Problem Chosen	2025 MCM/ICM Summary Sheet	Team Control Number
C		2512703

Medals in the Making: Unveiling Olympic Predictions Through Data

Summary

The prediction of the Olympic medal table holds significant implications for strategic planning in sports development, resource allocation, and the evaluation of event influence. This study, centered on the 2028 Summer Olympics in Los Angeles, constructed two medal prediction models, namely the information entropy weighted model and the TOPSIS method, and build an Influence Quantification Model to analyzed the "star coach" effect.

For Problem 1, we propose an interpretable influencing factor analysis model based on the theory of weight entropy: the **Information Entropy Weighting Model (IEWM)**. In IEWM, the athlete's status value and the standard project value are set as the core independent variables. We further propose a novel method that integrates ID3 algorithm and entropy weight method to estimate the weight of the above influencing factors. **Compared to the deep learning model Long Short-Term Memory (LSTM)**, which is just trained on historical data, our proposed model performs **closer to real data results** and has **better interpretability**. The model predicts that China, Japan, South Korea, Australia, and other countries may experience **a decline** in their medal table rankings, while Britain, France, the Netherlands, Germany, and others are expected to **improve**. Our results are shown in Table 5.

In Problem 2, our **TOPSIS Prediction Model** analyzes and predicts which countries are likely to win their first gold medal by considering the number of participants in 2024 and the number of times they have participated in the Olympic Games since 2000 as independent variables. Our results show that MLI, GUI, and ANG have a very high probability of winning their first gold medal in the next Olympic Games.

For Problem 3, we applied the **TOPSIS** algorithm to analyze the specialty events of each country, and presented the results in Table 6. We concluded that specialty events can significantly enhance a country's ability to win medals.

With regard to Problem 4, we propose a model to quantify the influence of "great coaches" on athletes' medal-winning abilities using **the Momentum Metric.** The Momentum Metric quantifies the coach's influence by fitting a linear model and obtaining the normalized coefficient of the great coach parameter. Additionally, we conducted an ablation experiment using **AutoRegressive Integrated Moving Average**(**ARIMA)** (specifically ARIMA(2, 1, 1)). The results demonstrate that our metric effectively quantifies the impact of the Great Coach.

As for Problem 5, through **Exploratory Data Analysis (EDA)** and our aforementioned **IEWM**, we revealed some special phenomena of the Olympic games, such as smaller countries breaking the medal monopoly of stronger nations in emerging events, and the increasing gender balance in Olympic medal achievements.

At the very last, we analyze the **strengths** and **weaknesses** of our model as well as its **sensitivity**.

Keywords: Olympic medal table prediction; Information entropy weighted model; TOPSIS method;LSTM; Star coach effect; ARIMA;Exploratory Data Analysis (EDA)

Contents

1 Introduction

1.1 Problem Background

In recent years, the global attention to the Olympic Games has grown, with far-reaching impacts on politics, economy, society, and culture. Fans closely follow the Olympic medal table, and some small countries have made breakthroughs in winning medals, makingpredictions about Olympic results a hot topic. Olympic results reflect a country's competitive strength, and nations are increasingly focused on prediction models to gain insights. Predicting Olympic medal outcomes can provide valuable information to Olympic committees, helping them optimize resource allocation, design targeted training programs, and make data-driven decisions to improve overall performance.

1.2 Restatement of the Problem

In this problem, we are given the data of Information about the Summer Olympics, including: national medal tables, athlete information, Summer Olympics hosts, Summer Olympics program information. We will solve the following problems based on the historical Olympic Data:

1. Predict the medal table of each country at the 2028 Los Angeles Olympics in the United States, predict which countries are relatively progressive, and which countries are relatively regressive;
2. Which countries will win their first medal at next year's Olympics, with an estimated probability;
3. The relationship between the project and the country's medal award, and explore how the project chosen by the host country will affect the outcome of the award;
4. Explore the phenomenon of great coaching and list projects that have invested in three countries that have introduced a great coach;
5. Provide additional insights into the number of gold medals won at the Olympics based on the model.

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). The study applied models such as linear regression, decision trees, and neural networks to predict medal counts (gold, silver, bronze) and identified factors affecting success.

Julia Bredtmann, Carsten J. Crede, and Sebastian Otten published in Significance, Issue 3, 2016, "Olympic medals: does the past predict the future?" examined whether past Olympic medal counts, sports participation, and government investment could predict future medal outcomes. They used regression analysis, time series analysis, and random forests to study

how unequal resource distribution impacts countries' chances of winning medals.

1.4 Our Work

We propose a model framework as shown in Figure 2, which mainly includes three models, namely **IEW Model**, **TOPSIS Prediction Model**, **Impact Quantification Model**, and some evaluation criteria, such as **association rules**, **linear correlation criteria.**

The **IEW Model** is used to predict the number of national medals, the **TOPSIS Prediction Model** is based on the improvement of the TOPSIS algorithm to finely evaluate whether a country can win the next medal.

Regarding the impact of great coaches, we use the **Influence Quantification Model**, which combines the **momentum metric** and the **linear metric**, to quantify the influence of great coaches.

