The 2028 Olympic Medal Prediction Model Based on Entropy Weight – TOPSIS

In the global sports landscape, the prediction of Olympic medal table has become a key research hotspot in the field of sports, which has far-reaching significance for the strategic planning of sports development, the rational allocation of resources and the evaluation of the influence of sports events. This study focuses on the 2028 Summer Olympic Games in Los Angeles, USA, aiming to build an accurate and effective medal counting model, and deeply analyze the changing trend of the medal table ranking and related influencing factors.

This study comprehensively uses the information entropy weighting method and TOPSIS method to construct two medal prediction models. In the process of constructing the information entropy weighting model, the ID3 algorithm and the entropy weight method are innovatively integrated to accurately determine the weight of each influencing factor, and the athlete's status value and the standard project value are set as the core independent variables. After a large amount of data calculation and multiple rounds of model testing, the model can better reflect the actual situation. Although there is a certain error (imprecision of 0.19), it still has high reference value. According to this model, in the 2028 Olympic Games, China, Japan, South Korea, Australia and other countries may decline in the medal table rankings, while Britain, France, the Netherlands, Germany and other countries are expected to improve. At the same time, through in-depth analysis, the important sports of each country are clarified, which provides a strong basis for countries to optimize the layout of the event and enhance the competitiveness of sports.

The TOPSIS method prediction model takes another way, selects the number of participants in 2024 and the number of participating in the Olympic Games since 2000 as independent variables, and predicts that MLI, GUI, ANG and other countries have a high probability of winning MEDALS for the first time in the next Olympic Games through scientific calculation of relative proximity. Among them, the probability of the first three countries winning the first medal is nearly 100%, which provides a new perspective for exploring the diversity and dynamic changes of the Olympic Games medal distribution.

In addition, a paired sample t-test was used to deeply analyze the "great coach" effect. The results showed that after Lang Ping coached the US women's volleyball team, Napieva coached the Chinese women's gymnastics team, and Bela Karolyi coached the US women's gymnastics team, the competitive status of the athletes was significantly improved, which strongly proved the key role of "great coaches" in the improvement of the team's performance and the positive contribution to the growth of the overall medal number.

The model constructed in this study shows the outstanding advantages of innovation, objectivity and intuitionistic, and performs well in data processing and prediction. However, it is also found that the weight of the TOPSIS method is difficult to estimate accurately, and the entropy weight method is susceptible to the interference of outliers. Looking forward to the future, follow-up studies can consider incorporating more evaluation factors such as the level of national economic development and the differences in athletes' living habits, and introduce more advanced and complex mathematical models such as neural network models to further improve the accuracy and comprehensis of medal table prediction. At the same time, the ideas and methods of model construction in this study have wide universality. After expansion and optimization, it is expected to be applied to weather prediction, stock market analysis, employment situation prediction and other fields, providing innovative ideas and effective methods for prediction research in different fields.

Keyword**:** Prediction of Olympic medal table; Information entropy weighted model; TOPSIS method; Great coach effect; Sensitivity analysis

Contents

1.Introduce

1.1Problem Background

In recent years, the global attention to the Olympic Games has been continuously rising, and its hosting has far-reaching impacts in politics, economy, society and culture. To ensure the smooth running of the Games, the host city will adjust policies, budgets and resource allocation, improve infrastructure such as transportation, accommodation and sports facilities, create a large number of job opportunities, and enhance the city's level and promote national fitness after the event. The Olympic Games can also attract tourists, boost tourism and investment, enhance the international status and influence of the host country, and promote international cultural exchanges.

Fans pay close attention to Olympic events and the medal table. Some small countries have also achieved a breakthrough in Olympic medals, making the Olympic Games full of infinite possibilities and making the prediction of Olympic results and medal tables a hot topic.

As a global sports event, Olympic results and medals reflect a country's competitive strength and willpower. Countries attach great importance to the prediction of results, and related research has been heating up. Various prediction models are constantly developing and improving.

The aim of prediction science is to reduce the decision-making risks brought by uncertainty. It has developed rapidly worldwide over the past 40 years. Scientific sports prediction is very necessary for the development of Chinese sports. It can improve the foresight of work, promote scientific training, provide scientific basis for decision-making, help coaches take the initiative in competitions, and provide key data and information when formulating sports plans and work plans.

ranking	Country	gold medal	silver medal	bronze	total
1	United States	40	44	42	126
2	China	40	27	24	91
3	Japan	20	12	13	45
4	Australia	18	19	16	53
5	France	16	26	22	64
6	Netherlands	15	7	12	34
7	United Kingdom	14	22	29	65
8	Korea	13	9	10	32
9	Italy	12	13	15	40
10	Germany	12	13	8	33
11	New Zealand	10	7	3	20
12	Canada	9	7	11	27
13	Uzbekistan	8	2	3	13
14	Hungary	6	7	6	19
15	Spain	5	4	9	18
16	Sweden	4	4	3	11

Figure 1 The medal table for the 2024 Olympic Games

1.2Restatement of the Problem

Given the background information and constraints identified in the restatement of the problem, we need to address the following questions:

Develop a medal count model for each country (gold and total MEDALS) and a projection of the medal tally for the 2028 Summer Olympics in Los Angeles, USA, saying whether certain countries will improve or decline and why.

Predict how many countries will win their first medal at the next Olympics and the probability, what sports each country is strong in, and whether the selection of the host country will cause any fluctuations in the medal count for each country.

Because the coach has the initiative to choose the coach's team, there will be a "great coach" to coach other national teams, which will lead to changes in the performance of the national teams in the Olympic Games. Combined with the relevant data, it is analyzed whether there is a "great coach" effect.

1.3 Literature Review

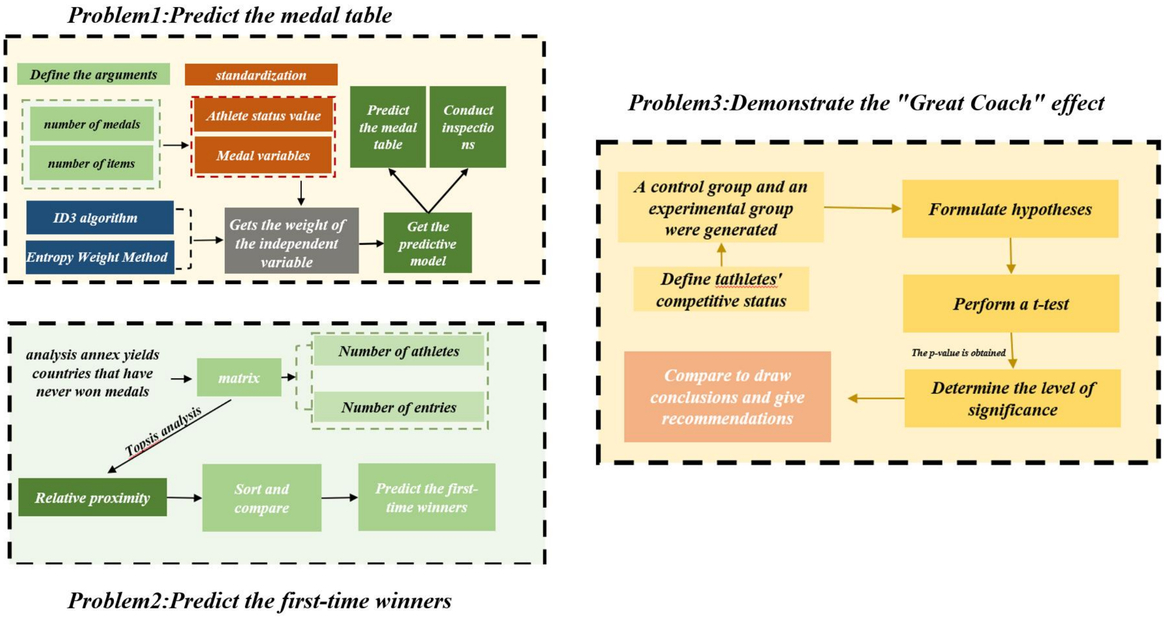
Muller, M., Gollner, B., & Tschernutter, C. In 2020, International Journal of Sports Science & Coaching,[1] "Predicting Olympic Success: "A Machine Learning Approach" features national historical performance, athletes' personal records and background information, the intensity of competition in the sports, and the resources invested (such as training facilities, financial support). Linear regression, decision tree, random forest, support vector machine (SVM), neural network and other models are used to conduct cross-validation. This paper discusses how to accurately predict the number and category of MEDALS (such as gold, silver, bronze) for each country in the Olympic Games, and what factors can affect the success rate of athletes and countries.

