Problem Chosen	2025	Team Control Number
C	MCM/ICM Summary Sheet	2516010

Olympic Medal Prediction: Multi-Model Fusion Unlocks the Password of Medal Table

As the most influential sports event in the world, the Olympic Games constitute a dynamic and complex large-scale system from the perspective of system science. It is of wide and far-reaching practical significance to scientifically and accurately construct the prediction model of competition results and reveal the law behind the medals. Therefore, this study focuses on the prediction of Olympic medal table, aiming to build a comprehensive and accurate prediction model, analyze the relevant influencing factors in depth, and provide practical strategic suggestions for national Olympic committees. We established three models: Model I: Basic medal prediction integrated model; model II: "great coach" effect validation model; model III: Olympic strategy insight model.

Model I is mainly devoted to medal list prediction. We extensively collect multi-source data, And robust standardization and iterative missing value filling technology are used to clean the data and carry out feature engineering. In view of the shortcomings of traditional methods, we integrate GBM, RF etc., to build an integrated model and improve the prediction ability by reasonable weighting. The model is used to predict the medal table of the 2028 Olympic Games, analyze the relationship between the project setting and the number of medals and the advantages of the host country, and analyze the medal data from multiple dimensions.

Model II revolves around the "great coach" effect. After outlier processing, design matrix construction and singular value decomposition, the performance consistency evaluation model is established. By calculating the trend line, residual analysis and consistency score, the effectiveness of the coach's guidance in key projects is judged. Then, in the improvement potential model, the relative gap, trend factor and historical factor are defined, the comprehensive score and final score are calculated, the improvement potential of specific projects in different countries is measured, and the time period of main participating countries and coach effect can be analyzed.

Model III focuses on deep mining of Olympic medal data. The clustering algorithm is used to divide the national sports development level, the slope of the data point is calculated to clarify the trend of the number of medals, and the Gini coefficient is introduced to measure the equilibrium degree of medal distribution. Analyze data from multiple dimensions to help the Olympic Committee adjust resource allocation; calculate the medal concentration to make it clear the competitive position; identify emerging and recessionary countries to provide empirical warnings; regional analysis helps to combine regional characteristics planning and development; calculate various scores to assist in assessing their own development trend.

These three models comprehensively reflect the dynamic complexity of Olympic medal acquisition. The combination of in-depth data analysis and advanced modeling technology provides a new perspective and method for Olympic medal prediction, which has important guiding significance for the National Olympic Committee to formulate a scientific and reasonable sports development strategy.

Keywords: Basic Medal Prediction Integrated Model; "Great Coach" Effect; Olympic Strategy Insight; Dynamic Complexity

Contents

1 Introduction

1.1 Problem Background

The Olympic Games, a globally renowned sports event, are a platform for countries to show their athletic strength. Nations strive for excellent results. In the 2024 Paris Olympics, the medal standings' top positions got much attention, and some small countries overcame their zero-medal situation. Predicting the Olympic medal table matters as it reflects a country's sports power and affects its international status, so it's crucial for national Olympic committees. But predicting medal numbers is tough due to multi-dimensional and non-linearly related influencing factors. Thus, a more comprehensive prediction model is needed.

1.2 Restatement of the Problem

Our objective is to develop a quantifiable and highly accurate prediction model for the Olympic medal table. This model aims to comprehensively explore the advantages enjoyed by the host country, the presence of the "great coach" effect, and the impacts of various factors, including event-specific data and athlete-related aspects. The research will be conducted through the following steps:

- Model Construction and Evaluation Verification: Gather relevant data from various countries, process it with specific techniques, build a multi-model integrated prediction model, determine weights via multiple regression, explore causal relationships, use cross-validation, calculate error metrics for uncertainty analysis, and provide a prediction interval.
- Results Analysis and Application: Predict the 2028 LA Olympics medal table, compare with history to assess countries' performance changes, predict new medal-winning countries, analyze event-medal impacts, and assist national Olympic committees in strategy-making and event-selection.
- Verification of the "Great Coach" Effect: Establish a hypothesis-testing framework, collect national medal data before and after the tenure of renowned coaches, quantify the performance impact, verify the existence and extent of the "great coach" effect, and provide data-based support for talent development strategies.
- Strategic Recommendations: Develop an Olympic strategy insight model, summarize insights related to Olympic medal counts revealed by the model, and offer decision-making references for national Olympic committees.

1.3 Our work

To tackle this challenge, we integrate GBM, RF, XGBoost, and LightGBM to build an integrated model. Key features are extracted from multi-source Olympic data. Model parameters are optimized, and the models are combined with cross-validation training for better prediction. We refine the model via historical data backtracking and new data verification, creating a basic medal prediction model. A "great coach" effect model, composed of performance consistency and improvement potential sub-models, is then

developed for stability and accuracy. Lastly, we merge KMeans, Theil-Sen regression, and Gini coefficient models into an Olympic strategic insight model. The final model accurately predicts medal counts and analyzes influencing factors, offering support for national sports strategies and enhancing Olympic competitiveness.