In addition, we also conducted comparative experiments to prove the effectiveness of the model, such as using **LSTM** to predict the number of national medals directly based on time compared to our **IEW Model**, and using ***ARIMA(p, d, q)*** (specifically *ARIMA(2,1,1)*)to compare data excluding the great coach factor with data containing the great coach factor.

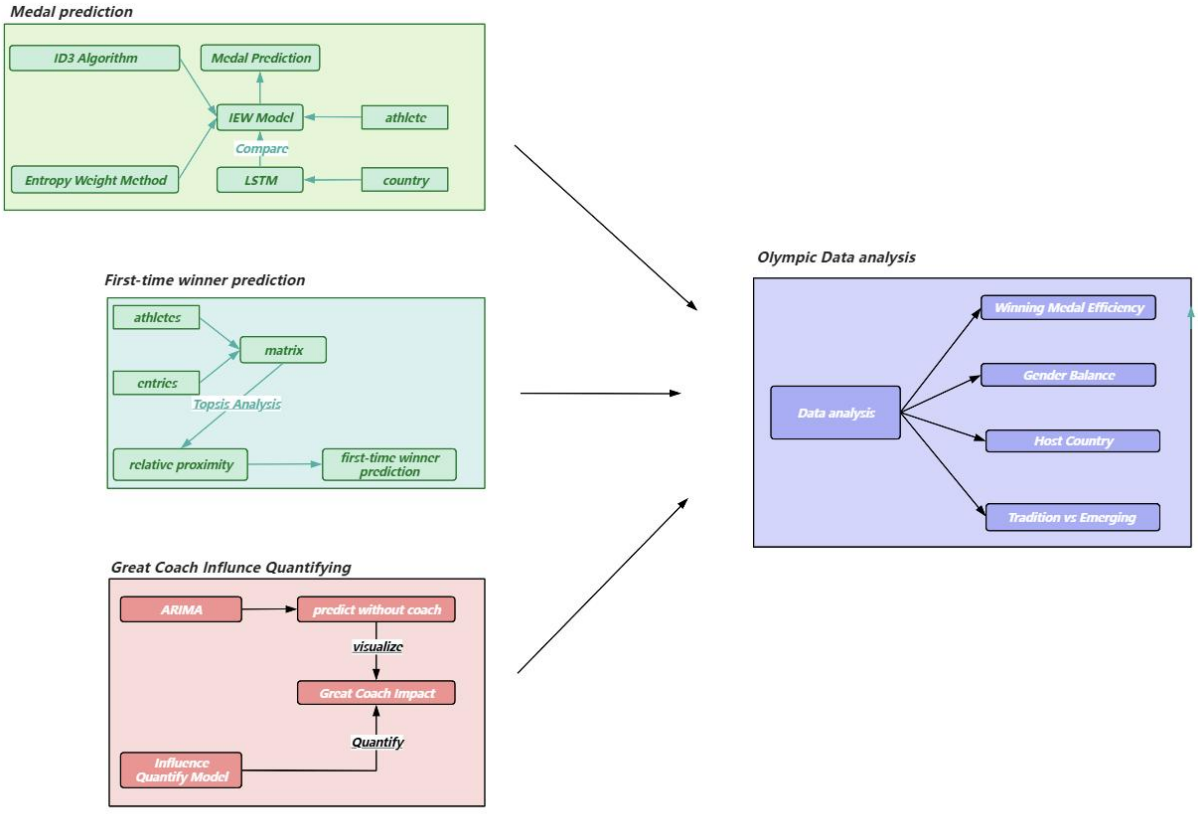


Figure1 our work

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 other reasons were not considered.

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.

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.

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.

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
φ	Standard item ratio
K	Maximum ambient capacity
r	Natural growth rate
A	Residual growth space ratio
PI	imprecision
D	Distance
C	Relative approximation
nae	The number of athletes in the project
P	The ability of a great coach himself
R	Inertia of team capabilities
O	The number of sessions of the Olympiad

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.

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

To analyze the performance of athletes from a specific country, we refer to their historical performance and conduct regression analysis based on the country's results in previous Olympic Games to reflect the current state of its athletes. As shown in the attachment, athletes participate in every Olympic Games. The gold medal winning rate is calculated by dividing the number of gold medals won by the total number of events. Since the status of the three types of medals varies, the gold, silver, and bronze medal winning rates are weighted and summed, with weights of 3, 2, and 1 respectively, to obtain a new variable. We define this as the athlete status value X, as follows:

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

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

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. 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, φ=1 and

$$\varphi = \frac{E}{300} (4.3)$$

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

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

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.