Julia Bredtmann, Carsten J. Crede, and Sebastian Otten published in Significance, Issue 3, 2016, "Olympic medals: does the past predict the future?" Taking the total medal count of the past four Olympic[2] Games, sports participation in each country, and government investment and support for sports as important features, the model is evaluated by regression analysis, time series analysis, classification model, random forest and other methods such as cross-validation to study whether unequal resource allocation between different countries leads to differences in the chances of winning MEDALS.

Christoph Schlembach, Sascha L. Schmidt, Dominik Schreyer, and Linus Wunderlich in Technological Forecasting and Social "Forecasting the Olympic medal distribution - A socioeconomic machine learning model[3]" in Change, Volume 175, 2022 to GDP (GDP), per capita income, population size, education level as important features, through a variety of machine learning algorithms to build forecasting models, Cross-validation and other statistical evaluation methods (such as mean square error, R-squared value) were used to evaluate the performance of the model. Socioeconomic factors such as the economic strength of the country, the investment in sports facilities, and the proportion of the population participating in sports activities had a significant impact on medal allocation.

1.4 Our Work

Considering the background of the problem and the previous research results, we analyzed different problems:



2 Assumptions and Justifications

Assumptions related to athlete status: When constructing the information entropy weighted prediction model, it was assumed that the athlete status was only affected by the performance of the past Olympic Games, and the influence of the innovation of training methods, changes in diet and nutrition, and fluctuations in psychological factors on the current competitive status of athletes was ignored. At the same time, it is assumed that the state of athletes in different sports is independent of each other. For example, the state change of swimming athletes will not interfere with the state of track and field athletes, and the transmission effect of cross-sport training or event experience exchange on athlete state is not considered.

The **change hypothesis of the number of** events: The change trend of the number of events in the Olympic Games is assumed to be not affected by major adjustments in sports organization policies, changes in global sports culture, and the rise of new sports. The change is only based on the rule described by the Logistic population model. It is believed that the adjustment of the number of sports in the future Olympic Games will be smooth and predictable, and there will be no sudden increase or decrease of large-scale sports or subversive change of rules.

Data Representative hypothesis: This study uses the data of the Olympic Games since 2000, and assumes that these data can comprehensively and accurately reflect the development trend of sports in various countries and the change law of athletes' competitive level. The impact of missing data, recording errors or special years (such as the special circumstances of the host country affecting the fairness of the event) on the data quality is ignored, and it is regarded as a fully representative sample for the construction and validation of the model.

Weight assumption of TOPSIS method: In the prediction model of TOPSIS method, it is assumed that the influence weight of the number of entries and the number of participants on the probability of winning MEDALS is equal, which is 0.5. Because in the countries that often participate in the Olympic Games and often win MEDALS, there are a large number of athletes who are never absent, and it is impossible to quantify the weight, and through daily experience, the two have little difference in the degree of influence on the results.

"Great Coach" effect hypothesis: When studying the "great coach" effect, it is assumed that the coach has a direct and unique influence on the athletic state of the athletes, and the synergistic or countervative effects of other coaches, training team members, or changes in the effort level of the athletes themselves on performance during the same period are not considered. At the same time, it is assumed that the coaching effect of "great coach" is consistent in different countries and different programs, and the influence of cultural differences and sports foundation differences on coaching results is not considered.

3 Notations

Symbol	Description
G	Number of gold MEDALS (with duplicates)
S	Number of silver MEDALS (with duplicates)
B	Number of bronze MEDALS (with duplicates)
SUM	Total number of games
p_i	Medal win percentage
X	Athlete status value
E	Total number of projects
M_i	Number of MEDALS
d	Index of variation
e	Information entropy
m_G	Medal variables
P	Standard item ratio
K	Maximum ambient capacity
r	Natural growth rate
A	Residual growth space ratio
PI.	imprecision
d	Distance
C	Relative approximation

Some of the symbols in the table above illustrate: One event can have multiple athletes, and each athlete can also participate in multiple competitions, so the total number of games here does not refer to the total number of events, nor the number of athletes. Suppose N events, on average, n athletes participate in each event, SUM = N × n. Because athletes participate in teams, and a champion only corresponds to a medal in the medal table, G, S and B represent the number of MEDALS in this repeated situation, and represent the number of MEDALS recorded in the medal table.M_G

4. Information entropy weighted prediction model

4.1 Data analysis and establishment of prediction model

4.1.1 Discussion of independent variables affecting the number of MEDALS

The number of Olympic MEDALS is generally not referred to the total number of historical MEDALS, so the number of historical MEDALS is not considered, and it is used as the dependent variable of the function to test the model. In most cases, the current status of the athletes can directly affect the final results of the number of MEDALS. To analyze the performance of the athletes of a certain country, we refer to the historical performance of the athletes and carry out regression analysis based on the performance of a country in the previous Olympic Games to reflect the state of the athletes of that country. As can be seen from the attachment, there are athletes participating in every Olympic Games. Most of the athletes did not win MEDALS, while a small number of athletes won gold MEDALS, silver MEDALS and bronze MEDALS. The gold medal winning rate was obtained by dividing the number of gold MEDALS won by the total number of events. Since the status of the three kinds of MEDALS was different, the gold medal, silver medal and bronze medal winning rate were weighted and added, and the weights of 3,2 and 1 were assigned respectively to obtain a new variable. We define as the athlete status value X, here we take the United States as an example to study, where

$$p_1 = \frac{G}{SUM}, \ p_2 = \frac{S}{SUM}, p_3 = \frac{B}{SUM} \#(4.1)$$

$$X = 3p_1 + 2p_2 + p_3 \#(4.2)$$

Where the value is between 0 and 1, X can directly affect the final medal count.