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

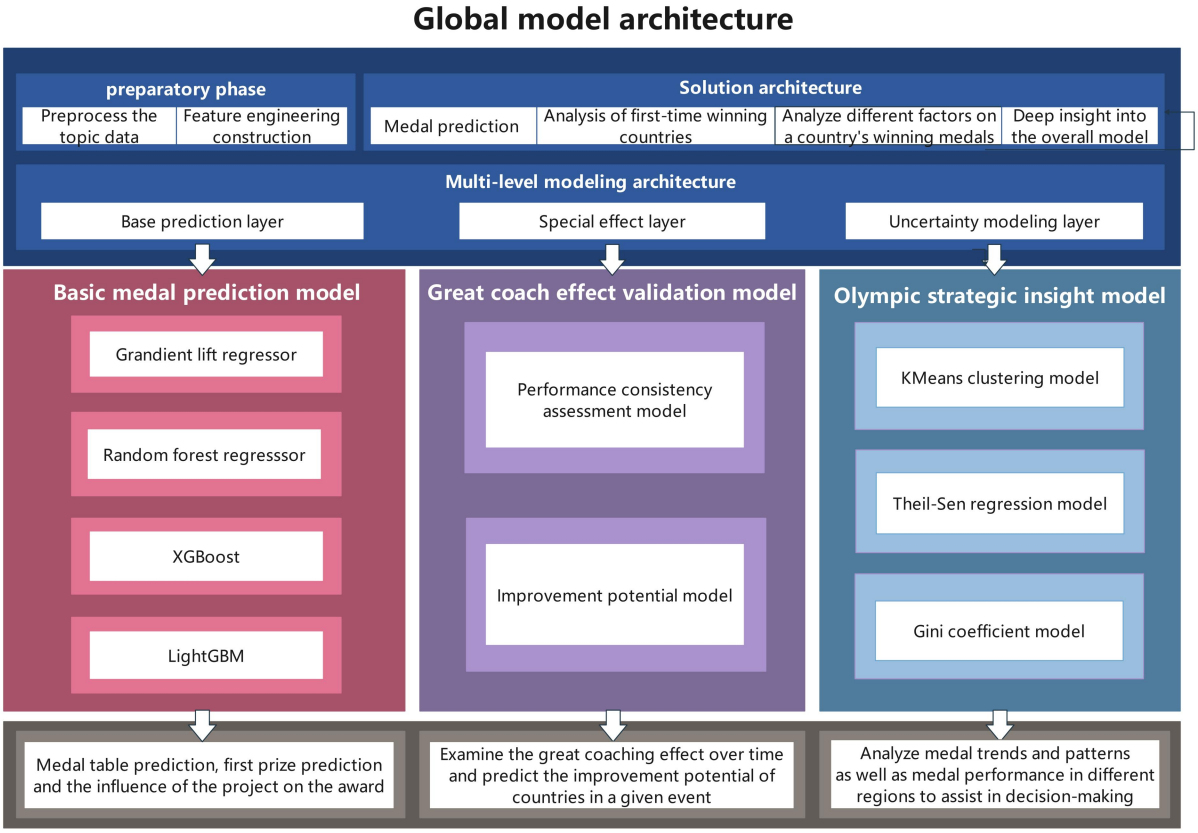

Figure 1: Model Overview

2 Assumptions and Justifications

To simplify the problem, we make the following basic assumptions, each of which is properly justified.

- **Assumption 1: In-cycle Sports Condition Stability.**

Justification: We assume sports development conditions in each country stay stable within an Olympic cycle. This simplifies medal-count change analysis, letting us focus on event factors. Ignoring short-term fluctuations helps isolate direct determinants.

- **Assumption 2: Inferring National Sports Strength from Historical Olympic Data.**

Justification: We postulate a country's sports strength can be inferred from historical Olympic data. Analyzing it helps identify key strength factors. Quantifying them aids in establishing objective indicators for medal prediction.

- **Assumption 3: Accuracy of Research Data.**

Justification: We assume the research data is accurate, without significant errors or fabrication. Inaccurate historical medal records can bias model predictions, and obtaining data for the 2028 Olympics is challenging. This assumption is crucial for accurate medal-count prediction.

- **Assumption 4: Stability of the External Environment**

Justification: We assume the external environment remains stable during the study and forecast. External factors like economic, trade, and sports-tech changes can disrupt events. If not considered or assumed stable in the model, they will undermine medal-count prediction accuracy.

3 Notations

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

Table 1: Notations used in this paper

Symbol	Description
n	Total number of countries
i	The number of times a country has participated in the Olympic Games
c	The number of different countries
$t_{c,\ i}$	The year in which the country c participated in the Olympic Games for the *i*-th time
$T_{c,t_{c,i}}$	The total number of medals won by the country c in that year
b	The average moving window size
$\hat{y}_i$	Medal prediction value
y_i	The true value of the medals
$\bar{y}_i$	The average of the actual number of medals for each country
m	The number of decision trees

4 Model Preparation

4.1 The Data

Since the amount of data is large and not intuitive, we directly visualize some of the data for display.

4.1.1 Data Collection

The data we used mainly include the 2028 project settings, the population of each country, the economic situation, and the resume of the "great coach" data. The data sources are summarized in Table 2.

