Problem Chosen	2025 MCM Summary Sheet	Team Control Number
C		2504787

Prediction of Medals and Quantification of the Great Coaches

Summary

The Olympic Games is one of the largest and most influential sports events in the world. In the upcoming 2028 Los Angeles Summer Olympics, how to allocate resources effectively to enhance Olympic performance will present significant challenges and hold substantial practical implications. Therefore, a modeling analysis was conducted.

Before constructing the model, we conducted data cleaning and normalization, and compiled the historical medal totals for each country. This preliminary work laid the foundation for the model construction and solution.

Firstly, to address the first question, we developed a **Medal Prediction and Correlation Model.** The model is based on the random forest algorithm to predict the total medal count for each country and performed a reliability analysis of the results with a 95% confidence level. We also predicted which countries' performances are likely to improve and which may decline, and forecast the medal standings for the 2028 Los Angeles Olympics. Next, we applied three regression methods—**polynomial regression**, **gradient boosting regression trees**, and **support vector regression** to analyze countries that have not yet won medals. After comparison, we selected the gradient boosting regression method with the best fit, predicting that four countries will win their first Olympic medals in 2028. Finally, we employed Pearson correlation analysis to examine the relationship between specific Olympic sports and medal counts, concluding that sports such as swimming, shooting, and gymnastics are strongly correlated with the countries' medal counts. This finding also confirmed that selecting sports aligned with a country's strengths or traditional culture can enhance performance, supporting the existence of the "Host Effect."

Secondly, to find evidence for the "Great Coach Effect," we used the **Analytic Hierarchy Process (AHP)** to quantify the performance levels of countries in the event. **The Mann-Kendall trend analysis method** was applied to identify years with significant changes, and we then identified all possible time points when "Great Coaches" could have moved between countries. After comparing with actual results, we confirmed the high reliability of our model. To improve the accuracy, we constructed a growth model with a volatility adjustment factor. By introducing a factor representing fluctuations in medal counts into the logistic growth model, we removed the natural growth effect caused by the overall increase in events, thereby quantifying the impact of the "Great Coach Effect." Based on this model, we provided quantitative analyses of the sports that require scientific investment and offered investment recommendations for each country.

Finally, by synthesizing all models, we systematically proposed a theory on Olympic medal counts, providing reasonable advice for National Olympic Committees (NOCs) on optimizing resource allocation, selecting potential sports, and improving medal performance.

Keywords: Medal Prediction；Random Forest；Gradient Boosting Regression Tree；Pearson Correlation Analysis；Mann-Kendall Trend Analysis；Logistic Growth Model；Resource Allocation

Contents

1 Introduction

1.1 Problem Background

The Olympic Games, as one of the largest and most influential sporting events globally, not only showcase the physical prowess and charm of athletes but also highlight the strength of various countries in different sports disciplines. Furthermore, the Games reflect the comprehensive political, economic, and cultural power of nations. After the Olympics, the medal standings often attract significant attention. Analyzing the medal table can provide insights into the changing political and economic influence of countries, as well as inform decisions regarding resource allocation in sports.

Taking the medal standings of the 2024 Paris Summer Olympics as an example, the United States, as a major political and economic power, topped the medal table. Traditional sports powerhouses such as China, Japan, Australia, and the United Kingdom also secured prominent positions. France, as the host country, also performed admirably. Furthermore, smaller nations such as Dominica and Saint Lucia, which won their first-ever gold medals, added diversity to the medal table and exemplified the Olympic spirit.

The changes in the medal table are influenced not only by the personal efforts of athletes but also by various factors such as national support and investment, sporting traditions, coaching, and the international environment. Predicting how countries will perform in future Olympics is both a challenging and practically significant task. In the upcoming 2028 Los Angeles Summer Olympics, factors like sport selection, coaching adjustments, and athlete changes will likely affect the performance of nations. Establishing a model to analyze these impacts will be crucial for countries to rationally adjust resource allocation and enhance their Olympic performance.

Figure 1 Olympic games

1.2 Restatement of the Problem

Considering the background information and restricted conditions identified in the problem statement, we need to solve the following problems:

Here is the translation of the outlined problems:

- Problem 1: Predicting and Analyzing Medal Counts
  - Predict the medal count for each country in the 2028 Los Angeles Summer Olympics and the overall medal standings. Identify which countries are likely to improve and which countries may decline.
  - Estimate the number of countries that will win their first-ever Olympic medals in the upcoming Games and calculate the probability of this estimate.
  - Investigate the relationship between sports events and medal counts, and explore the impact of sports selection on the performance of the host country.
- Problem 2: Evidence for the "Great Coach Effect"
  - Identify possible evidence for the existence of the "Great Coach Effect."
  - Quantify the impact of the "Great Coach Effect" on a country's medal count.
  - Develop investment strategies based on the "Great Coach Effect" and analyze their potential influence.
- Problem 3: Additional Insights from the Model
  - Use the constructed models to analyze any additional conclusions or insights that can be derived.

1.3 Our Work

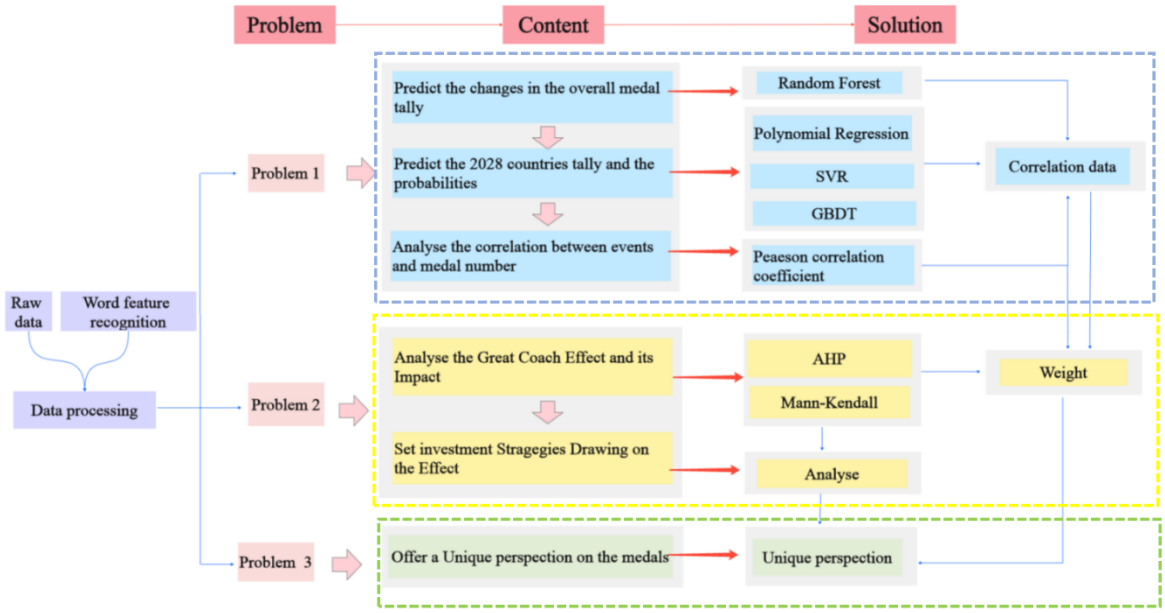


Figure 2 Flowchart

2 Assumptions and Justifications

● **Assumption 1**: Assume that there is a correlation between historical medal counts and future medal counts, and that historical medal counts can serve as a predictor for future medal counts.

● **Justification**: Historical medal counts reflect the past training levels of athletes, financial investments, coaching arrangements, and other political and economic factors of a country. Additionally, the political and economic stability of a country tends to remain stable over time.