In addition, the more sports there are in the Olympic Games, the more MEDALS will be distributed throughout the Games, and the higher the probability that each country will win a medal. As you can see from the image:

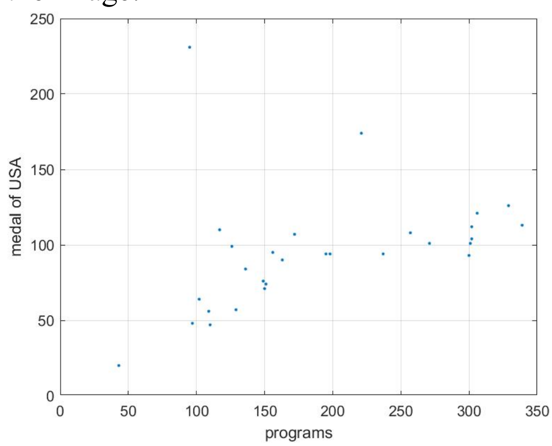


Figure 2 The number of sports versus the total number of MEDALS for the United States

Obviously, the total number of Olympic Games will also affect the number of MEDALS for a country, and the total number of Olympic Games directly determines the total number of Olympic Games MEDALS, but the contribution degree of the two to the number of MEDALS is obviously different. The total number of events E and X are too different, so E should be standardized. When E=300, P=1 and

$$P = \frac{E}{300} \#(4.3)$$

Therefore, the medal variable can be defined as follows:m_G

$$m_G = k_1 X + k_2 E = k_1 (3\frac{G}{SUM} + 2\frac{S}{SUM} + \frac{B}{SUM}) + k_2 P \#(4.4)$$

4.1.2 Selection of algorithm and overview of model

Decision tree model ID3[4], as a greedy algorithm, plays the role of constructing decision trees. ID3 algorithm originates from Concept Learning System (CLS). At each node, it selects the attribute with the highest information gain as the partition criterion, and then continues this process until the generated decision tree can perfectly classify the training examples.

The entropy weight method[5] is used as the weight determination method. Based on the information entropy, the weight of each index is determined by calculating the entropy value of each index and comparing the degree of difference between the indexes. The entropy value is inversely correlated with the difference between the indicators, which can show that the index is more important in decision-making.

In this paper, the two methods are combined. Firstly, the algorithm of information entropy and information gain in ID3 algorithm is used to obtain the disorder degree of decision attribute and the influence degree of each attribute on decision attribute. This method can combine the advantages of the two methods to make the weight determination more reasonable and comprehensive.

We divide the athlete state X into three parts. Taking the United States as an example, the state is set as follows. The total number of Olympic events E is divided into three levels: more, medium and less. The number of gold MEDALS won by the United States is divided into three levels: more, medium and less. The proportion of each level is calculated and the information gain of E and X is calculated by ID3 algorithm.$M_G M_G$

About the specific use of the idea of entropy weight method: the process of entropy weight method to standardize the sample index is to control the value of the described sample in [0,1], and use it to calculate the entropy value, and the scope of the information entropy algorithm is, this paper uses the entropy weight method to calculate the weight of the information entropy, and in the entropy weight method to calculate the variation index of the entropy value: $[0, \log_2 3\,]$d= 1-E, where "1" is the maximum value of the value of the sample after standardization, then when we calculate the variation index of the information entropy, the formula should be:

$$d = \log_2 3 - e\#\#(4.5)$$

The corresponding weights k1 and k2 are calculated from the formula of weight obtained by processing the variation index in the entropy weight method:

$$k_i = \frac{d_i}{\sum_{i=1}^{n} d_i} \#(4.6)$$

Finally, the formula for the total number of gold MEDALS in the United States is obtained:

$$m_G = \frac{\log_2 3 - e_1}{\sum_{i=1}^{n} d_i} \times (3\frac{G}{SUM} + 2\frac{S}{SUM} + \frac{B}{SUM}) + \frac{\log_2 3 - e_2}{\sum_{i=1}^{n} d_i} \times P\#(4.7)$$

4.2 Solving the model

4.2.1 Calculation of specific parameters

Because of the rapid development of all countries in the world, the results calculated by the data of the Olympic Games since 2000 have more reference value for today. Through the screening and calculation of the attachment, we obtain the table:

Table 1 Calculation of athlete status in the United States in recent years

Year	SUM	Gold	Silver	Bronze	X
2000	776	130	61	51	0.725515
2004	735	117	75	71	0.778231
2008	768	127	110	80	0.886719
2012	698	145	57	46	0.852436
2016	726	139	54	71	0.820937
2020	856	113	110	75	0.740654
2024	854	131	96	94	0.795082

Use the United States itself as a reference to classify athlete status, specifying that an athlete status value between 0.75 and 0.8 is fair, greater than 0.8 is good, and less than 0.75 is bad. By observing the number of Olympic events since 2000, the number of stipulated events between 310 and 320 is medium, less than 310 is less, more than 320 is more, and the following table is obtained. Similarly, the number of stipulated gold MEDALS between 37 and 45 is medium, less than 37 is less, and more than 45 is more:

Table 2 Decision attribute Table

Year	X	E	M_G
2000	Bad	less	General
2004	General	less	less
2008	Good	less	less
2012	Good	less	more
2016	Good	less	more
2020	Bad	more	General
2024	General	more	General

According to the above table, the calculation results of information gain and information entropy can be calculated:

Table 3 takes the gold medal of the United States as an example, and the data results of all indicators

Metrics	Results obtained
Unconditional entropy	1.5566567074628228
Information gain and entropy of athlete states	(1.138724635395457, 0.27709636376654934)
Information gain and information entropy of the total amount of the project	(0.18150882310232208, 1.2343121760596842)
k_1, k	(0.6756010197686305, 0.32439898023136954)

The results are as follows:

$$m_G = 0.6756 \times (3\frac{G}{SUM} + 2\frac{S}{SUM} + \frac{B}{SUM}) + 0.3244 \times P\#(4.8)$$

The medal constant is defined based on the number of gold MEDALS in the United States in $2000 Q_G$

$$m_G \times Q_G = M_G\#(4.9)$$

$$m_G \times Q_G = 37 \rightarrow Q_G = 45.4\#(4.10)$$

$$M_G = 45.4 \times (0.6756 \times (3\frac{G}{SUM} + 2\frac{S}{SUM} + \frac{B}{SUM}) + 0.3244 \times P)\#(4.11)$$

Plug in the data to find the number of medals.Similarly, we can calculate:

Table 3 Data results of all indicators for silver and bronze MEDALS

Indicators	Silver Medal	Bronze
Unconditional entropy	1.556656707	
k_1, k	0.45730441975063163, 0.5426955802493684	0.826933753115614, 0.17306624688438585
Q	27.4	41.4

$$M_S = 27.4 \times (0.4573 \times (3\frac{G}{SUM} + 2\frac{S}{SUM} + \frac{B}{SUM}) + 0.5427 \times P)\#(4.12)$$

$$M_B = 41.4 \times (0.8270 \times (3\frac{G}{SUM} + 2\frac{S}{SUM} + \frac{B}{SUM}) + 0.1731 \times P)\#(4.13)$$