Table 2: Data source collation

Database Names	Database Websites Data	Type
Olympics	https://olympics.com/	Physical、News
UNSD	http://unstats.un.org/unsd/default.htm	Report

4.1.2 Data Cleaning

When reviewing historical Olympic data, we identified missing values in economic investment, athlete numbers in niche events, and sport-specific competitiveness indicators. For numerical data, we use mean imputation for minor missing values. For major ones, we estimate based on a country's economic and sports development. For athlete-and event-related data, we first try to supplement missing values with relevant materials. If unavailable, we interpolate using trends and data from similar countries.

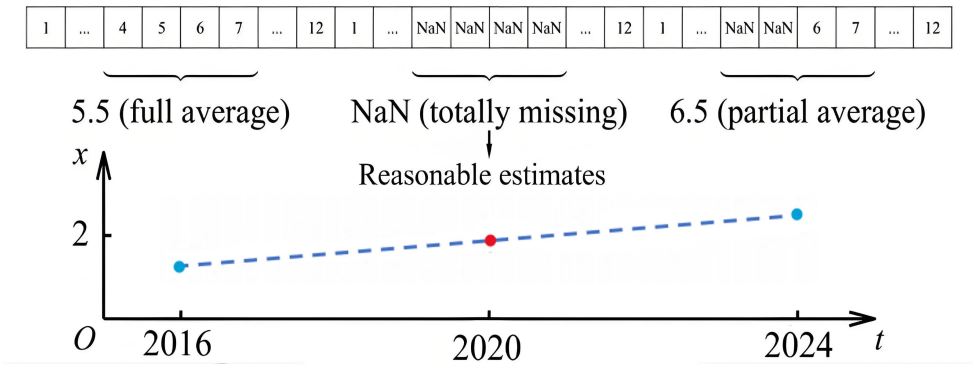

Figure 2: Data cleaning

5 Model I: Basic Medal Prediction Model

The basic medal prediction model, a core part of the Olympic medal prediction system, needs to accurately forecast the number of gold and total medals per country. Analyzing historical data and literature reveals that medal-winning factors are multi-dimensional with time-series traits. A single model can't meet the need for capturing static and dynamic

features, so we propose multi-model integration. We combine methods like Gradient Boosting Regressor (GBM), Random Forest (RF), XGBoost, and LightGBM. By integrating their predictions and using an optimization algorithm to set weights, the final model can better capture data features. This improves accuracy and robustness, making it more suitable for complex medal-number predictions.

5.1 Gradient Boosting Regressor

The Gradient Boosting Regressor uses decision trees as base learners. Based on the gradient boosting algorithm, it iteratively fits the residuals of the previous model to improve performance. Each iteration corrects the previous model's prediction errors, getting the model closer to the true value. The model formula is

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \tag{1}$$

Where $F_m(x)$ is the model's predicted medal count (gold or total medals) for a country in a specific Olympics after m iterations. $F_{m-1}(x)$ is the previous iteration's result. The γ learning rate is set to 0.05 in the code for more accurate predictions. $h_m(x)$ is the m-th decision tree. The decision tree analyzes factors like athlete numbers and event settings to learn their relationships with medal counts, providing a basis for model prediction.

After obtaining the predicted values, the loss function is used to measure the difference between the predicted values and the true values. In regression problems, a commonly used loss function is the Mean Squared Error, which is expressed in the following form

$$L(y, F(x)) = \frac{1}{2}(y - F(x))^2 \tag{2}$$

For gradient boosting, we need to calculate the gradient of the loss function with respect to the current model $F_{m-1}(x)$. That is, we take the derivative to represent the gap between the predicted value $F_{m-1}(x)$ and the true value y. The formula is expressed as

$$\nabla F_{m-1}(x) = -(y - F_{m-1}(x)) \tag{3}$$

Then, we fit a new model $h_m(x)$ to minimize the gradient. This model is usually a simple decision tree, which fits the gradient at a given data point to update our prediction model. That is to say, we go back to the first formula to complete one iteration.

5.2 Random Forest Regressor

This is an ensemble-learning-based algorithm. It constructs multiple decision trees and combines their prediction results for the final prediction. When building each decision tree, the random forest uses random feature selection and sample sampling. This gives the decision trees some independence, strengthening the model's generalization ability and lowering the over-fitting risk. Its prediction formula is:

$$F(x) = \frac{1}{M}\sum_{m=1}^{M} T_m(x) \tag{1}$$

Here, $F(x)$ is the final prediction value of the model, which is obtained by integrating the prediction values $T_m(x)$ of M decision trees.

5.3 XGBoost

It is similar to the Gradient Boosting Regressor. The difference lies in that it introduces a regularization term into the objective function. The addition of the regularization term can effectively control the complexity of the model, preventing the model from over-fitting during the training process, and thus improving the stability and generalization ability of the model. The formula of the objective function is:

$$Obj = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \quad (1)$$

$\sum_{i=1}^{n} l(y_i, \hat{y}_i)$ is the loss function, which is used to measure the difference between the predicted medal value $\hat{y}_i$ and the true value y_i, and still uses the SME. $\sum_{k=1}^{K} \Omega(f_k)$ is the regularization term, which is used to control the complexity of the model. f_k represents the k-th decision tree. By constraining the complexity of the decision tree, the model can be prevented from over-fitting the training data.