● **Assumption 2**: Assume that the type and number of Olympic events are correlated with the number of medals won by each country.
● **Justification**: The host country may adjust the events for the current Olympics, and changes to events may also occur due to other factors. Different countries may excel in different events, and assuming a correlation between event selection and medal counts helps improve the accuracy of medal predictions.
● **Assumption 3**: We assume that when a coach moves from one country to another, they continue to coach the same events.
● **Justification**: Coaches typically remain stable in the events they coach during sports competitions.
● **Assumption 4**: Assume that the pre-processed data is reliable.
● **Justification**: This assumption is made to ensure the accuracy of the model solution.

3 Notations

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

Table 1 Notations used in this paper

Symbol	Description
M	Total Medals
MS	Total Medals in a Sport
MAS	Total Medals in All Sports
S	Sports
N	The Number of Sports
H	Historic Total Medals
δ	Host
E	Events
A	Athlete's Performance
Y	Year
R	The Best Result

4 Data Description and Preprocessing

4.1 Data Preprocessing

(1) Data Checking: We checked for duplicate data in each dataset to ensure the uniqueness and accuracy of each athlete and medal data.
(2) Data Imputation: We checked for missing values in each dataset and chose to fill missing values with 0 for medal counts or the number of events, and used "no medal" to fill in missing medal data.
(3) Outlier Detection: We checked for outliers in each dataset and identified and corrected any garbled data or special characters.

4.2 Data Description

By organizing the historical Olympic event data, we can observe that with the passage of time, the total number of Olympic events, disciplines, and sports has shown an increasing trend, albeit with fluctuations. The trend for Total Events is the most noticeable, and its visualization is shown below.

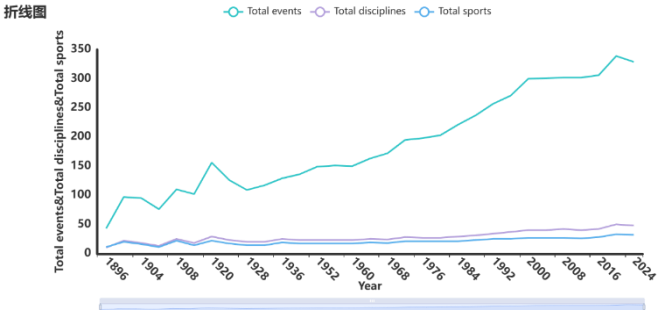


Figure 3 Total Events, Total Disciplines, and Total Sports Trend Over Time

Additionally, we compiled line charts showing the total number of medals won by each country in past Olympic Games over time. It is evident that as the number of Olympic events increased over time, the total number of medals won by countries also generally increased. Some countries that had never won medals before began to earn medals with the introduction of new events, while sports powerhouses continued to perform well in recent years. The following chart illustrates the total number of medals won by the United States, China, Japan, and Mexico over the years.

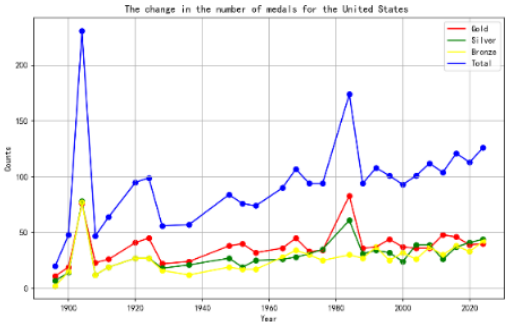


Figure 4 The change in the number of medals for the United States

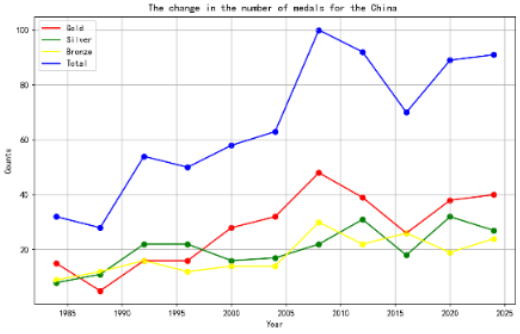


Figure 5 The change in the number of medals for the China

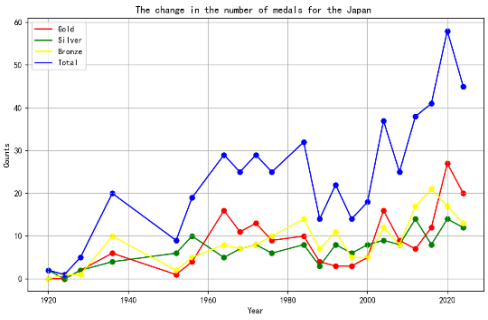


Figure 6 The change in the number of medals for the Japan

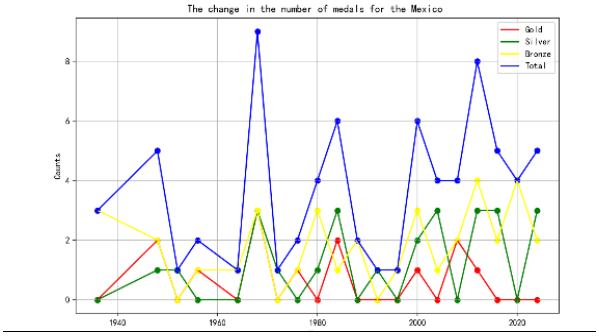


Figure 7 The change in the number of medals for the Mexico

5 Medal Prediction and Correlation Model

5.1 Random forest medal prediction model

5.1.1 Construction of random forest medal prediction model

(1) Model Construction Approach

Firstly, since the task explicitly mentions that directly predicting the total future medal count based on historical medal data is inaccurate, we aim for a more precise prediction by focusing on the list of current athletes who will participate in the upcoming Olympics. Therefore, we consider evaluating the contribution of each athlete to the country's total medal count. Additionally, since each athlete may participate in different events, we decided to analyze the contribution of all athletes to the country's medal count by organizing them by sport. The total medal count for each country across all events it participates in will then be used to predict the country's medal count for that year. The pattern for statistics and predictions is shown in the diagram below.

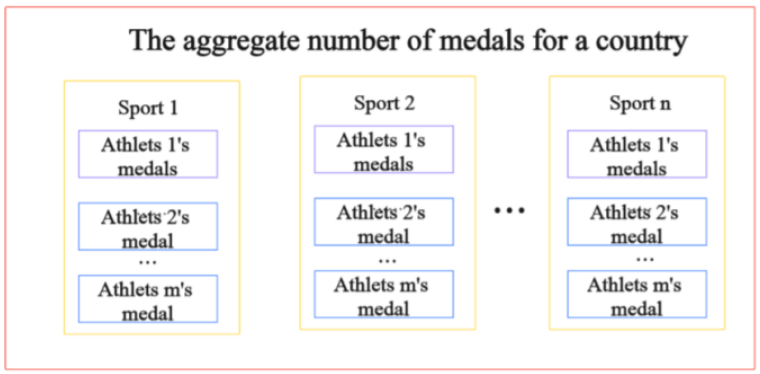


Figure 8 Flowchart

Further, we first need to identify the indicators that measure the contribution of an event (Sport) to the total medal count, i.e., selecting different types of indicators to assess the number of medals a country is likely to win in this event.

(2) Selection of Indicators

After comprehensive consideration, we use the total number of medals historically won by the country in this Sport and the highest level of medals historically achieved as factors to measure the overall level of the country's athletes. The total number of athletes participating in the Sport that year serves as an indicator of the country's investment and attention to the event, while external factors include the year, the name of the Sport, the total number of times this Sport has been held, and whether the country is the host for that year. Based on the background of the problem and the research data, here are the reasons for selecting some of these features:

Therefore, we statistically calculated the total number of medals each country has historically won in each event, the highest level of medals, the total number of athletes participating in each Sport at each Olympics, and whether the country was the host, to make further predictions.

