VarName2' , 'VarName3' ,
'VarName4' };
trainedClassifier.HowToPredict = sprintf( 'To make predictions on a new table,
T, use: \n yfit = c.predictFcn(T) \nreplacing ''c'' with the name of the varia-
ble that is this struct, eg '' trainedModel''. \n \nThe table, T, must contain
the variables returned by: \n c.RequiredVariables \nVariable formats (eg ma-
trix/vector, datatype) must match the original training data. \nAdditional var-
iables are ignored. \n \nFor more information, see <a
href="matlab:helpview(fullfile(docroot, ''stats'', ''stats.map''), ''ap-pclas-
sification_exportmodeltoworkspace'')">How to predict using an exported model<
/a>.' );
inputTable = trainingData;
predictorNames = { 'VarName1' , 'VarName2' , 'VarName3' , 'VarName4' };
predictors = inputTable(:, predictorNames);
response = inputTable.VarName5;
isCategoricalPredictor = [false, false, false, false];
partitionedModel = crossval(trainedClassifier. ClassificationTree, 'KFold' ,
5);
validationAccuracy = 1 - kfoldLoss(partitionedModel, 'LossFun' , 'Clas-sifEr-
ror' );
C = confusionmat(group,validationPredictions, 'Order' ,ClassNames);
stats = statsOfMeasure(C);
MICRO_F1_SCORE(i,j,q) = stats.microAVG(end);
now_num = now_num+1;
mystr=[ 'calculating...' ,num2str(100*now_num/TOTAL_NUM), '%' ];
waitbar(now_num/TOTAL_NUM, mywaitbar, mystr);
        end
    end
end
close ; figure(8) ;
plot(x1,y1, 'r' , 'Linewidth' ,1.5);hold on
plot(x2,y2, 'g' , 'Linewidth' ,1.5);plot(x3,y3, 'b' , 'Linewidth' ,1.5);
string1 = [ 'gdi' , ', AUC = ' , num2str(round(auc1,3))];
string2 = [ 'twoing' , ', AUC = ' , num2str(round(auc2,3))];
string3 = [ 'deviance' , ', AUC = ' , num2str(round(auc3,3))];
legend(string1,string2,string3, 'Location' , 'Best' ) ;
xlabel( 'FPR' );ylabel( 'TPR' );
```