5.4 LightGBM

LightGBM uses histogram algorithms for high training efficiency and low memory use. With large-scale Olympic data, it converges fast and predicts accurately. The GOSS sampling in this paper accelerates training by keeping large-gradient samples and some small-gradient ones, cutting load while maintaining performance.EFB bundles mutually exclusive features for feature selection and dimensionality reduction. The formula is:

$$gain = \frac{1}{n} \sum_{i=1}^{n} (g_i^2 / h_i) \quad (1)$$

$gain$ represents the importance of features, used to evaluate the contribution of features to the model's performance. g_i indicates the gradient of the i-th sample, reflecting the error of the model on that sample. h_i indicates the second derivative of the i-th sample, reflecting the curvature of the model on that sample.

5.5 Integrated model

To determine the weights of each base model, an optimization objective needs to be defined. This objective is to minimize the difference between the predictions of the ensemble model and the true values. The formula is as follows:

$$min_w \sum_{i=1}^{n} (y_i - \sum_{k=1}^{K} w_k F_k(x_i))^2 \quad (1)$$

x_i is the feature vector of the i-th sample. $F_k(x_i)$ is the prediction of the k-th base model for the input x_i. Additionally, the weights must satisfy the following constraints, that is,

the sum of all weights must equal 1 and be non-negative:

$$\sum_{k=1}^{K} w_k = 1, w_k \geq 0 \tag{2}$$

Finally, integrate the model to generate the final prediction results. The formula is as follows:

$$F_{ensemble}(x) = \sum_{k=1}^{K} w_k F_k(x) \tag{3}$$

$F_{ensemble}(x)$ is the final medal prediction of the ensemble model for the input x. K is the number of base models. w_k is the weight assigned to the k-th base model.

5.6 Results

By applying the above-mentioned ensemble model, we can predict and conduct in-depth analyses of the total medal counts and gold medal counts that countries will achieve in the 2028 Olympic Games.

5.6.1 Predictions and Comparisons

We selected the top 10 countries with the predicted number of medals for display, as shown in Figures 3 and 4, in which the numbers represent the intervals for medal count predictions. In gold medal predictions, the US, UK, China, etc., rank high. They're also predicted to get many total medals, showing their recognized sports strength.

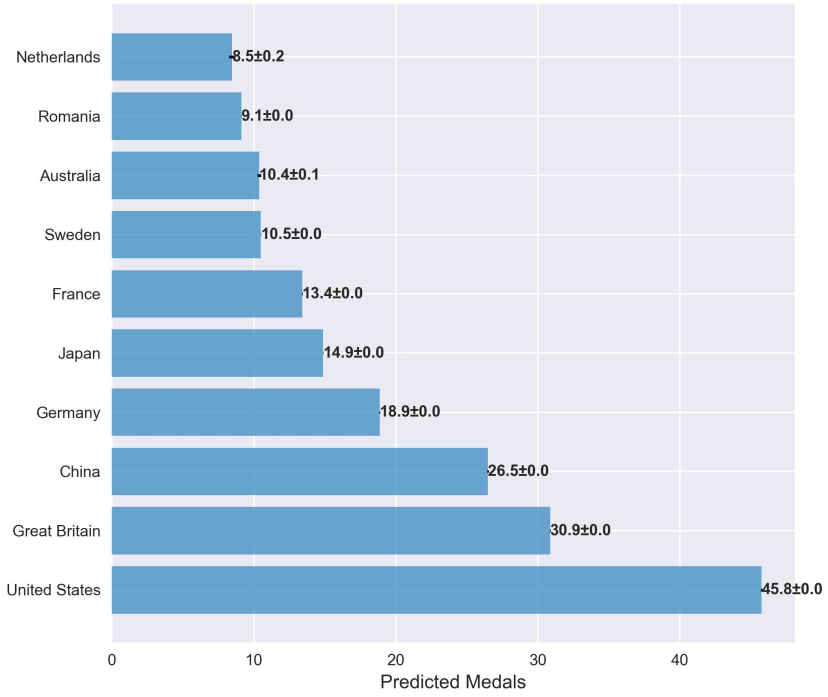

Figure 3: 2028 Gold Medal Prediction

We then analyzed countries' medal-count trends to assess sports development. The figure below shows the medal trends of the Soviet Union, CIS, the US, etc., from 1900-2020. Clearly, medal counts fluctuated greatly. The Soviet Union was outstanding at times, while the US stayed at a stable high level, reflecting the dynamic change of countries' sports strength over time, which vividly reflects the dynamic evolution of the sports capabilities of different countries over time.