(3) Dataset Construction

In this study, our goal is to predict the types of medals and total medal count for each country in the 2028 Olympics. We will predict each type of medal and the total medal count separately. First, we need to determine the input features and target variables for the model. Let the year be denoted as Year, the event name as Sport, the host status as HostFlag, the total number of participants from the country in this event as TotalParticipants, the total number of medals historically won by the country's athletes as TotalMedal, the best historical achievement as BestMedal, and the total number of times this event has been held as HeldTimes.

Note x_i =[Year i,Sport i, HostFlag i ,TotalParticipants i,TotalMedal i,BestMedal i,HeldTimes i]， BestMedal =[M],BestMedal =$\begin{cases}1, gold\\2, silver\\3, bronze\end{cases}$,

$$HostFlag = \begin{cases}0, yes\\1， no\end{cases} \quad (1)$$

In this case, **i** represents the i-th sport event that a country participates in during that year's Olympics. Since we are predicting the number of gold, silver, and bronze medals separately, we have performed predictions for each type of medal and the total medal count using the data mentioned above. The following descriptions refer to the total medal count prediction as an example, with the methods for predicting other types of medals being similar.

(4) Construction and Training of the Random Forest Model

First, construct a single decision tree. Starting from the root node, for categorical data such as Year i and Sport i, the subtree is directly constructed based on the different categories. For quantitative data, let the variance of the dataset at the root node be a. The values of y_i at the root node are sorted in ascending order, and a division point is selected in the middle to split the data into two parts, denoted as b and c. The split is made $\{b+c\}_{\min} < a$ at the position where $\min\{\Delta y_i\}$. In other words, the selected criterion **x** divides the data at the root node into two parts, creating two subtrees. This process is repeated to construct a single decision tree.

Under the premise of no data shuffling, no cross-validation, and not considering the maximum feature ratio during splits, the Mean Squared Error (MSE) is used as the evaluation criterion for node splitting. In each iteration, 80% of the data is randomly selected, and 70% of the features (i.e., 4 features) are randomly chosen to build a single decision tree. At the same time, the minimum number of samples at a leaf node is set to 1, and decision trees are built with a maximum of 50 leaf nodes and a maximum depth of 10. A total of 100 trees are constructed. The general structure of this process is shown in the diagram below.

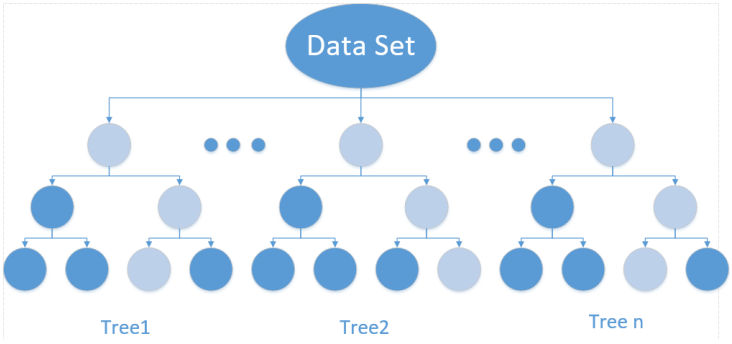


Figure 9 Random Forest

To ensure consistency in the data distribution between the training and testing sets, and to avoid significant biases caused by factors such as time span between the training and testing data, we divide 80% of the data for training and 20% for testing.

(5) Solving and calculating the results

Finally, we average the results independently provided by each random tree to obtain the final prediction result.

$$f(x)=\frac{1}{N}\sum_{i=1}^{N}T_i(x) \tag{2}$$

(6) Estimation of the confidence interval

$$\mu=\bar{x}\pm Z_{\frac{\alpha}{2}}\times\frac{\sigma}{\sqrt{n}} \tag{3}$$

The confidence interval is:

$$\left[\bar{x}-Z_{\frac{\alpha}{2}}\times\frac{\sigma}{\sqrt{n}},\bar{x}+Z_{\frac{\alpha}{2}}\times\frac{\sigma}{\sqrt{n}},\right] \tag{4}$$

5.1.2 The Solution of the Random Forest Model

(1) Solution of the result

Based on the random forest model we built, we predict the total number of medals each country will win with 95% confidence, and round the predicted medal count up. A portion of the results is displayed as follows:

Table 2 the total number of medals won by each country

Country	Predicted Medals in 2028	Lower Bound (95% CI)	Upper Bound (95% CI)	Predicted Medals in 2028
United States	126.39	117.5039952	134.4960048	126
China	90.01	83.45399524	100.4460048	90
East Germany	82.05	73.55399524	90.54600476	83
Soviet Union	71.97	63.47399524	80.46600476	72
Unified Team	62.52	54.02399524	71.01600476	63
…	…	…	…	…
Zambia	1.20	-7.296004756	9.696004756	2
Peru	1.16	-7.336004756	9.656004756	2

(2) Visualization of the results

Based on our predicted data, we have created a virtual medal table for the top 12 countries in the 2028 Los Angeles Olympics. This medal table is primarily sorted by the total number of medals, with gold, silver, and bronze medals as secondary, tertiary, and quaternary keywords, respectively. Additionally, we considered several factors and removed countries that historically

existed but no longer exist in the contemporary world, as well as countries that, due to political or war-related reasons, are unlikely to participate in the 2028 Olympics, such as Russia, which is banned by the International Olympic Committee. Furthermore, we have also created statistical charts predicting the total number of gold, silver, and bronze medals each country is expected to win.

Rank	Country/ Region				Total
01	United States	40	44	42	126
02	China	36	27	27	90
03	France	16	25	22	63
04	Great Britain	13	16	23	52
05	Australia	17	16	16	49
06	Japan	20	12	14	46
07	Italy	12	13	16	41
08	Nether-lands	14	9	12	33
09	Germany	13	13	17	33
10	Canada	10	7	10	27
11	New Zealand	10	7	4	21
12	Republic of Korea	6	4	10	20

*The number of medals is only a projection

Figure 10 2028 medal table

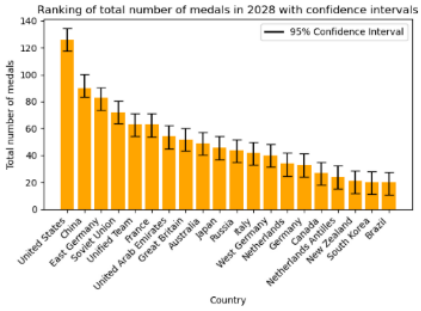


Figure 11 Total number of medals

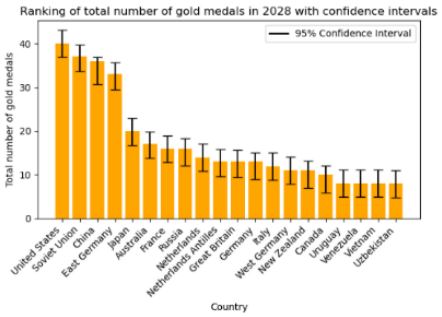


Figure 12 Total number of gold medals

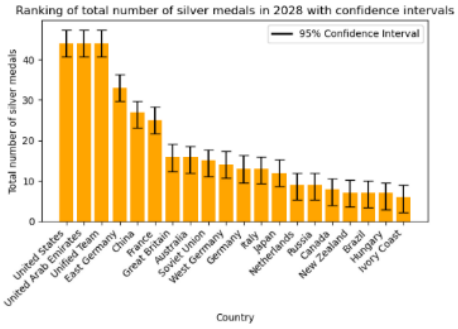


Figure 13 Total number of silver medals

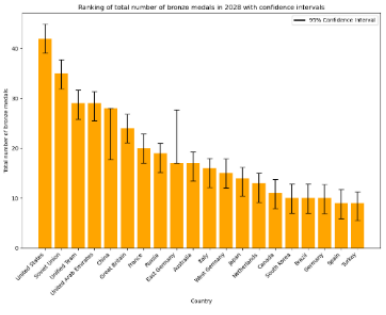


Figure 14 Total number of bronze medals

5.1.3 Model testing and evaluation

(1) Model testing

Based on common evaluation metrics, we selected five samples—Australia, the United States, France, Italy, and China—to assess the performance of the model.