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:

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

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

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

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

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

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	1.556656
Information gain and entropy of athlete states	(1.138724, 0.277096)
Information gain and information entropy of the total amount of the project	(0.181508, 1.234312)
k_1, k	(0.675601, 0.324398)

The results are as follows:

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

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

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

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

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

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

Table 4 Data results of all indicators for silver and bronze MEDALS

Indicators	Silver Medal	Bronze
Unconditional entropy	1.556656	
k_1, k	0.457304, 0.542695	0.826933, 0.173066
Q	27.4	41.4

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

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

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 φ. We can reflect the φ value by two-dimensional image analysis of the total number of events E and the year:

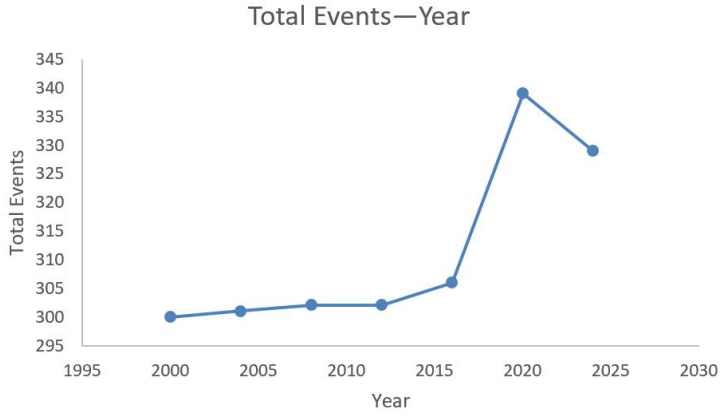


Figure 2 Relationship between total number of events and year

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:

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

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 5 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:

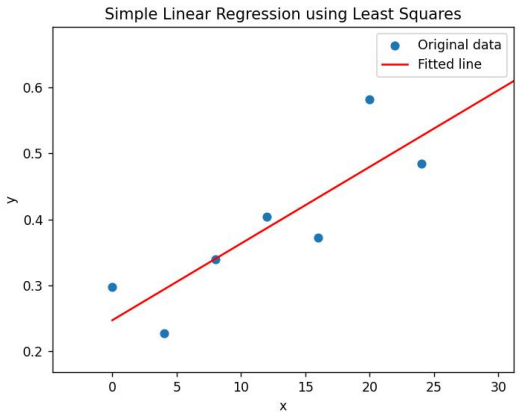


Figure 3 Regression analysis of athlete status values for France

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 6 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 7 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 Comparision Test

We use LSTM to make predictions based on historical national award data, and compare them with our proposed IEW model, and the experimental results are as follows:

Table 8 Comparison of prediction results

Actual situation	IEWM	LSTM
United States	United States	United States
China	China	China
Netherlands	Japan	Britain
Japan	Australia	France
France	France	Australia
Australia	Netherlands	Japan
South Korea	Britain	Germany
Germany	South Korea	Italy

Britain	Italy	South Korea
Italy	Germany	Netherlands

The MAE of the ranks of these countries is:

$$MAE_{IEWM} = 2.4 \qquad MAE_{LSTM} = 10.0$$

Experiments have shown that our IEW Model can model the state of athletes in a more refined manner, so as to make more accurate predictions.

4.4 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. $\log_2 3 \approx 1.5850$.

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 9 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 10 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. The model has passed the reliable test.

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

The medal-winning efficiency varies across countries. The figure below shows the relationship between the number of participants and the number of medals in each country:

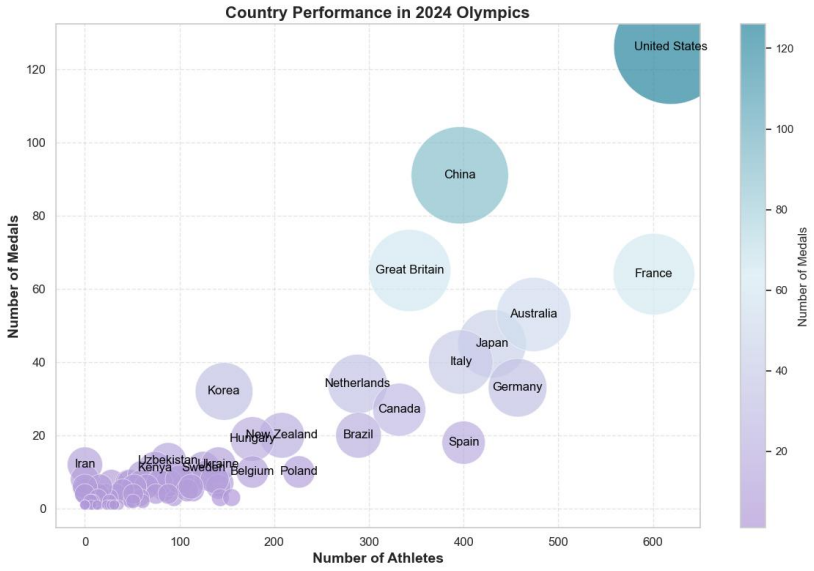


Figure 4 relationship between the number of participants and the number of medals in each country

Additionally, greater familiarity with the rules, resulting from frequent participation, can also improve the medal-winning rate. 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, we select the countries that have never won a medal from the table, and then count their number of participants in the 2024 Olympics as well as the number of times they have participated in the Olympics since 2000. There are 68 countries that have not won a medal but will participate in the 2024 Olympics. By combining the number of participants and the number of Olympic appearances, we can create a 68x2 matrix. Below is a sample of this matrix:

Table 11 Number of Olympic Games participated by countries and 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 TOPSIS method involves constructing the positive and negative ideal solutions for a decision problem, representing the best and worst values for each attribute. Then, the distances from each alternative to these ideal solutions are calculated, with the solution closer to the positive ideal and farther from the negative ideal being considered better. Below 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 countries that frequently win medals but are almost absent in the Olympics, and where the number of athletes per country is large, it's challenging to measure the contribution of these indicators. Thus, we assume the contribution of the number of entries and participants is equal, with a weight of 0.5, and the ideal solution is the maximum value. , and the negative ideal solution is the minimum value, then there is a distance d$max_j min_j$

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

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

It can be concluded that the relative closeness

$$C = \frac{D^-}{D^+ + D^-}$$

comparing the size of the relative approximation can get the most likely to get the medal of the country.