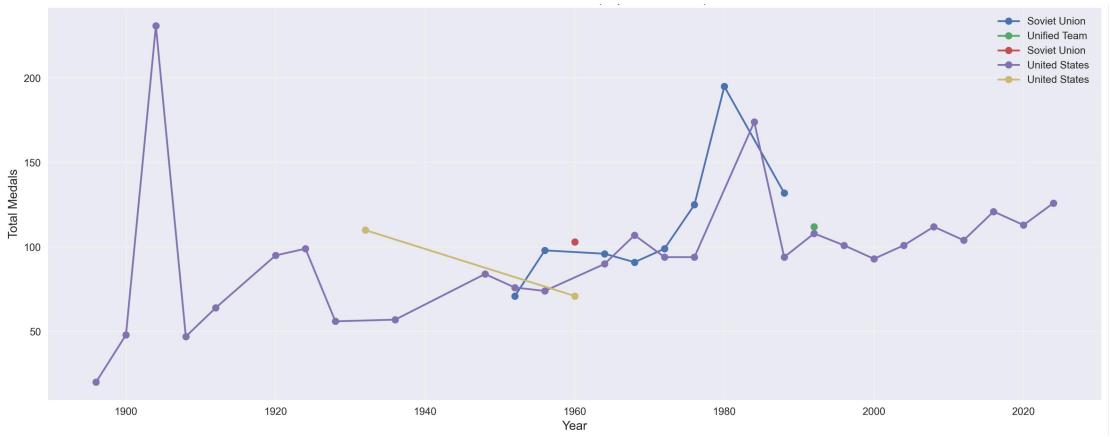

Figure 4: Historical Medal Trend (Top 5 Countries)

By comparing the actual medal counts in 2024 with the predicted ones for 2028, we found that among ten countries like the US, UK, and China, the US, UK, Germany, etc., are predicted to make progress in the next Olympics, while China, France, Japan, and Australia are expected to perform worse. Sweden is projected to make the greatest progress, with its total medal count increasing by 222.4%, as shown in the figure 5. This graph visually presents the predicted trends and percentage changes in the medal counts of various countries.

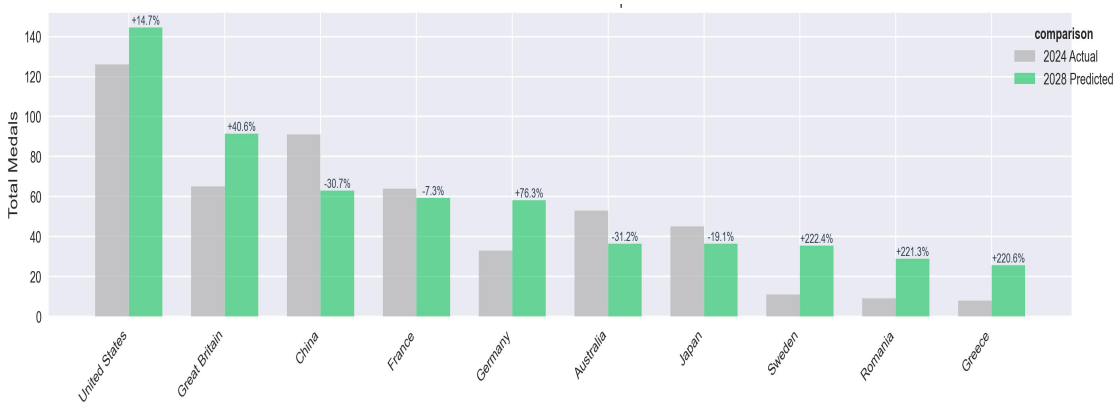

Figure 5: 2024 vs 2028 Medal Comparison

5.6.2 Zero Medal Breakthrough Prediction

To forecast countries likely to break the zero-medal barrier, data on historical Olympic participation, total athlete numbers, sports diversity, and recent performance indicators were gathered for 16 countries expected to compete in the 2028 Olympics yet with no prior medal wins. All data were standardized to eliminate dimensional disparities and ensure

comparability, as depicted in the figure presenting the correlation among multiple Olympic-related features.