Table 3 Model evaluation results

Countries	MSE	MAE	R^2
Australia	1.364305435	0.808586957	0.960733938
America	0.025857955	0.034886364	0.999515284
France	2.413717949	0.935384615	0.92332202
Italy	2.287594444	1.005	0.877174102
China	4.583110959	1.075205479	0.891337588

The above results indicate that R^2 of our model is very close to 1, suggesting that the model has a good fit.

(2) Error analysis :Select four samples: Australia, the United States, France, and China, to conduct error analysis.

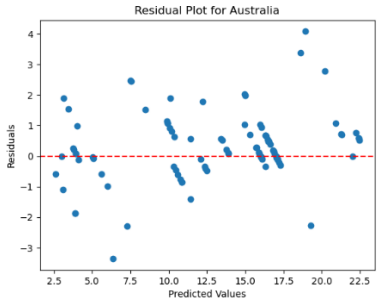


Figure 15 Residual plot for Australia

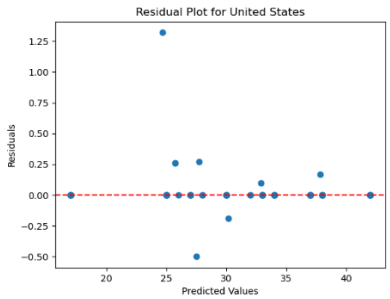


Figure 16 Residual plot for the United States

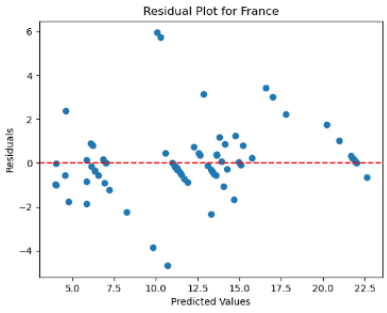


Figure 17 Residual plot for France

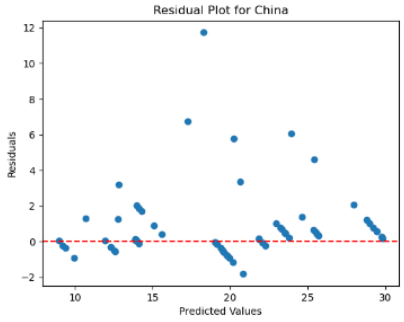


Figure 18 Residual plot for China

Based on the error analysis results, the model generates small prediction errors and exhibits good fit.

(3) Analysis of countries with progress and decline

We will compare the total number of medals predicted by our model for the 2028 Olympics with the number of medals each country won in the 2024 Olympics. After excluding countries that did not participate due to political or war-related reasons, we identify countries that are likely to improve their performance and those that may experience a decline. A portion of the results is displayed as follows:

Table 4: Model changes' results

Country	Predicted Medals in 2028	Round up.	Medals in 24	Changes
Czech Republic	17.46	18	5	13
Saint Lucia	10.87	11	2	9
Chile	7.86	8	2	6
Belgium	14.27	15	10	5
Turkey	11.78	12	7	5
Georgia	11.66	12	7	5
…	…	…	…	…
Moldova	1.67	2	4	-2
Australia	48.7	49	53	-4
South Korea	19.63	20	32	-12
Great Britain	51.61	52	65	-13

5.2 First-won medal countries model

5.2.1 The construction of First-Won Medal Countries Models

Based on the fact that the total number of countries participating in the Olympics remains unchanged, while the number of countries winning medals for the first time in each Olympic Games continues to decrease, we subtract the number of countries that have already won medals for the first time from the total number of participating countries each year, denoted as M**.** The number of countries winning medals for the first time in that particular Olympics is denoted as N. Use $\frac{N}{M}$ as p , and p belongs to(0,1)。We predict the probability $p_{first-won}$ of countries winning a medal for the first time in the next Olympic Games. We then use the same method to predict the number M_{next} of countries that have not yet won medals in the next Olympic Games, and the product of these two values represents the predicted outcome.

Additionally, considering the significant difference in scale between the year and probability, this might increase the prediction error. Normalizing both variables ensures that the scales of features and targets match, which generally improves the performance of the regression model. Therefore, we normalize both the year and the probability. Let the variable to be normalized be xx, then the normalized value is:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{5}$$

In addition, normalization helps to accelerate model training, especially for models optimized using gradient descent methods, such as support vector regression and gradient boosting regression. It can prevent certain features with large values from dominating the training process.

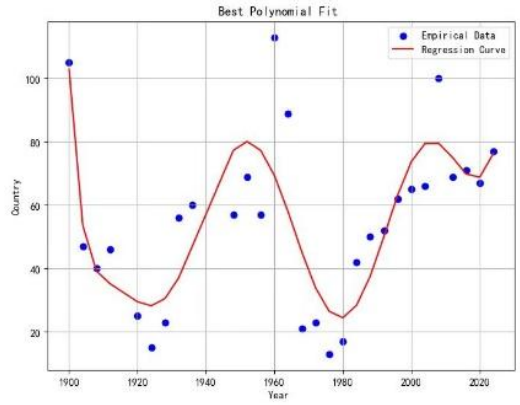
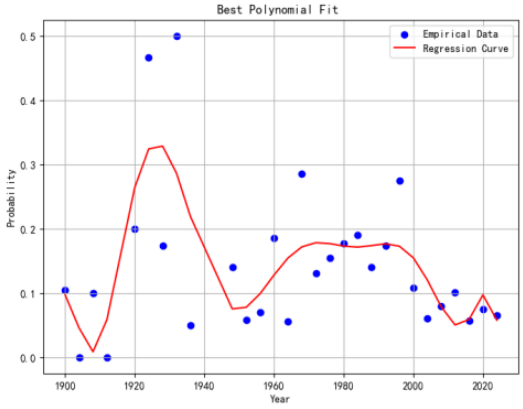
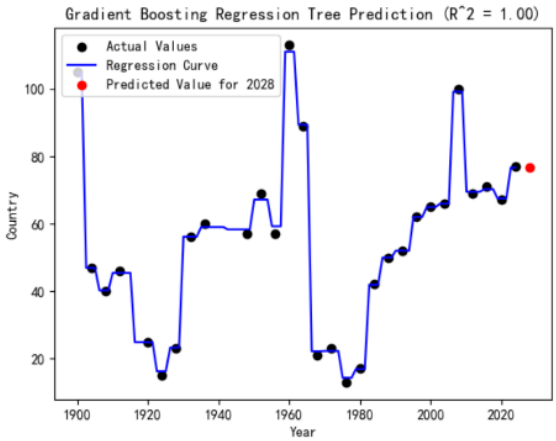
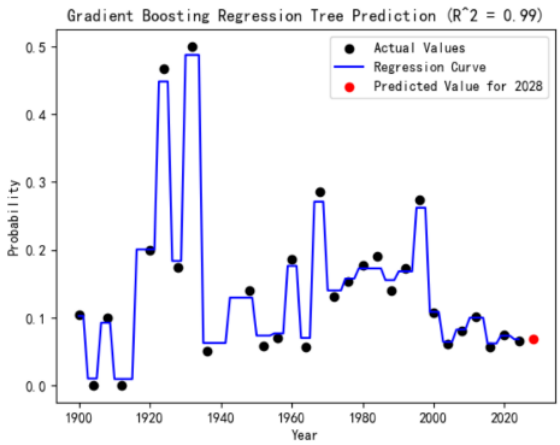
After performing normalization, we predict the normalized probabilities, and then apply de-normalization to obtain the predicted probability values. For the prediction of the number of countries that have not yet won medals in the next Olympic Games, we directly use the best regression method to make the prediction.