4.2.2 Predictions for the medal table in 2028.

The two main independent variables in equations (4.11), (4.12) and (4.13) are the athlete status value X and the standard event value P. We can reflect the P value by two-dimensional image analysis of the total number of events E and the year:

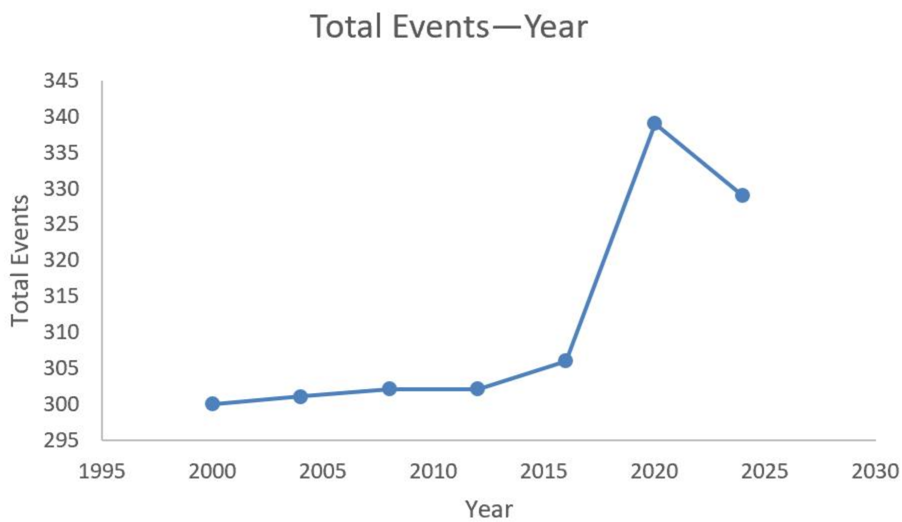


Figure 3 Relationship between total number of events and year

As shown in the figure, starting from the year 2000, the first segment of the image is similar to the exponential image, and the second half begins to decline but is relatively gentle. Such images can be approximately described by the Logistic population model[8], and the expression of the Logistic population model is:

$$E-300=\frac{K-300}{1+Ae^{-r(t-2000)}}, \quad A=\frac{K-E(0)}{E(0)} \#(4.14)$$

It is reasonable to set K=350 and E(0)=300 for the observation image, then A=0.17, and r=0.0105 can be calculated by substituting the value (2016,306) into equation (4.14). namely$E=\frac{50}{1+0.17e^{-0.0105\times(t-2000)}}+300\#(4.15)$

Substituting t=2028 into equation (15), E=344, and P=1.1467.

Next, take the top ten MEDALS table in 2024 as the sample to calculate the athlete status value of each country:

Table 4 Athlete status values for each country

Year / Country	2000	2004	2008	2012	2016	2020	2024
United States	0.725515	0.778231	0.886719	0.852436	0.820937	0.740654	0.795082
China	0.380368	0.389068	0.473236	0.524345	0.427562	0.460317	0.543296
Japan	0.23416	0.374396	0.246696	0.35589	0.254587	0.39136	0.279264
Australia	0.472081	0.552413	0.446208	0.363813	0.312741	0.362069	0.355932
France	0.297872	0.227766	0.339326	0.403756	0.373047	0.581784	0.484395
Netherlands	0.552901	0.509363	0.53169	0.630631	0.273556	0.409207	0.616927
Britain	0.284337	0.322222	0.402878	0.368421	0.686192	0.485401	0.515947
South Korea	0.342618	0.313953	0.519757	0.352564	0.209125	0.215976	0.45968
Italy	0.269147	0.407173	0.156863	0.319372	0.320802	0.26145	0.278689
Germany	0.358025	0.466216	0.342342	0.413725	0.559701	0.249581	0.307573

Using the least squares[7] method, we can predict and plot the athlete status of each country in 2028, taking France as an example:

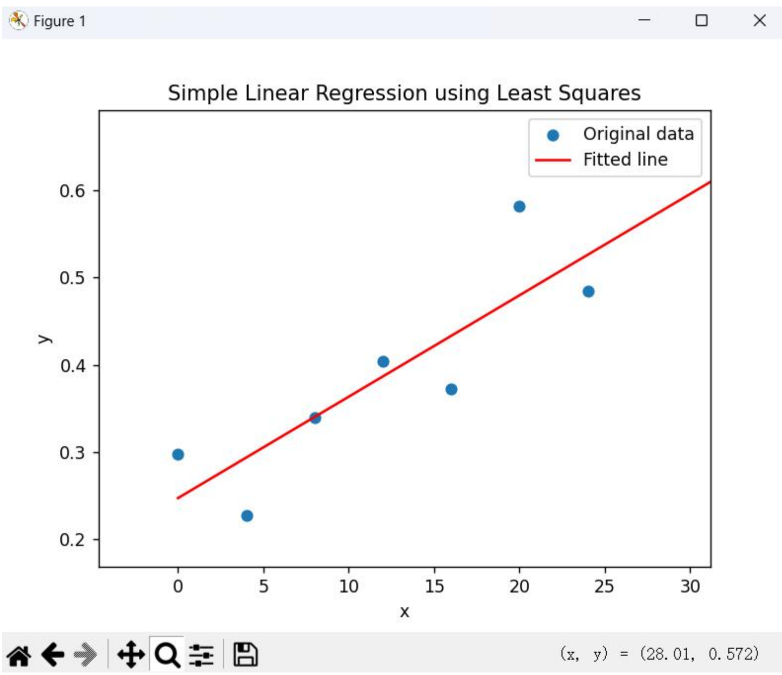


Figure 4 Regression analysis of athlete status values for France

The abscissa of the figure above is the year -2000, and the ordinate is the athlete status value.

After this calculation, the new medal table is obtained:

Table 5 Medal tally forecast for 2028

2024 Rankings	Projected 2028 Rankings	M_G	M_S	M_B
United States	United States	42	27	37
China	Britain	36	24	31
Japan	France	34	24	29
Australia	China	33	23	27
France	Netherlands	31	22	25
Netherlands	Germany	27	21	20
Britain	Japan	27	21	20
South Korea	Australia	26	21	19
Italy	Italy	25	20	18
Germany	South Korea	22	19	15

A comparison of the new medal tally with the old one shows that China, Japan, South Korea and Australia will fall back in 2028, while Britain, France, the Netherlands and Germany will improve.

4.2.3 Evaluation of the importance of each country's events

For the athlete state value X defined by us, athletes in different countries have different performances in different kinds of sports. That is to say, by discussing the athlete state value of one kind of sports separately and comparing the X value of each sport in a country, the sport corresponding to the largest athlete state value is the most important sport in a country. Analyzing the data, here is a list of the most important sports for a subset of countries:

Table 6 table of events that each country excels in

Countries	Project	X
United States	Swimming	1.244094
China	Swimming	0.57732
Japan	Judo	1.5
Australia	Swimming	0.87395
France	Judo	1.896552
Netherlands	Hockey	3
Britain	Rowing	1.714286
South Korea	Archery	2.357143
Italy	Volleyball	1.5
Germany	Athletics	0.04

If a host country wants to improve its medal ranking, hosting more events it is good at will bring other countries that are good at these events up with it, while countries that are not good at these events will cause its ranking to fall.