5.3 Solving the model

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

Table 12 Relative approximation for each country

NOC	MLI	GUI	ANG	SAM	LBR	ESA	NEP	MAD	NCA	GAM
C	0.0018 69	0.0018 63	0.0018 55	0.0018 54	0.0018 48	0.0018 44	0.0018 38	0.0018 35	0.0018 35	0.0018 28

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. An analysis of the situation of the first award-winning country in recent years was carried out and it was found that nearly 100 percent of the top three countries in relative proximity were the first winners of the yearthe 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 Great Coach Impact

In the study of the impact of great coaches on a team, we first use PACF and ACF to determine the parameters of the ARIMA model, and then apply the ARIMA(p, d, q) model to predict the data before the coach's tenure, which helps to exclude the influence of the great coach, allowing for a comparison with the actual data that accounts for the coach's impact. Next, we construct the Influence Quantification Model, which uses linear criteria to quantify the coach's impact on the team. Finally, we use the t-distribution test to assess the accuracy of the model.

6.1 Model Establishment

6.1.1 Establishing the model equation

Considering the volatility of medal counts, we use the variation in athletes' average scores to reflect the influence of the great coach on the team. First, we define the average score of athletes as follows, with the convention that the weight for gold medals is 3, silver medals is 2, and bronze medals is 1:

$$score_{ave} = \frac{3 \times G + 2 \times S + 1 \times B}{nae} \quad (6.1)$$

where $score_{ave}$ represents the average score of the athletes and Num represents the number of athletes. Clearly, $0 <= score_{ave} <= 3$.

We model the impact of a great coach on a team by examining their influence on the team's average score (considering that the number of medals varies with the number of events). The coach's effect on the team's score can be quantified as follows:

$$Impact = \frac{P}{R} \quad (6.2)$$

Here, P represents the coach's own ability, and R represents the team's performance inertia, which refers to the resistance to the coach's influence.

6.1.2 ARIMA Parameter Selection

We use the ARIMA model to predict the data without considering the influence of the great coach, in order to compare it with the real data. Taking the influence of Béla Károlyi on the U.S. women's artistic gymnastics team as an example, we first select the ARIMA parameters based on PACF and ACF:

Based on the PACF plot, we choose the first significant cutoff point as p, as shown in the figure, with p = 2. And based on the ACF plot, we select the first significant cutoff point as q, as shown in the figure, with q = 1.

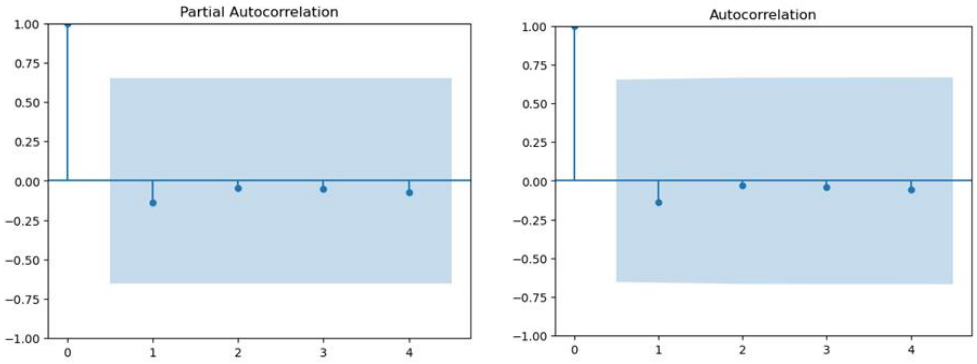


Figure 5 Partial Autocorrelation and Autocorrelation of the Series

Additionally, since the data is a non-stationary series, we set d = 1.

Therefore, we will use the ARIMA(2, 1, 1) for prediction in the next section.