5.2.2 The Solution of First-Won Medal Countries Models

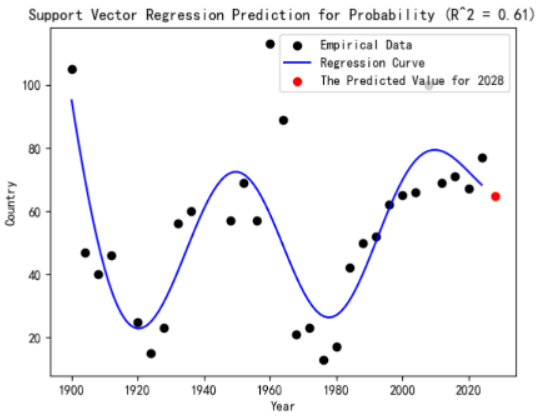
To ensure the accuracy of the model's predictions, we selected three regression methods—polynomial regression, gradient boosting regression trees, and support vector regression—to compare and identify the optimal regression method.

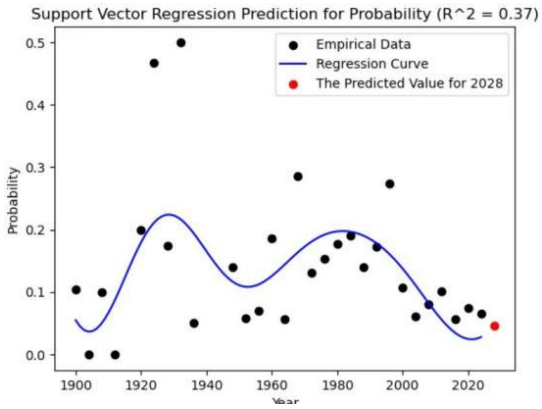
(1) The raw data, fitted curve, and regression probability are visualized as follows：

Table 5 raw data, fitted curve, and regression probability

Regression Method	Raw Data and Fitting Curve	Return Probability
Polynomial regression		
Gradient Descent Regression Tree		

Support Vector Regression





(2)The calculation results are as follows: Let the country's predicted value be F, the country's confidence interval be C, the probability prediction value be P, and the probability confidence interval be D. Since F and P are independent, the product confidence interval M can be calculated using the Delta Method.

Table 6 calculation results

	Polynomial regression	Gradient Descent Regression Tree	Support Vector Regression
F	91.43874611	76.60127434	64.83406965
C	[-6574.29,1068.6]	[66.951,77.006]	[11.200,92.922]
P	-0.39445605	0.06802101	0.04551444
D	[-18.449,15.537]	[0.057,0.080]	[-0.0630,0.218]
M	[-9592.2,16845.4]	[4.082,6.641]	[-3.515,14.202]

(2) The results of the model testing and evaluation are shown in the table below:

Table 7 results of the model testing and evaluation

Regression Method	Forecast the Count Value for 2028	R^2	MSE	RMSE
Polynomial regression	0.8486540578134014	0.46	540969.723540619	735.506440176
Gradient Boosting Regression	5.028608847971755	1.00	0.00	0.01
Eigenvector Regression	4.166487039158179	0.41	0.01	0.09

Finally, by rounding down the predicted number of countries, the predicted number of countries that will win a medal for the first time in the next Olympic Games is 4.

5.3 Sports' impact on the MAS

5.3.1 An introduction of the method

Correlation analysis is one of the commonly used methods in data analysis. By analyzing the relationships between different features or data, it helps identify key influences and driving factors of variables and predict the development of those variables.

Pearson correlation analysis is a method used to measure the strength of the linear relationship between two variables. It is typically used to assess whether there is a correlation

between two continuous variables. It expresses the linear relationship between the variables by calculating the Pearson correlation coefficient.

The formula for calculating the Pearson correlation coefficient rr is as follows:

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}} \tag{6}$$

The value of the Pearson correlation coefficient rr ranges from -1 to 1:

• r=1: Perfect positive correlation, indicating that as one variable increases, the other variable increases in exact proportion.

• r=−1: Perfect negative correlation, indicating that as one variable increases, the other variable decreases in exact proportion.

• r=0: No correlation, indicating that there is no linear relationship between the two variables.

The Pearson correlation coefficient provides a simple quantitative measure that intuitively reflects the strength and direction of the linear relationship between two variables. We use Pearson correlation analysis to quantify the impact of project selection on the number of medals. The closer rr is to 1, the stronger the correlation between the project and the country's medal count, meaning the project is more important to that country.

5.3.2 The Application of the Method

(1) We have compiled the number of medals each country has won in different events, where SPSP represents the number of medals in specific sports, and TMATMA represents the total medal count. Using Pearson correlation analysis, we analyze the relationship between a country's chosen events and its medal count, identify the events that each country excels in, and explore how to adjust strategies by selecting the right events to achieve better results.

(2) We focus on analyzing the impact of the events chosen by the United States, China, Denmark, and France on their final results. The closer the color is to red, the stronger the positive correlation, while the closer the color is to blue, the stronger the negative correlation. The correlation data for these four countries is visualized as follows:

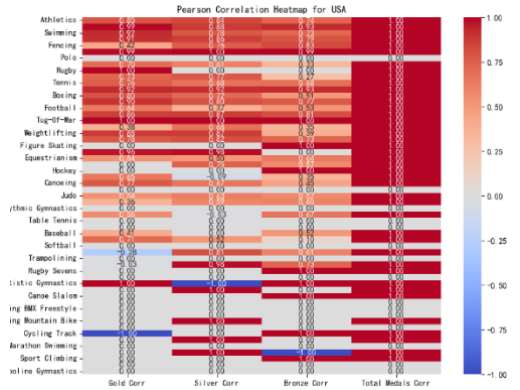


Figure 19 Pearson correlation on Heatmap for USA

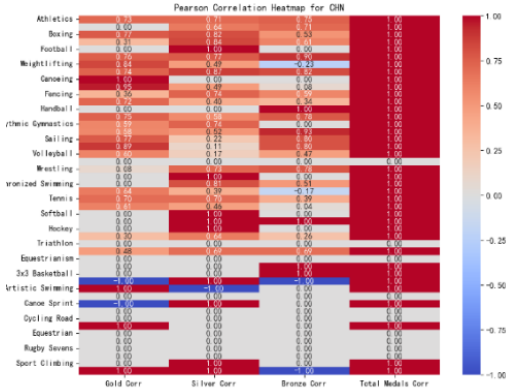


Figure 20 Pearson correlation on Heatmap for CHN

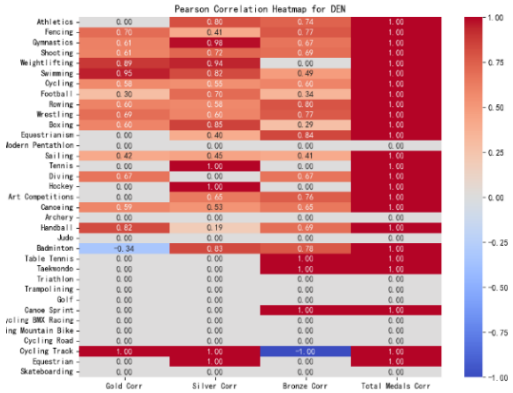


Figure 21 Pearson correlation on Heatmap for DEN

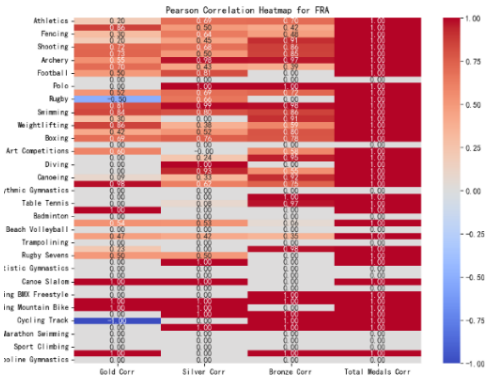


Figure 22 Pearson correlation on Heatmap for FRA

• United States (USA) Correlation Analysis:

In the correlation analysis for the United States, gymnastics, swimming, and weightlifting have correlation values of 0.80, 0.92, and 0.88 with the number of gold medals, respectively. This indicates that the United States' strength lies mainly in gymnastics, swimming, and weightlifting. The correlation between fencing and gold medals is only 0.42, suggesting that more investment is needed in fencing.

• China (CHN) Correlation Analysis:

In China's correlation analysis, weightlifting and synchronized swimming have correlation values of 0.84 and 1.00 with the number of gold medals, indicating that China has made significant progress in both weightlifting and synchronized swimming. The correlation between fencing and gold medals is only 0.36, which suggests that fencing remains a weak area for China, requiring further investment and more effective training strategies.

• Denmark (DEN) Correlation Analysis:

In Denmark's correlation analysis, weightlifting and swimming have correlation values of 0.89 and 0.95 with the number of gold medals, indicating that Denmark has consistently maintained its strength in these areas. The correlation between badminton and gold medals is -0.34, indicating that Denmark needs to increase investment in badminton.

• France (FRA) Correlation Analysis:

In France's correlation analysis, swimming and weightlifting have correlation values of 0.84 and 0.86 with the number of gold medals, highlighting that France's strengths are concentrated in swimming and weightlifting. The correlation between gymnastics and gold medals is only 0.20, and fencing has a correlation of only 0.30, suggesting that France still needs more focus on gymnastics and fencing.

(3) We selected data from the 2012-2024 Olympic Games for six traditional sports powerhouses—United States, South Korea, the United Kingdom, France, China, and Australia—focusing on their participation in events and the number of medals won. The visualization uses color: the closer the color is to dark blue, the more frequent the participation; the closer the

color is to light yellow, the less frequent the participation. In 2024, South Korea did not participate in the listed events.

The analysis leads to the following conclusions: Swimming, shooting, and gymnastics have strong correlations with the number of medals won by the countries, making these key areas to focus on for training. Weightlifting and table tennis, on the other hand, show weaker correlations with medal counts, and countries might achieve significant improvements by increasing investment in these areas.

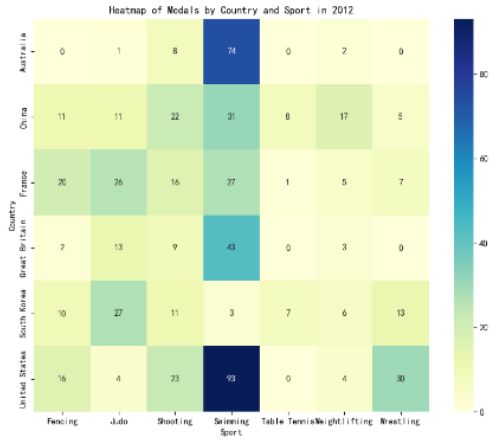


Figure 23 Heatmap of medals by country and sport in 2012

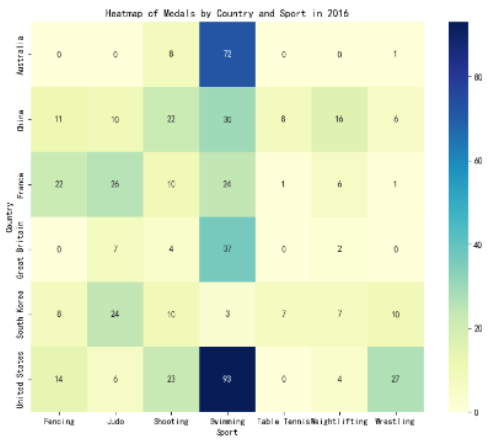


Figure 24 Heatmap of medals by country and sprot in 2016

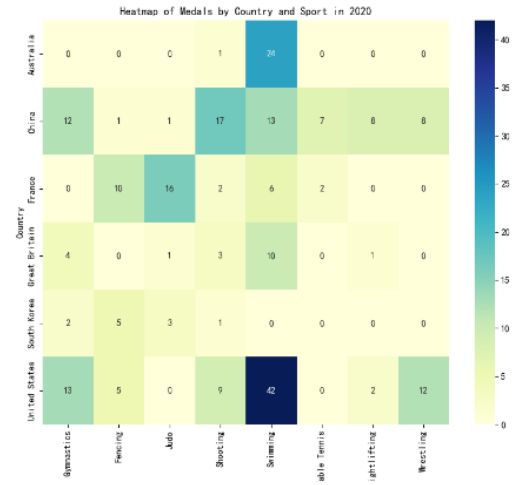


Figure 25 Heatmap of medals by country and sport in 2020

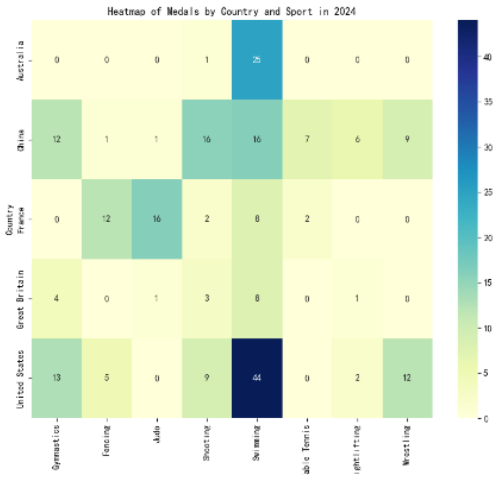


Figure 26 Heatmap of medals by country and sport in 2024

(4)Based on the fact that host countries can introduce events that reflect their traditional history and culture, and select events that highlight their national strengths, we further explore the impact of the events chosen by host countries on the final medal count. We perform a correlation analysis between the events chosen by the host countries and the medals won, then calculate the average values and visualize the results. Additionally, we perform a correlation analysis

for all events and their corresponding final medal counts, calculate the average values, and visualize the results. A comparison of the results is presented below:

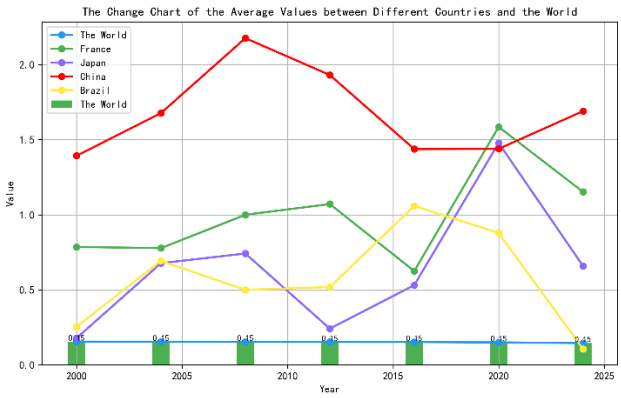


Figure 27 The change chart of the average values between different countries and the world

After analyzing the data and visualizations, the following conclusions can be drawn: The overall correlation between the selected events and the number of medals remains largely unchanged across each Olympic Games. The correlation between the events chosen by the host country and the medals won is stronger than the correlation between all events and medals won overall, indicating that the introduction of events where the host country excels contributes to the host country winning more medals.