4.3 Reliability Test

According to the meaning of unconditional entropy, the number of decision attributes selected by the model is 3, so the maximum value of unconditional entropy is according to the formula. Compared with the maximum value of unconditional entropy, the ratio is about 98.2%, which shows that the random sampling results of the established model are successful, which can better reflect the actual situation and contain more information. $\log_2 3 \approx 1.5850$The information gain of the condition attribute is the reflection of the degree of influence on the decision attribute, and the information entropy of the condition attribute is the reflection of its irrelevance to the decision attribute. It is observed that the contribution of athlete status value and the total number of events to the decision attribute is obvious, because their information gain values are relatively high, which means that they are important classification features.

The model of medal quantity is tested on the sample of China since 2000: The following table is obtained by calculation:

Table 7 Actual indicators of China

Year	X	M_G	M_S	M_B	E
2000	0.380368	28	16	14	300
2004	0.389068	32	17	14	301
2008	0.473236	48	22	30	302
2012	0.524345	39	31	22	302
2016	0.427562	26	18	26	306
2020	0.460317	38	32	19	339
2024	0.543296	40	27	24	329

Substitute X into the model and obtain the result:

Table 8 Predicted results of China's MEDALS

M_G	M_S	M_B
26	20	20
27	20	21
30	21	23
31	22	25
29	21	22
29	23	24
33	23	26

Comparing the two tables, it can be seen that the model has certain reference value for the prediction of this sample, but there are still some inaccuracies. We take the quotient of the difference between the predicted value and the actual value and the actual value to reflect the inaccuracy of the result

$$\pi = \frac{\sum_{i=1}^{21}(x_i - x)}{21} = 0.19\#(4.16)$$

The error of the model is small, which has a good reference value, but there is a small deviation from the actual situation. This is not considering the situation of data outside the attachment given by the title. When other sample data are obtained, the model can be improved to achieve smaller error.

5 TOPSIS method prediction model

5.1 Data analysis and establishment of prediction model

5.1.1 Discussion of independent variables and selection of algorithms

Observe the table, in the top of the Olympic Games medal list are the strong comprehensive strength of the big countries, not only are regular visitors to the Olympic Games, there are many participants, more participants, naturally have greater competitiveness in the competition, participate in the Olympic Games more times, The figure below shows the relationship between the number of participants and the number of medals in each country

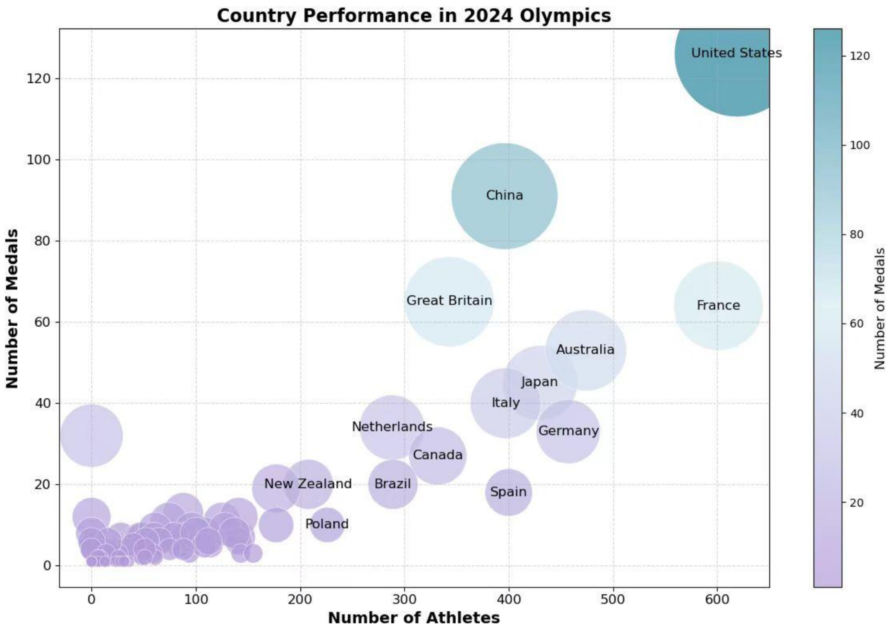


Figure 5 relationship between the number of participants and the number of medals in each country

but also more familiar with the rules, will also improve the medal rate, so, Therefore, the number of participants in 2024 and the number of participating in the Olympic Games since 2000 of each country are used as independent variables to establish a prediction model for 2028 through the TOPSIS comprehensive evaluation method[6].

Firstly, countries that have never won a medal are selected from the table, and their number of participants in 2024 and the number of participants since 2000 are counted. There are 68 countries that have not won a medal and are participating in the 2024 Olympics. A 68*2 matrix can be created by combining the number of participants and the number of games played. Here's a small list:

Table 9 Number of Olympic Games participated by countries and the number of athletes participating

NOC	Number of appearances	Number of participants
LIE	19	1

MAD	14	7
MAW	12	3
MDV	10	5
MHL	5	4
MLI	15	24
MLT	18	5
MTN	11	2
MYA	19	2
NCA	14	7
NEP	15	7
NRU	8	1

5.2 Overview and solution of the model

The basic principle of TOPSIS method is to construct the positive ideal solution and negative ideal solution of the decision problem, that is, the optimal solution and the worst solution of each attribute index, and then calculate the distance between each alternative and the positive ideal solution and the negative ideal solution, and determine the order of the pros and cons of each alternative by comparing the size of these distances. The closer the solution is to the positive ideal solution and the farther the solution is to the negative ideal solution, the higher the relative superiority is. The following is the calculation procedure

Positive transform the data

$$x_{ij}^{'} = \frac{1}{x_{ij}} \#(5.1)$$

Re-normalize the data

$$\tilde{x}_{ij} = \frac{x_{ij}^{'}}{\sqrt{\sum_{i=1}^{m} (x_{ij}^{'})^2}} \#(5.2)$$

For the case that countries that often win MEDALS are almost absent in the Olympic Games, and these countries that often win MEDALS will have a large number of athletes to participate in each time, it is difficult to measure the contribution value of these two indicators, so it is assumed that the contribution value of the number of entries and the number of participants is the same, that is, the weight is 0.5, the ideal solution is the maximum value, and the negative ideal solution is the minimum value, then there is a distance d$max_j min_j$

$$d^{+} = \sqrt{\sum_{j=1}^{m} (0.5\tilde{x}_{ij} - max_j)^2} = \sqrt{\sum_{j=1}^{m} \left(\frac{0.5}{x_{ij}\sqrt{\sum_{i=1}^{n} \left(\frac{1}{x_{ij}}\right)^2}} - max_j\right)^2} \#(5.3)$$

$$d^{-}=\sqrt{\sum_{j=1}^{m}(0.5\tilde{x}_{ij}-min_j)^2}=\sqrt{\sum_{j=1}^{m}\left(\frac{0.5}{x_{ij}\sqrt{\sum_{i=1}^{n}\left(\frac{1}{x_{ij}}\right)^2}}-min_j\right)^2}\#(5.4)$$

It can be concluded that the relative closeness C=, comparing the size of the relative approximation can get the most likely to get the medal of the country.$\frac{d^{-}}{d^{+}+d^{-}}$

5.3 Solving the model

The relative approximation of each country is calculated. Here are the top ten relative approximation countries:

Table 10 Relative approximation for each country

NOC	MLI	GUI	ANG	SAM	LBR	ESA	NEP	MAD	NCA	GAM
C	0.001869	0.001863	0.001855	0.001854	0.001848	0.001844	0.001838	0.001835	0.001835	0.001828

By analyzing the annex, we can get that in recent decades, the minimum number of countries winning MEDALS for the first time in the Olympic Games is 8, and usually more than 10. The probability of the ten countries in the above table winning MEDALS for the first time in the next Olympic Games is very high, and the probability of the first three countries winning MEDALS can be almost 100%.