6.1.3 Metrics

First, we use the ARIMA(2, 1, 1) model to predict the data based on the performance before the coach's tenure, and the difference between the ARIMA forecast and the actual data is shown in the figure below, illustrating the changes in the average performance of the U.S. women's artistic gymnastics team:

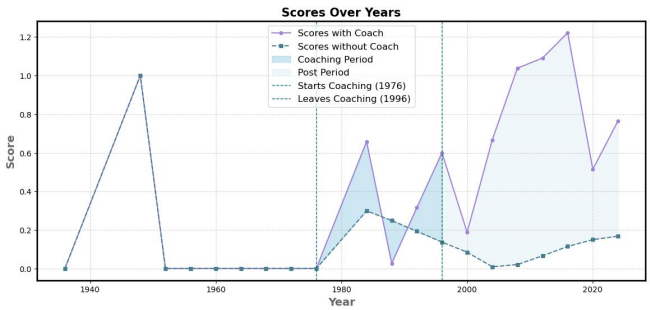


Figure 6 Comparison of ARIMA(2, 1, 1) Forecast and Actual Performance of U.S. Women's Artistic Gymnastics Team

As seen in the figure, Béla Károlyi brought a significant improvement to the average score of the U.S. women's artistic gymnastics team. Next, we will use two criteria to quantify this impact:

Linear Metric: We believe that the average performance of the team members in the current Olympic Games is related to the performance of athletes in the previous Games, the team's historical performance, the coach's influence, and other constant factors, such as nutrition. Therefore, we model the problem with the following equation:

$$score_{ave_i} = C + ascore_{ave_{i-1}} + \beta O_i + \gamma T_i (6.3)$$

Where C represents constant factors, Oi represents the historical performance of the team in the i-th Olympic Games, and Ti is an indicator variable representing whether there was a great coach in the i-th Games.

Using the above equation, we fit a linear regression model to the average score, which allows us to calculate the coefficient of the indicator variable Ti, representing the influence of the coach on the team. And in the above linear relationship, the team's resistance to the influence of the great coach, denoted as *re*, is:

$$re = |a + \beta| (6.4)$$

We apply a linear regression model to the average performance data of the U.S. women's gymnastics team and obtain the parameters as shown in the figure:

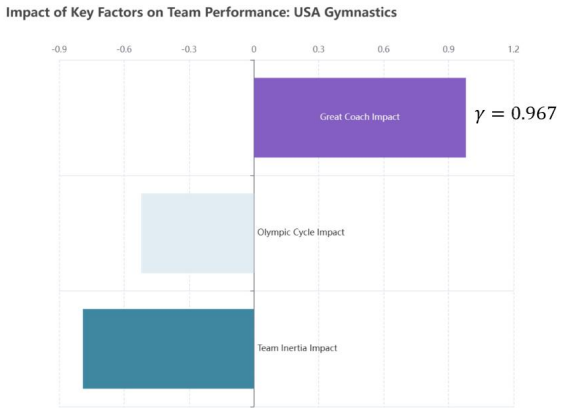


Figure 7 Coefficients of the Linear Model

It is clear that the coach's influence is positive and significantly higher than the other two coefficients, indicating that Coach Béla Károlyi indeed brought a tremendous improvement to the U.S. women's gymnastics team. This aligns with the facts and validates the effectiveness of our model.

6.2 Making decisions

We conducted the above analysis on three countries and three events to reveal how introducing great coaches in these events could lead to significant improvements.

6.2.1 Japan Women's Volleyball

Assuming that the Japanese team could hire Coach Lang Ping in the future, we calculate the coach's impact factor on the Japanese team based on the above linear criteria: $\gamma = 0.957$. Using the existing data, we obtain the performance change of the Japanese team, comparing it to the forecasted results without the coach using ARIMA(2, 1, 1), as shown in the figure below:

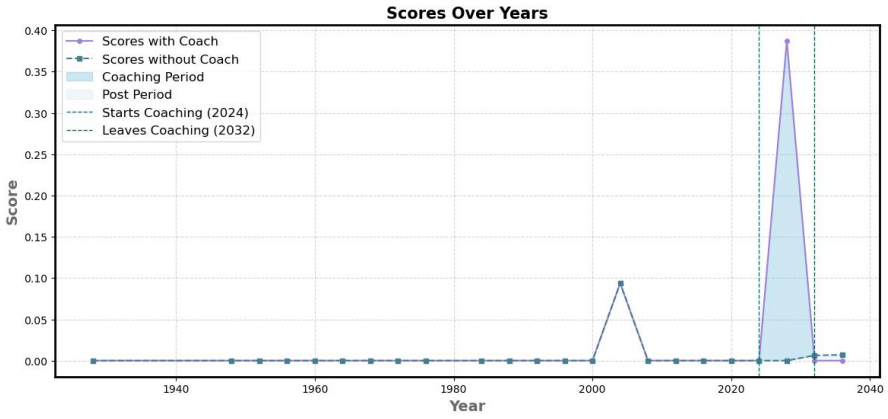


figure 8 Comparison of ARIMA Forecast and Actual Performance of Japan Women's Volleybal Team

6.2.2 French Women's Artistic Gymnastics

Assuming that the French Women's Artistic Gymnastics team could hire coach Béla Károlyi in the future, we calculate the influence factor of Coach Béla Károlyi for this event using the aforementioned linear method: $\gamma = 0.388$.