6 'Great Coach'Effect Identifying Model

6.1 The Construction of the Model

(1)For each event, identify significant progress in a country's performance in a specific event at a particular Olympic Games due to coaching changes.
(2)Since the total number of medals is not a representative measure, we construct an index HHH composed of the types of medals and their related factors to assess a country's level in a specific event. The formula for calculating H is as follows:

$$H = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 + \alpha_5 \mathrm{E} \tag{7}$$

Where x_1 is the number of gold medals won, x_2 is the number of silver medals won, x_3 is the number of bronze medals won, x_4 is the total number of medals won, and E is the number of athletes participating.
Additionally, the value of HH should be referenced for no more than three Olympic Games.

(3)Let H_i be the value of a specific event for a country at the i-th Olympic Games. When the following condition is met: $H_i << H_{i+1}$.It indicates that the country achieved significantly better results in the current Olympic Games compared to previous ones, suggesting that a "great

coach" has entered the country. When the following condition is met: $H_i >> H_{i+1} \& H_i >> H_{i-1}$. In other words, when a country achieves better results in the current Olympic Games than both the previous and subsequent ones, and after a certain year, the countries' performance in the event significantly declines, it can be interpreted as the departure of a "great coach." This corresponds to a situation where the value of H shows a significant upward trend followed by a subsequent downward trend (with the downward trend not necessarily being significant). The time period where this trend occurs is the potential departure period of the "great coach," and the peak value of H corresponds to the "departure point."

Similarly, a time period where H shows a significant upward trend is considered the potential arrival period of the "great coach" at another country. The initial moment of the upward trend is defined as the "arrival point."

(4)Let all possible "departure points" where the "great coach" effect may occur be denoted as ξ_i, and all possible "arrival points" where the "great coach" effect may occur be denoted as ζ_j.

We then need to determine if there exists a time interval between these points where the following conditions are met:

$$\xi_i - \zeta_j \leq 4 \tag{8}$$

(5)Based on the strong correlation between this problem and time features, we choose the Mann-Kendall method to analyze the historical trend of each country's medal count in order to assess the potential impact of the "great coach effect" on the medal count. This method is suitable because the impact of coaching changes typically exhibits a delayed effect.

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sign(x_j - x_i) \tag{9}$$

Then calculate its variance:

$$Var(S) = \frac{n(n-1)(2n+5) - \sum t(t-1)(2t+5)}{18} \tag{10}$$

(6)Normalize the S value to obtain the Z value. If the Z value is positive and greater than 1.96, we consider the time series data to exhibit a significant upward trend; if the Z value is negative and less than -1.96, we consider the data to exhibit a significant downward trend; if the Z value is close to zero, it indicates no significant trend

6.2 Analysis Process

6.2.1 Finding the "Great Coach" Effect

(1)We use the Analytic Hierarchy Process (AHP) to calculate the vector weights. AHP is a structured decision-making method that determines the importance or weight of various factors or options by constructing a hierarchical structure and performing pairwise comparisons.

● First, we take the index H as the goal layer and x_1, x_2, x_3, x_4 and E as the criterion layers.

● Suppose the importance of x_1 relative to x_2 is 5, the importance of x_1 relative to x_2 is 9, the importance of x_1 relative to x_2 is 13, and the importance of x_1 relative to x_2 is 18. The judgment matrix is constructed as follows:"

$$A=\begin{bmatrix} 1 & 5 & 9 & 13 & 18 \\ \frac{1}{5} & 1 & 5 & 9 & 15 \\ \frac{1}{9} & \frac{1}{5} & 1 & 3 & 9 \\ \frac{1}{13} & \frac{1}{9} & \frac{1}{3} & 1 & 3 \\ \frac{1}{18} & \frac{1}{15} & \frac{1}{9} & \frac{1}{3} & 1 \end{bmatrix} \tag{11}$$

After normalizing the judgment matrix, calculate the average value of each row. The following results are obtained:

$$\lambda = 5.41, \alpha_1 = 0.6183, \alpha_2 = 0.2478, \alpha_3 = 0.0833, \alpha_4 = 0.0344, \alpha_5 = 0,0162$$

$$H = 0.6138x_1 + 0.2478x_2 + 0.0833x_3 + 0.0344x_4 + 0.0162\mathrm{E} \tag{12}$$

To assess the rationality and consistency of the judgment matrix, we perform a consistency check. The indicators and results are as follows：

Table 8 The indicators and results

Index	Formula	Result
Consistency Indicators	$CI = \frac{\lambda_{\max} - n}{n-1}$	0.127
Consistency Ratio	$CR = \frac{CI}{RI}$	0.0917

(2)After calculation, the maximum eigenvalue of the $\lambda_{\max}$ we constructed is 5.41, which is approximately equal to the matrix dimension of 5. Since CR<0.1CR < 0.1CR<0.1, it indicates that the model has good consistency

Using the volleyball event for China and the United States, and the gymnastics event for Romania and the United States as examples, we analyze the changes in the H values for China's and the United States' volleyball events, as well as the changes in the H values for Romania's and the United States' gymnastics events. The visualized results are as follows:

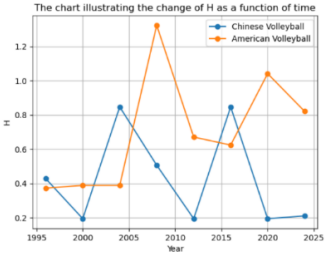


Figure 28 the H values for volleyball events

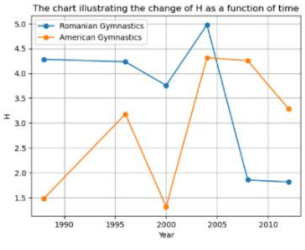


Figure 29 the H values for gymnastics events

(3)After Lang Ping left China in 1999, China's performance declined in 2008. In 2005, Lang Ping began coaching the United States, helping the U.S. volleyball team achieve excellent results in 2008. Based on this, Lang Ping can be considered to have the 'great coach' effect.

Bela Karolyi left Romania in 1977, which led to Romania's disappointing performance in the 1996 Olympics. Bela Karolyi began coaching the United States in 1981, which resulted in the U.S. gymnastics team's outstanding performance in 1996. Based on this, Bela Karolyi can be considered a coach with the 'great coach' effect.

(4)Based on the model analysis, we have organized all the possible time points (Year) when coaches with the 'great coach' effect left or entered a country. We denote departure as O and arrival as I. The partial results we obtained are shown in the table below:

Table 9 The partial results

Country	Sport	Point Type	Year
Australia	Athletics	O	1920
Australia	Athletics	O	1924
Australia	Athletics	O	1952
…	…	…	…
Uzbekistan	Taekwondo	I	2012
Uzbekistan	Taekwondo	I	2016
Serbia	Tennis	I	2016

(5)Through our query results on the official website, we found real-world examples of the 'great coach' effect and the specific transfer times of coaches. We discovered that the results we calculated include these examples, which proves that our model can predict and find evidence of the possible 'great coach' effect.