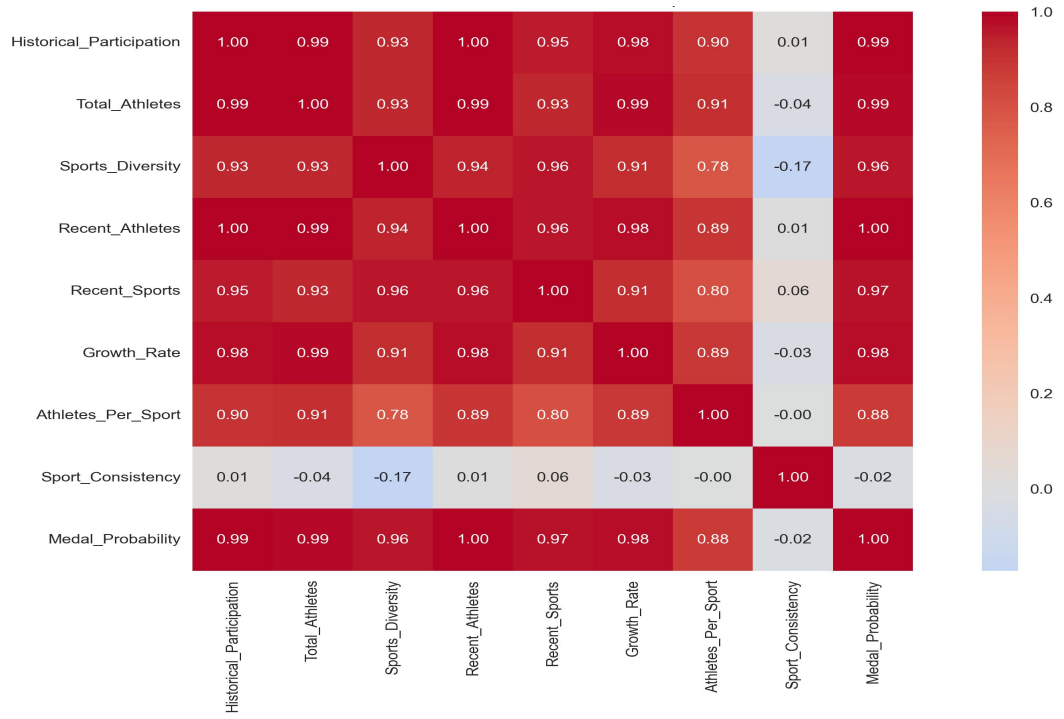

Figure 6: Feature Correlation Heatmap

Through in-depth analysis of the relevant data of each country and prediction calculations, a series of valuable outcomes were derived, and corresponding reports were compiled. The top five countries with the highest probability of winning their first medal, along with some of their pertinent indicators, are presented in the following table:

Table : The Top 5 Countries Most Likely to Win the First Medal

NOC	Historical Participation	Total Athletes	Recent Athletes	Growth Rate	Medal Probability
11 CHE	1.00	1.000000	1.000000	1.000000	1.000000
2 MDA	0.16	0.065132	0.166065	0.011140	0.182990
5 SAM	0.04	0.035260	0.115523	0.062762	0.129876
14 BOL	0.00	0.037218	0.061372	0.040795	0.095199
0 MCO	0.00	0.033301	0.028881	0.029812	0.093528

As is evident from the table, Switzerland (CHE) excels in numerous key indicators, with its probability of winning the first medal reaching as high as 1.000000, making it the most promising nation. Besides Switzerland, Moldova (MDA) and Samoa (SAM) also demonstrate high potential.

Key performance metrics indicate that Switzerland ranks first in historical participation,

recent activity levels, sports diversity, and development velocity, all of which are closely associated with medal-winning potential. The distinct advantages of Switzerland can perhaps be attributed to its stable policies, abundant resources, and well-established talent cultivation system. Countries with a low probability of winning medals, such as Bolivia (BOL) and Monaco (MCO), may have deficiencies in sports event selection and infrastructure construction. It is advisable to focus on Switzerland, Moldova and Samoa, as they possess advantages in medal-winning probability and sports event diversity. Sustained attention to these countries will contribute to a deeper understanding of the successful experiences and underlying laws of sports development. Simultaneously, it is recommended that countries with low sports diversity expand their scope of participation, enhance participation levels and the number of athletes, and reinforce their training systems.

5.6.3 Correlation Between Countries and Events

In the analysis, we noticed that some countries' medal-winning is concentrated in certain events. The figure 7 shows the medal distribution of different countries in key sports events. The horizontal axis represents countries, and the vertical axis represents sports events. Different colors indicate the number of medals each country won in the corresponding events. This graph clearly demonstrates the strengths and weaknesses of various countries in different sports events. For example, the UK is strong in athletics. The differences in medal distribution among countries in various events reflect the focus and characteristics of each country's sports development.

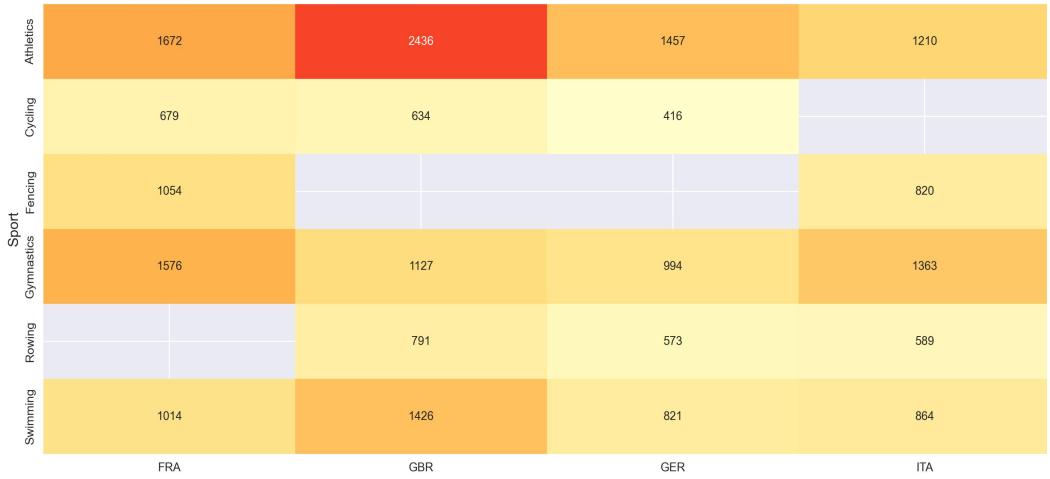

Figure 7: Key Sport Distribution by Country

Therefore, we further investigated the relationship between sports events and the number of medals won by countries. The figure 8 shows the correlation coefficients between some countries and specific events. It can be seen that there is a relatively high correlation between them. For example, the correlation coefficient of China is 0.858.