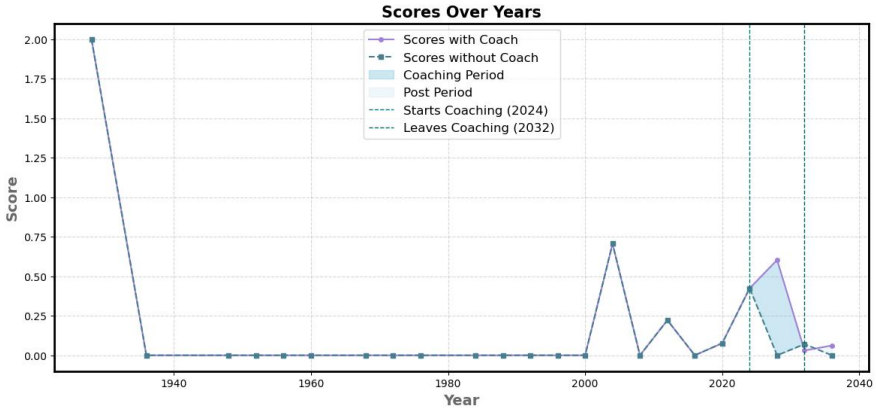
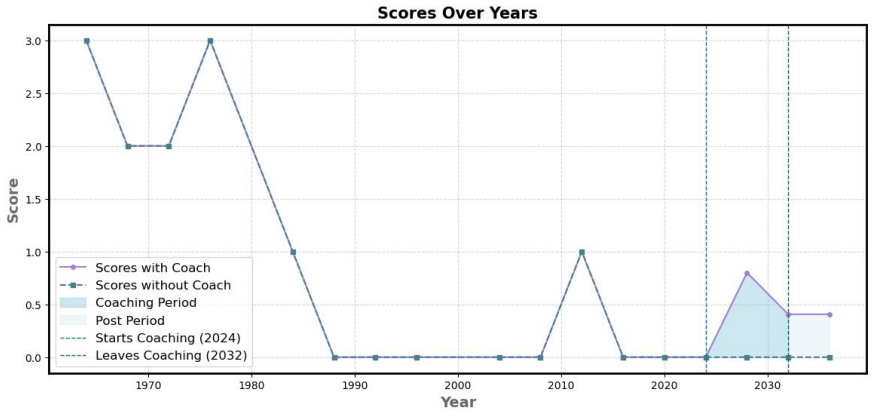


Figure 9 Comparison of ARIMA Forecast and Actual Performance of France Women's Artostic Gymnastics

6.2.3 Italy Women's Artistic Gymnastics

Assuming that the Italian Women's Artistic Gymnastics team could hire coach Béla

Károlyi in the future, we calculate the influence factor of Coach Béla Károlyi for this event



using the aforementioned linear method: $\gamma = 0.667$.

Figure 10 Comparison of ARIMA Forecast and Actual Performance of Italy Women's Artostic Gymnastics

7 Key Insights and Implications for Olympic Medal Predictions

According to the model in this paper, it is found that the Olympic medals are related to some variables, and we put forward the following insights for the prediction of the Olympic medals:

7.1 Winning the award for efficient countries and inefficient countries

We use our IEW Model to predict the medal outcomes for each country in the 2028 Olympics, compare them with the results from the 2024 Olympics, and fit the data using a linear regression model. The results are as follows:

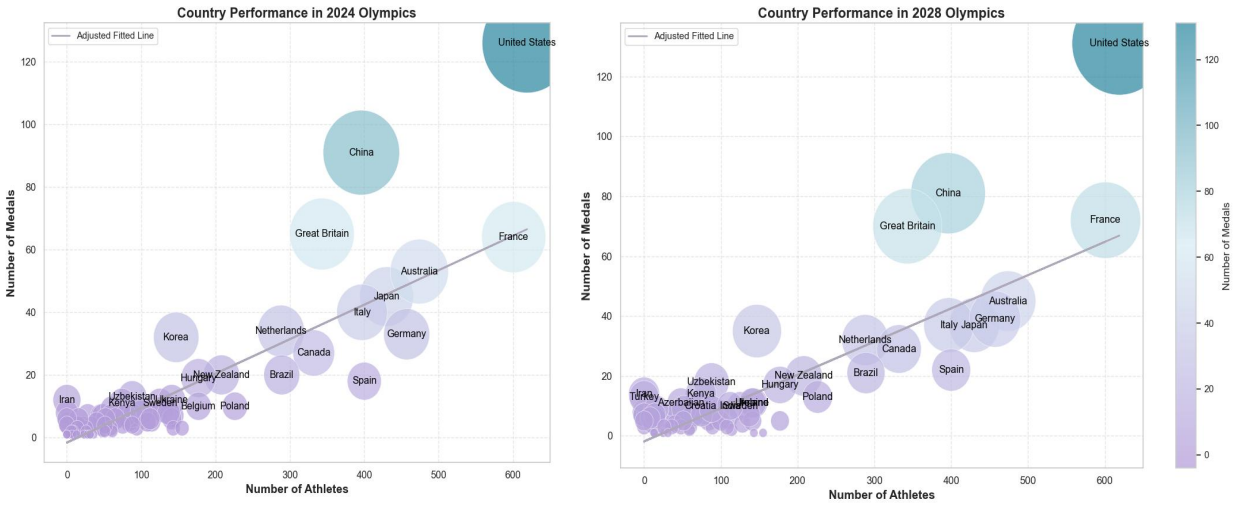


Figure 11 Comparison of Medal Efficiency Between 2024 and 2028

We define countries above the line as high-efficiency medal-winning countries and those below the line as low-efficiency medal-winning countries (of course, medal

count is not the primary goal of the Olympics, this is only for research purposes)

We found that countries like the United Kingdom and France have seen an improvement in medal-winning efficiency, while countries like China have experienced a decline in their efficiency.