6 Star coach discussion on the influence of the number of MEDALS

6.1 Data analysis and algorithm selection

In the Olympic Games, the role of coaches is not only reflected in technical guidance and tactical development, but also in stimulating the potential of athletes, improving team cohesion, and optimizing training plans. Unlike athletes, coaches are easier to transfer between countries, which opens the possibility of a "great coach" effect. In most cases, the current competitive state of athletes can directly affect the final results of the number of MEDALS, and how to analyze the influence of star coaches on athletes, we refer to the historical performance of athletes, and use the performance of a country before and after the hiring of star coaches to reflect the overall competitive state of the country's athletes. Based on the attachment information, we can draw Figure 6,7

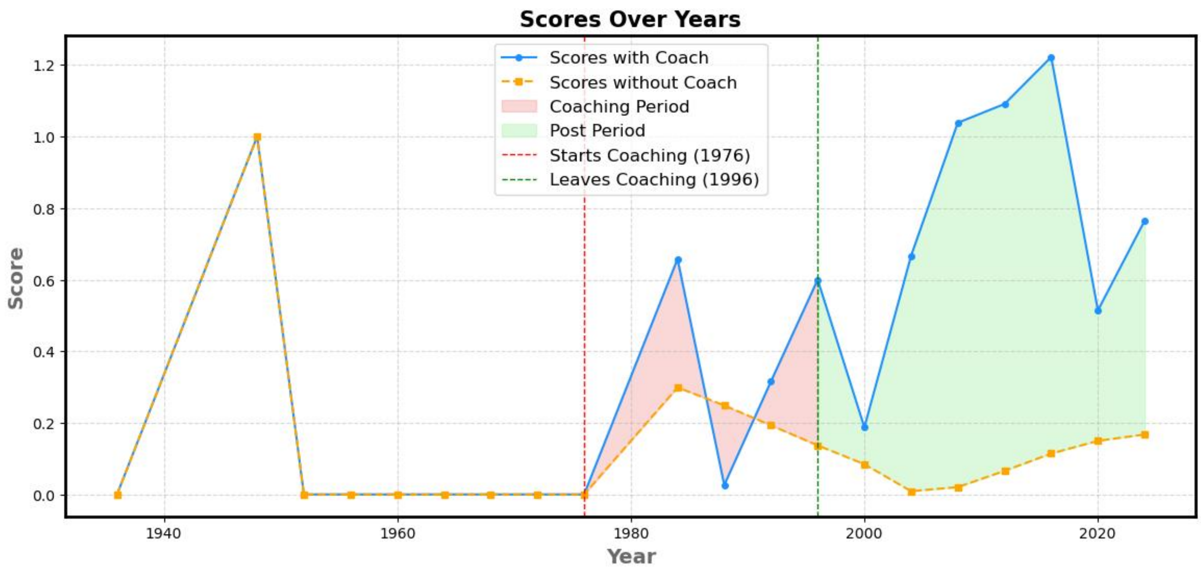


Figure 6 U.S. performance is influenced by great coaches

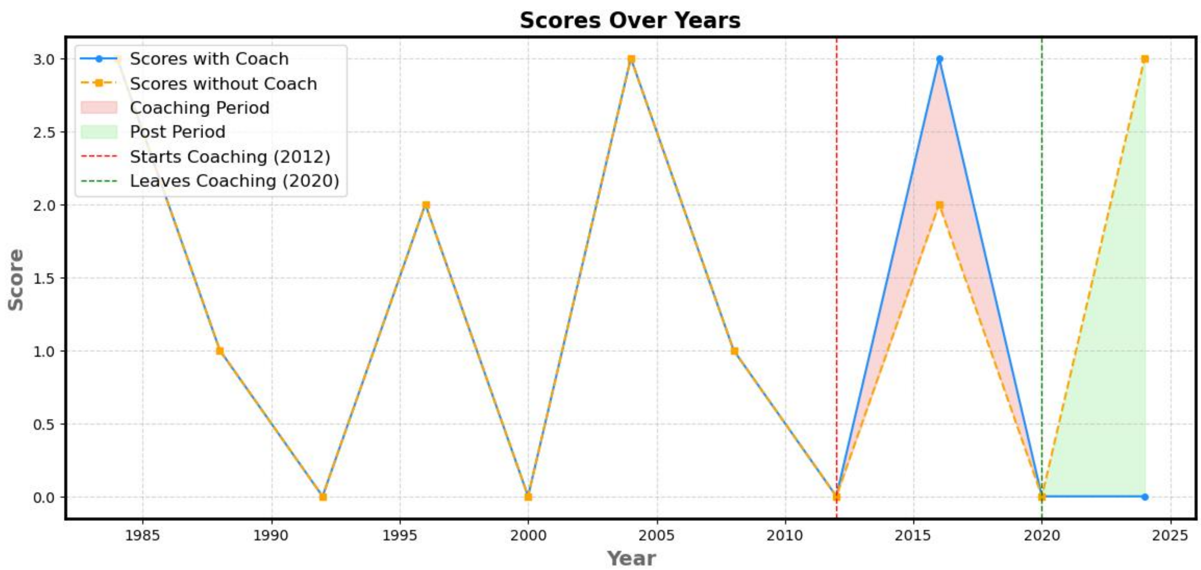


Figure 7 China's performance is influenced by great coaches

as can be seen from Figure 6,7 and annex, in some Olympic Games and major competition events, some countries will shine on the events they were not good at and win gold MEDALS, silver MEDALS and bronze MEDALS. In addition, for the four cases of bronze MEDALS, silver MEDALS, gold MEDALS and not winning MEDALS, we set the value of athletes' competitive status y_i based on the winning situation in the first game of the athletes, where the gold medal is won, The value is 3 when winning the gold medal, 2 when winning the silver medal and bronze medal, and 1 when not winning the medal.

Because the status values were compared across multiple games, we $Y = \{y_i | \mathrm{i} = 1, \ldots, n\}$calculated the difference between each pair of observations using a paired sample t-test, and then tested whether the mean of these differences was significantly greater than zero.

Selection and overview of test methods

Paired Sample t-test is a common statistical method, which can be used to compare the effects of different treatment conditions, and can also be used to compare the differences between different individuals in the same treatment condition to compare whether there is a significant difference in the mean of two groups of related samples.

In this article, the data is collected first, and the data of each paired sample is organized into two columns, one column is the value before the treatment and the other is the value after the treatment, or one column is the value of the experimental group and the other is the value of the control group.