Table 10 evidence of the possible 'great coach' effect

Coach	Sport	Year	ξ_i	ζ_j
Kim Chang Back	hockey	1999	Korea	China
Ping Lang	Volleyball	2008	China	United States
Anastasia Bliznyuk	rhythmic gymnastics	2022	Russia	China

6.2.2 Quantify the Impact of the “Great Coach” Effect

(1)Based on the overall growth trend of Olympic events in each edition, which naturally leads to an increase in the number of medals, we define the natural growth without changing coaches as P(t), and the impact brought by the 'great coach' as G. The calculation formula is as follows:

$$G = \Delta H - P(t) \tag{13}$$

(2)To quantify P(t), we introduce a factor in the logistic growth model that represents fluctuations in the number of medals. The growth model with a fluctuation adjustment factor is constructed as follows：

$$\frac{dP(t)}{dt} = r \bullet P(t)\left(1 - \frac{P(t)}{K}\right) - d(t) \bullet P(t) \tag{14}$$

P(t) is the number of medals won by the country at time t.
r is the medal growth rate, which represents the speed at which the number of medals increases over time.
K is the maximum carrying capacity of medals, i.e., the maximum number of medals the country can ultimately win.
d(t) is a time-varying adjustment factor that represents the decrease or negative influence on the number of medals. It may be related to factors such as declining performance, policy changes, insufficient training, etc. It is a function that changes over time.
t is time
(3)Similarly, using the countries and events (Sport) with the 'great coach' effect as examples, the partially obtained results are displayed as follows:

Table 11 the partially obtained results

Sport	ΔH **Country**	ΔH	P(t)	G
Volleyball	China	-0.3406	0.171034	-0.51163
Volleyball	United States	0.9349	-0.031291	0.966191
hockey	South Korea	-0.266	0.170335	-0.43634
hockey	China	0.5738	0.011427	0.562373
rhythmic gymnastics	Russia	-0.6778	-0.066923	-0.61088
rhythmic gymnastics	China	0.6689	-0.003489	0.672389

Where G represents the predicted contribution of the 'great coach' effect to the number of medals. From the results, it can be observed that, in general, the departure of all 'great coaches' has a negative contribution to the medal count of the country they leave, meaning the number of medals in that event will decrease in the leaving country. Conversely, the arrival of the 'great coach' has a positive contribution to the medal count of the country they arrive at, meaning the number of medals in that event will increase in the arriving country

6.2.3 Formulating Investment Strategies

Based on the model we constructed, we select the events with low HHH values in the last three Olympic Games, which should be the events to invest in that have the 'great coach' effect. We have compiled the HHH values for each event of all participating countries from 1896 to 2024, and we present the results for three countries as follows

Table 12 the results for three countries

Country	**Sport**	H_{2016}	H_{2020}	H_{2024}
Algeria	Weightlifting	0.0324	0.0162	0.0162
Canada	Shooting	0.0486	0.0324	0.0648
Vanuatu	Judo	0.0162	0.0162	0.0162

This indicates that Algeria should assign better coaches in the weightlifting event, Canada in shooting, and Vanuatu in judo, as this would help improve their performance in these events and maximize the returns.

7 Other Insights and Implications

(1)Medal Count is Affected by Multiple Factors
The model we constructed to predict medal counts shows that Olympic medal totals are influenced not only by historical performance and athlete numbers but also by factors such as training systems, coaching levels, and athletes' psychological resilience. The Olympic Committee should focus on improving coach quality and optimizing resource allocation rather than simply increasing the number of athletes. By analyzing potential events and formulating reasonable strategies, the medal count can be effectively increased.
(2)Resource Investment Affects Medal Distribution
The model we built to analyze events and medal counts shows that medal distribution reflects a country's sports development strategy. A diverse medal distribution indicates stable performance across multiple events. The Olympic Committee can adjust resource allocation based on the medal distribution, focusing efforts on more challenging gold medal events, stabilizing performance in established strength areas, and increasing investment in the events with the highest medal potential.
(3)Medal Gap is Affected by Long-Term Development Strategies
The model we constructed to predict first-time medal-winning countries shows that the medal gap is a result of long-term accumulation. This is particularly true for countries that have not yet reached the top level, where short-term improvements are difficult. Long-term investment in youth sports, coach development, and infrastructure construction is crucial. The Olympic Committee should focus on long-term development, allocate resources effectively, strengthen strong events, cultivate internationally competitive athletes, and improve overall medal counts.
(4)The 'Great Coach' Effect Has a Significant Impact on Medal Counts
The model we developed to quantify the 'great coach' effect shows that excellent coaches can greatly enhance athlete performance, especially in critical moments to unlock potential. The experience, tactics, and psychological guidance of coaches have a significant impact on medal counts. The Olympic Committee can use the analysis of the 'great coach effect' to identify and invest in high-quality coaching talent, driving growth in medal counts for relevant events.

8 Model Evaluation

8.1 Strengths

- **Comprehensiveness**

  In our model, we comprehensively considered factors such as the number of participants, historical best medal rankings, host country effect, and other factors as indicators to predict the number of medals. By taking multiple factors into account, we ultimately established the entire model, which is convincing.。

- **Innovation**

  Unlike conventional prediction methods, we used the Random Forest ensemble learning algorithm to predict the medal counts of countries in the 2028 Olympics. This method allows for the adjustment of the training set proportion and the sampling proportion to optimize the

training results and find the best solution, demonstrating innovation. Furthermore, using Spearman's rank correlation coefficient, we obtained the correlation between different events in each country and the country's medal counts. Based on this, we conducted an analysis to explore the importance of different events for each country, showcasing both scientific rigor and innovation.

- **Scientific Rigor**

When we used our model to predict the countries that will win their first medal in the next Olympics, we considered that countries that have already won medals in previous Olympics are unlikely to win their first medal again in the future. Therefore, we excluded these countries from the statistics and only counted the proportion of countries that will win their first medal and those that have participated but never won a medal. Considering the large difference in the order of magnitude between the independent and dependent variables, which could affect the regression performance, we normalized both to bring them to the same scale. Finally, we used multiple regression models for comparison and selected the model with the best fit for the regression, ultimately obtaining good results.

8.2 Weaknesses

- **Subjectivity**

In the model for evaluating and quantifying the effect of 'great coaches,' when considering the weight of each factor, we subjectively set the numbers in the judgment matrix. Therefore, there might be some flaws in terms of objectivity and rationality.

- **Idealism**

The actual Olympic medal count may also be influenced by a series of other factors, such as unexpected injuries to athletes, international factors, or political influences. However, we are unable to account for such detailed and comprehensive factors, so the predictions may have some discrepancies with the actual values.

References

[1] Jiang, L. (2007). *A model for predicting Olympic medal counts based on historical data*. Journal of Quantitative Analysis in Sports, 3(2), 89-101.

[2] Shi, Z., & Zhang, X. (2024). *Application of Grey Forecasting Models in Predicting Olympic Medal Numbers*. International Journal of Forecasting, 40(1), 45-58.

[3] Chien, C. F., & Tsai, H. H. (2012). *Predicting Olympic medal counts using regression models*. Computational and Mathematical Methods in Medicine, 2012, 1-8.

[4] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

[5] Breiman, L. (2001). *Random forests*. Machine Learning, 45(1), 5-32.

[6] Goh, M., & Lee, S. (2013). *Prediction of Olympic medals using machine learning techniques*. Proceedings of the 2013 IEEE International Conference on Data Mining, 1011-1016.

[7] Coyle, D., & Stocking, G. (2017). *Analyzing the impact of sports coaches on Olympic medal performance*. Journal of Sports Economics, 18(5), 522-539.

[8] Wang, Y., & Xie, Y. (2019). *Impact of coaching changes on the performance of national sports teams: A case study of Olympic medal counts*. International Review of Sports Economics, 13(4), 364-381.

[9] Li, Y., & Zhang, W. (2018). *Effectiveness of national coaching systems on Olympic performance: Evidence from multiple countries*. European Sport Management Quarterly, 18(2), 200-218.

[10] Langer, S., & Schwarz, P. (2014). *The role of sports coach effectiveness in Olympic success: A statistical analysis of medal counts*. Sports Science Review, 23(2), 107-121.