Additionally, we observed that most smaller countries have seen an increase in their medal-winning efficiency. This reflects the Olympics' shift towards greater diversity. This could be due to the introduction of new events in the Olympics, allowing some smaller countries to break the medal monopoly of sporting powerhouses in emerging events, leading to an increase in medal-winning efficiency. We will discuss this point further below.

7.2 Gender and Olympic

The Olympic spirit embodies the values of excellence, respect, friendship, peace, and fair play, promoting unity and understanding among nations through the power of sport.

The Olympics advocates the spirit of equality, which is why we have studied the number of male and female athletes participating in the Games, as shown in the figure below.

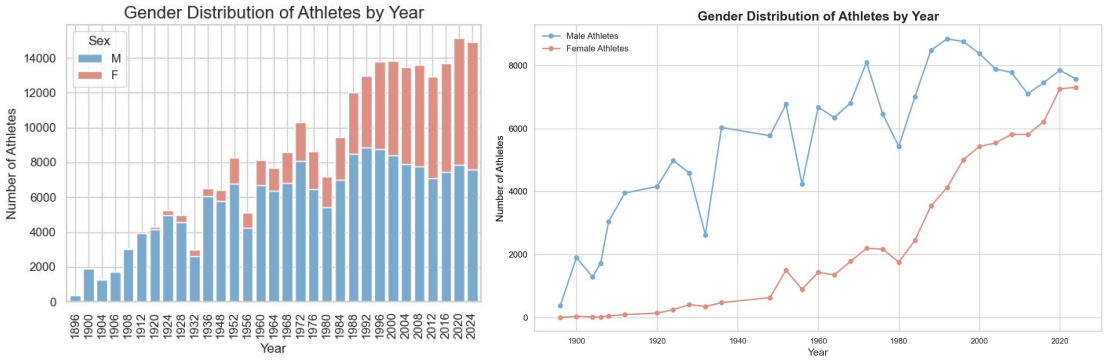


Figure 12 Number of men and women participating

We found that the number of female athletes participating in the Olympics is steadily increasing and is beginning to match that of male athletes. **This is a positive trend, signaling the increasing gender equality in sports and reflecting the growing recognition and opportunities for female athletes in the Olympic Games.**

7.3 Host Country as a Key Medal Factor

Our model analyzes whether being the host country is an important factor for a nation in

winning more medals at a particular Olympics. We first visualized the medal data for the United States and China, as shown in the figure below:

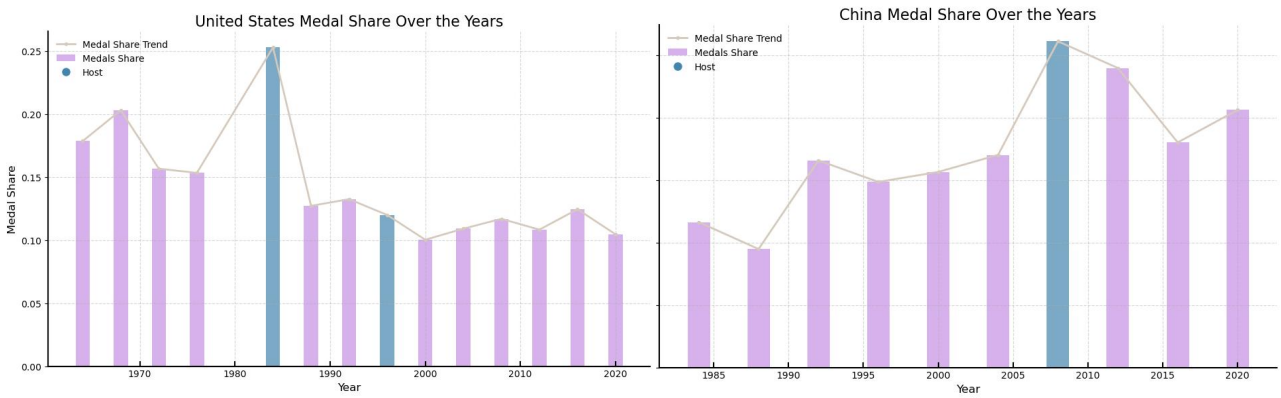


Figure 13 Host factor

It is clear that the host country has a significant impact on the nation's medal share. In Section 4, we used the TOPSIS method to predict the medal share of countries, and the results are shown above.

7.4 The difference between the awards of the sports powerhouses in traditional and emerging sports

In this section, we examine the medal performance of several countries in both traditional and emerging events at the 2024 Olympics and present the radar chart below(The data has been normalized).

The selected countries are the U.S., China, France, the U.K., Australia, and Lithuania, with the first five being sporting powerhouses and Lithuania having fewer medals.

The chosen events are volleyball, swimming, athletics, gymnastics, breakdancing, and skateboarding, with the first four being traditional sports and the latter two being emerging sports.