6.2 Formulating Hypotheses

Null hypothesis: It is usually assumed that the means of two paired samples are not significantly different, that is, where is the population mean of the difference between the two paired samples.$H_0: \mu_d = 0 \mu_d$

Alternative Hypothesis: Depending on the purpose of the study, the one-tailed alternative hypothesis or the two-tailed alternative hypothesis may be chosen. The one-tailed alternative hypothesis is often used to test whether a treatment is effective-for example, whether the mean of the experimental group is greater than the mean of the control group; The two-sided alternative hypothesis is often used to test whether two group means differ regardless of the direction of the difference.

Calculate the test statistic

Calculate the difference: Calculate the difference for each pair of samples, where and are the two observations for the second pair, respectively. $d_i = X_{1i} - X_{2i} X_{1i} X_{2i} i$

Calculate the mean and standard deviation: Calculate the sample mean and sample standard deviation of the difference as follows:$\bar{d} s_d$

$$\bar{d} = \frac{1}{n} \sum_{i=1}^{n} d_i,\ \ s_d = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (d_i - \bar{d})^2} \# (6.1)$$

Where, is the number of paired samples.n

Calculate the test statistic: Calculate the statistic t of the paired sample t-test according to the formula:

$$t = \frac{\bar{d}}{s_d / \sqrt{n}} \# (6.2)$$

Determine significance level and degrees of freedom

significance level: The significance level is usually chosen; a common value is 0.05 or 0.01. αThe significance level indicates the maximum probability that you are allowed to reject the null hypothesis.

Degrees of freedom: The degrees of freedom for a paired samples t-test are, where is the number of paired samples. $df = n - 1 n$

Make a decision to

find the critical value: Based on the significance level and degrees of freedom, look up the t-distribution table or use statistical software to find the critical value.$\alpha df t_{\alpha/2}(df)$

Compare statistics and critical values: Compare the calculated test statistics to the critical values.t $t_{\alpha/2}(df)$

6.3 Making Decisions

If, reject the null hypothesis that the means of the two paired samples are significantly different.$|t| > t_{\alpha/2}(df)$

If, the null hypothesis is not rejected and the means of the two paired samples are not significantly different.$|t| \leq t_{\alpha/2}(df)$

The competitive status values of the US women's volleyball players before and after Lang Ping coached the US women's volleyball team were respectively

$$Y_1 = \{1,1,1,2,2,1,2,1\}$$
$$Y_2 = \{2,2,2,2,3,2,3,2\}$$

The p value obtained by t test is 0.0011, which rejects the null hypothesis and shows a significant difference.

The competitive status values of Chinese gymnastics team athletes before and after Nabieva coached the Chinese gymnastics team were respectively

$$Y_1 = \{1,1,1,1,1,1,1,1,1\}$$
$$Y_2 = \{2,3,2,3,2,2,2,2,2\}$$

p value of t test: 0.0000, the result of rejecting the null hypothesis, there is a significant difference.

For Bela Karolyi before and after coaching the US women's gymnastics team, the competitive status values of the US women's gymnastics team athletes were

$$Y_1 = \{1,1,1,1,1,1\}$$
$$Y_2 = \{3,2,2,2,3,3\}$$

The p value of t test was 0.0002, which rejected the null hypothesis and showed a significant difference.

For the national team, the arrival of the star coach will strengthen the strength of the team and make the team's competitive ability stronger in the competition. In order to quantify the strength change brought by the star coach, we let be the value of the team's competitive state, and be the growth rate of the team's competitive state, so as to measure the influence of the star coach on the national team's competitive state.$|Y|\widehat{Y}$

$$|Y| = \sum_{i=1}^{n} y_i \#(6.3)$$

$$\widehat{Y} = \frac{|Y_2| - |Y_1|}{|Y_1|} \times 100\% \#(6.4)$$

Through calculation, it can be concluded that the competitive state value of the American women's volleyball team when Lang Ping $\widehat{Y}100\%$coached it is, the competitive state value of the Chinese women's volleyball team when $\widehat{Y}122.22\%$ Nabieva coached it is, and the competitive state value of the American women's gymnastics team when Bela Karolyi coached it is.$\widehat{Y}150\%$

The improvement of the team by the "great coach" is obvious. The competitive state of the athletes is almost doubled, so the influence of the "great coach" on the single project is doubled, and the influence on the overall medal is almost a direct addition of a medal. All countries should invest in "great coaches". This is a universal situation.

7 Sensitivity Analysis

We use MATLAB to give the influence of different k1 values on the image, here +=1, so it changes with the change of the other side, the stability test of one of them is the stability test of the other, we change the value to make a diagram, the step flow chart is as follows:$k_1 k_2 k_1 k_2 k_1$与$k_2 k_1$

Here the value is (0,0.25,0.5,0.75,1), let a country capacity value is 0.8, the number of items is 300, and several relationship images with M are obtained:$k_1 y_1$， y_2

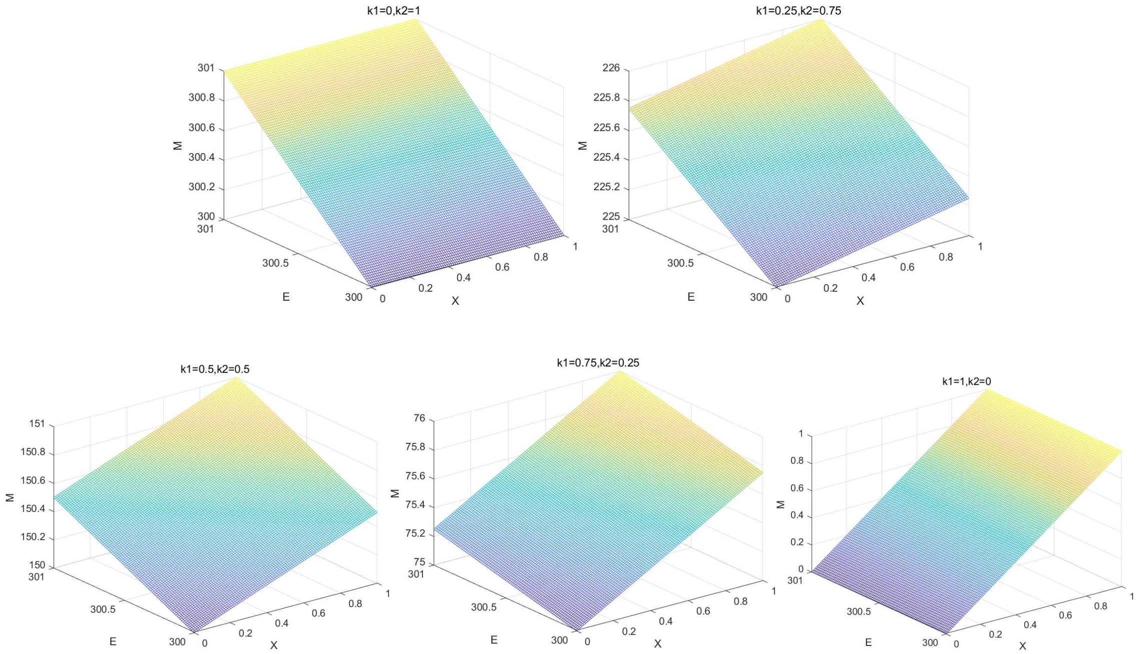


Figure 8 Sensitivity analysis of the model

By analyzing the five images, it is found that in non-extreme cases, the value of does not have a great influence on the value of M. However, when the difference between, and is too large, it still has a great influence on M. Therefore, the model has local stability, but has a certain sensitivity in a large range.$k_1 k_2 k_1 k_2$

8 Model Evaluation and Further Discussion

8.1 Strengths

1. The combination of ID3 algorithm and entropy weight method is innovative, and the results obtained by this calculation are objective and reliable.