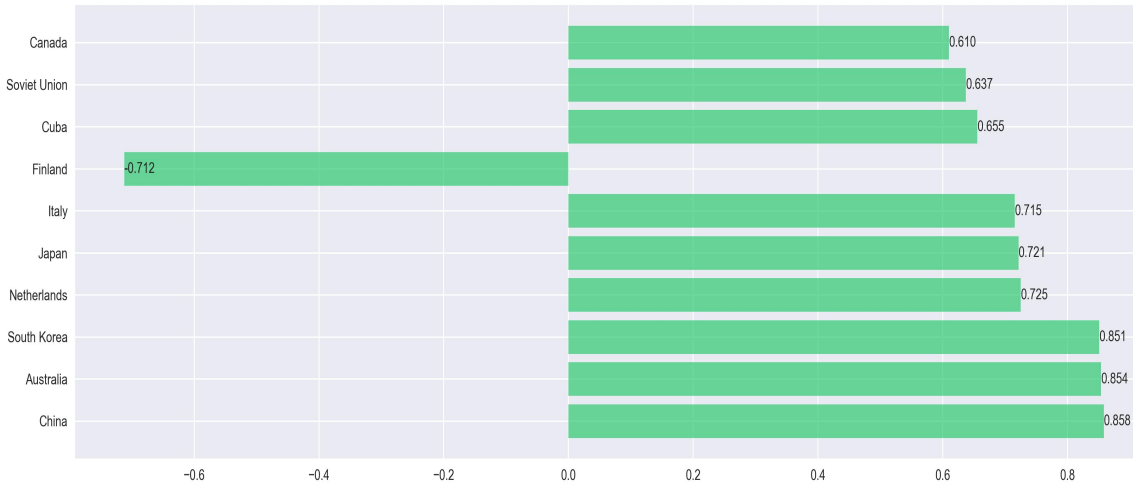

Figure 8: Country-Specific Events-Medals Correlation

Subsequently, we determine the relationship between the number of medals in each country and the project through the concentration of sports events. As shown in the figures 9 and 10, the higher the percentage of medals won by the top-ranked projects, the more concentrated the country 's advantages in these projects. The global ranking distribution reflects the international competitiveness of countries in different projects. The lower the ranking, the stronger the global competitiveness in the project. By comparing the two analyses, we can comprehensively evaluate the development balance and competitiveness of sports in various countries.