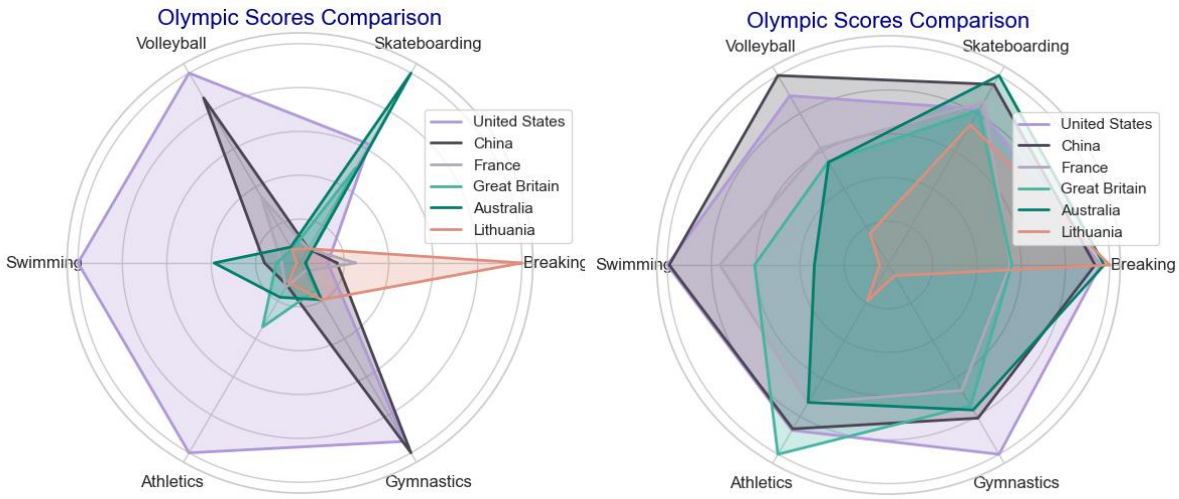


Figure 14 Olympic Scores Comparision

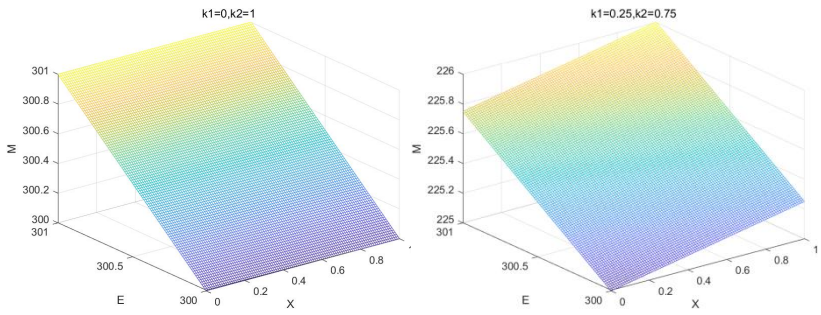
Interestingly, we found that although Lithuania's performance in traditional events is indeed not as strong as that of other sporting powerhouses, its performance in emerging events, especially breakdancing, surpasses that of other countries. This clearly demonstrates that in emerging events, traditional sporting powerhouses do not hold as significant an advantage over smaller countries, allowing smaller nations to break the monopoly of medals traditionally held by larger sports powers.

Additionally, we used our IEW Model to predict the medal outcomes for countries in 2028. We also plotted the radar chart for the same countries and events (with normalized data), and the results, shown in Figure 14, further validate the conclusions above.

8 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:$k_1 k_2 k_1 k_2 k_1$and$k_2 k_1$.

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



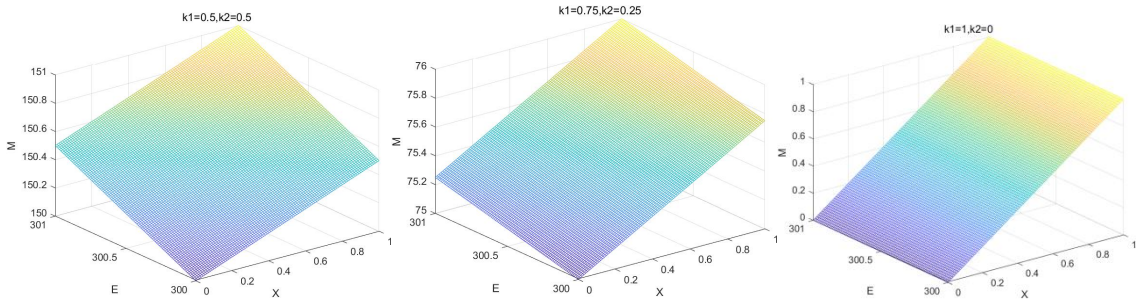


Figure 15 Sensitivity analysis of the model

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.$k_1 k_2 k_1 k_2$

9 Model Evaluation and Further Discussion

9.1 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.

9.2 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.

9.3 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.

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 relative proximity is that the probability of the first three countries winning the first medal is close to 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.

Report on use of AI

For the AI section, we used the GPT-04 model, which primarily handles translation tasks and the development of some tabulation programs.

Figure1.usage of ai