2. A lot of data is standardized, and more variables are defined, which not only makes the calculation simple, but also does not affect the final results, and the analysis of the problem is more clear and convenient to describe.

3. Using regression analysis and Logistic population model to process the data, making

the prediction results more credible.

4. Using TOPSIS comprehensive analysis method to process the data and predict the results, the results are intuitive, easy to understand, and can deal with a large number of data.

5. Use t test to verify the "great coach" effect, making the results intuitive.

8.2 Weaknesses

1. The weight cannot be estimated in TOPSIS method, which may bring bias.

2. The idea of entropy weight method is strongly dependent on data and may be disturbed by outliers.

3. The current low number of "great coaches" worldwide may result in an insufficient sample size for the data analysis here.

8.3 Further Discussion

This model can be further improved by considering the introduction of more evaluation factors, such as the level of national economic development and the difference in athletes' living habits, in order to comprehensively assess the medal competition ability of a country and predict the medal table more comprehensively. At the same time, more complex mathematical models, such as neural network models, can be used to more accurately describe the value of athletes' status and the relationship between events and MEDALS. The model in this paper can not only be used in medal prediction, but also has universality. As long as the actual situation can be quantified, such as "athlete status value" and "standard project ratio", the weight can be calculated to obtain the prediction model. After adding enough variables, more accurate predictions can be made in weather prediction, stock market, and employment situation.

Conclusion

Focusing on the prediction of the Olympic medal table, this study constructed the information entropy weighted prediction model and the TOPSIS method prediction model, and analyzed the "great coach" effect, and obtained the following conclusions:

Medal list prediction results: The information entropy weighted prediction model determines the weight by innovatively combining ID3 algorithm and entropy weight method, and uses the athlete state value and standard item value as the key independent variables to predict the medal list of the 2028 Olympic Games. The reliability test of the model shows that although there is a certain error (the inaccuracy is 0.19), it still has good reference value. The prediction results show that in the 2028 Olympic Games, China, Japan, South Korea, Australia and other countries may decline in the medal table rankings, while Britain, France, the Netherlands, Germany and other countries may improve. At the same time, the important sports of each country are identified, which provides a reference for each country to formulate the competition strategy.

First time winning country prediction: The TOPSIS method prediction model takes the number of participants in 2024 and the number of countries participating in the Olympic Games since 2000 as independent variables. By calculating the relative proximity, it is predicted that MLI, GUI, ANG and other countries have a high probability of winning a medal for the first time in the next Olympic Games. Among them, the probability of the first three countries getting the first medal is nearly 100%, which provides a new perspective for studying the diversity of Olympic medal distribution.

Verification of "Great Coach" effect: The paired sample t-test was used to analyze the effect of "great coach". The results showed that after Lang Ping coached the US women's volleyball team, Napiyewa coached the Chinese gymnastics team, and Bela Karolyi coached the US women's gymnastics team, the athletes 'competitive status was significantly improved, which proved that the "great coach" had an obvious effect on improving the team's performance. It also has a positive effect on the increase of the overall medal number.

References

[1] Muller, M., Gollner, B., & Tschernutter, C. (2020). Predicting Olympic success: A machine learning approach. *International Journal of Sports Science & Coaching*, 15(3), 345-359.

[2] Bredtmann, J., Crede, C. J., & Otten, S. (2016). Olympic medals: does the past predict the future? *Significance*, 13(3), 22-25.

[3]Schlembach, C., Schmidt, S. L., Schreyer, D., & Wunderlich, L. (2022). Forecasting the Olympic medal distribution – A socioeconomic machine learning model. *Technological Forecasting and Social Change*, 175, 121-134.

[4]Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81-106.

[5] Zhang, X., & Wang, Y. (2019). An entropy weight method for evaluating the performance of sustainable development goals. *Sustainability*, 11(22), 6265.

[6] Mardani, A., Jusoh, A., Zavadskas, E. K., & Raut, R. D. (2015). Application of TOPSIS method for supplier selection in supply chain management: A case study. *Journal of Business Economics and Management*, 16(6), 1134-1155.

[7] Seber, G. A. F., & Lee, A. J. (2003). *Linear regression analysis* (2nd ed.). Wiley.

[8] Verhulst, P. F. (1838). Notice sur la loi que la population suit dans son accroissement. *Correspondance Mathematique et Physique*, 10, 113-121.

Matrix used in the TOPSIS algorithm:

NOC	Frequency	2024 number of people	C
AND	13	4	0.001799
ANG	11	26	0.001855
ANT	12	5	0.00181
ARU	10	7	0.001817
ASA	10	2	0.001711
BAN	11	5	0.001805

BEN	13	5	0.001814
BHU	11	3	0.001765
BIH	9	6	0.001802
BIZ	14	1	0.001591
BOL	16	4	0.001808
BRU	7	3	0.001732
CAF	12	4	0.001795
CAM	11	3	0.001765
CAY	12	5	0.00181
CGO	14	4	0.001802
CHA	14	3	0.001777
COD	12	6	0.00182
COK	10	2	0.001711
COM	8	5	0.001784
ESA	14	9	0.001844
FSM	7	3	0.001732
GAM	11	8	0.001828
GBS	8	6	0.001794
GEQ	11	3	0.001765
GUI	13	25	0.001863
GUM	10	9	0.001826
HON	13	4	0.001799
ISV	2	5	0.001565
IVB	11	4	0.00179
KIR	6	3	0.001718
LAO	11	4	0.00179
LBA	12	6	0.00182
LBR	14	10	0.001848
LES	13	3	0.001774
LIE	19	1	0.001604
MAD	14	7	0.001835
MAW	12	3	0.00177
MDV	10	5	0.001799
MHL	5	4	0.001722
MLI	15	24	0.001869
MLT	18	5	0.001827
MTN	11	2	0.001716
MYA	19	2	0.001741
NCA	14	7	0.001835
NEP	15	7	0.001838

NRU	8	1	0.001557
OMA	11	4	0.00179
PLE	8	8	0.001806
PLW	7	3	0.001732
PNG	12	7	0.001827
RWA	11	8	0.001828
SAM	11	24	0.001854
SEY	11	3	0.001765
SKN	8	3	0.001743
SLE	13	4	0.001799
SOL	11	2	0.001716
SOM	11	1	0.001579
SSD	3	14	0.001698
STP	8	3	0.001743
SWZ	12	3	0.00177
TLS	6	4	0.001742
TUV	5	2	0.001648
UAE	1	16	0.001365
VAN	10	6	0.001809
VIN	10	4	0.001784
YEM	9	4	0.001777