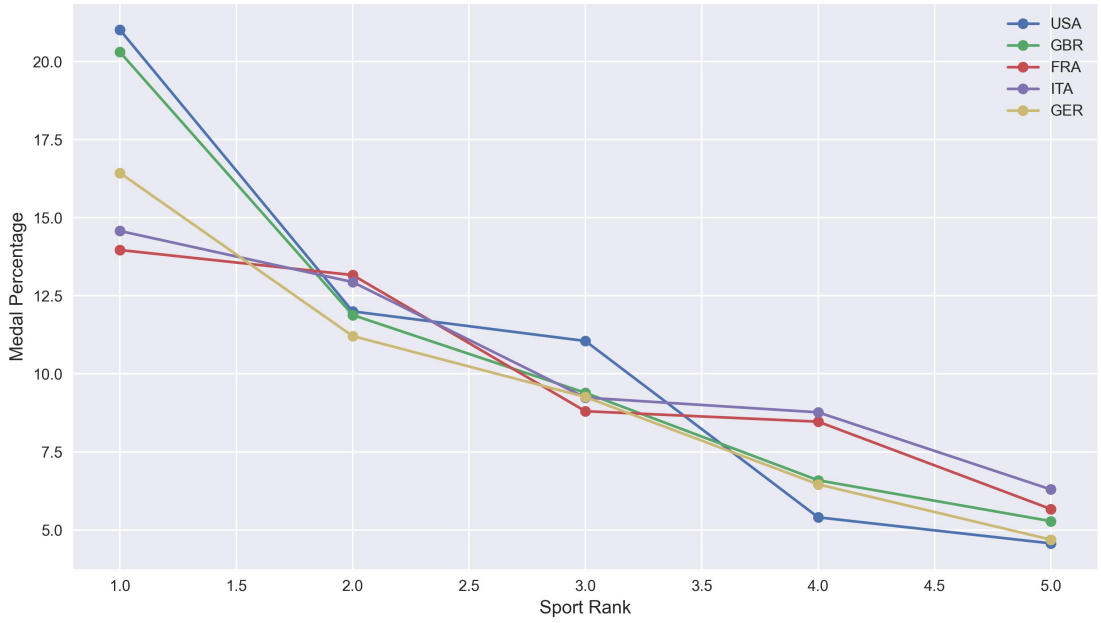

Figure 9: Sport Concentration Analysis

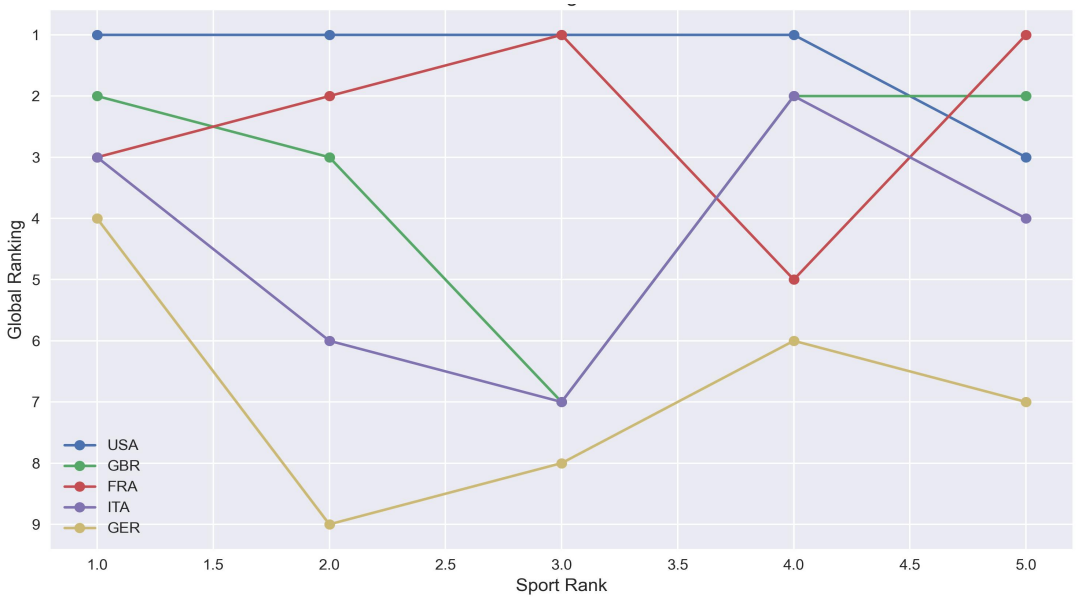

Figure 10: Global Ranking Distribution

6 Model II: Great Coach Effect Validation Model

Based on the analysis of the topic data, in order to analyze the main participating countries in each sport, to detect the time period of the coaching effect, and to calculate the improvement potential of different countries in specific sports, we divide it into the following parts to construct the model of the great coach effect validation.

6.1 Performance Consistency Assessment Model

Firstly, the outlier is handled, and then the design matrix X, $X = \begin{bmatrix} Y \\ 1 \end{bmatrix}^T$, where Y is the year vector, which includes all the years in which a country participates in the Olympic Games, and **1** is the all-1 vector. SVD decomposition of X: $X = USV^H$, where U and V are orthogonal matrices, S is diagonal matrix, and the coefficient vector of linear regression is C, and M_s. represents the medal data matrix composed of the gold, silver, bronze medals and total medals won by a country in each year of the Olympic Games, then there is:

$$C = V^T diag\left(\frac{1}{S}\right) U^T M_s \tag{1}$$

Defining the trend line Tr, there is:

$$Tr = C[0] \times Y + C[1] \tag{2}$$

Note that if $\frac{min(s)}{max(s)} > 10^{10}$, the matrix S is considered ill-conditioned. The Theil-Sen regression model is used to calculate Tr. Subsequently, residual analysis is performed, and the residuals R and the residual standard deviation σ_R are calculated as follows:

$$R = M_s - T \tag{3}$$

$$\sigma_R = std(R) \tag{4}$$

Finally, calculate the consistency score, let the base score of a certain country be B_c, the trend weight be W_{Tr}, and the final score be F_c, then we have:

$$B_c = \frac{1}{1+\sigma_R} \tag{5}$$

$$W_{Tr} = |corr(Y, M_s)| \tag{6}$$

$$F_c = 0.7B_c + 0.3W_{Tr} \tag{7}$$

The final score comprehensively reflects the advantages of stability and trendiness in key projects. The higher the final score, the more effectively it demonstrates the effectiveness of the coach's guidance.

6.2 Improvement Potential Model

After selecting the development project, we first define the following unknowns: T_r represents the recent performance of the target country (the country that needs improvement) on the project; B_r represents the recent performance of the benchmark country on the project; T_t represents the development trend of the target country on the project; B_t represents the development trend of the benchmark country on the project; T_h represents the historical fit of the target country on the project; B_h represents the historical fit of the benchmark country on the project. From this, we can deduce the relative gap R_g:

$$R_g = \frac{B_r - T_r}{B_r + 1} \tag{1}$$

Trend factor T_f:

$$T_f = tanh\big(max(0, B_t - T_t)\big) \tag{2}$$

Historical factor H_f:

$$H_f = \frac{B_h}{T_h + 0.1} \tag{3}$$

Comprehensive score P:

$$P = 0.5R_g + 0.3T_f + 0.2min\big(H_f, 2\big) \tag{4}$$

Final score F_p:

$$F_p = clip(P, 0, 1) \tag{5}$$

6.3 Results

6.3.1 Analysis and Verification Results of the Great Coach Effect

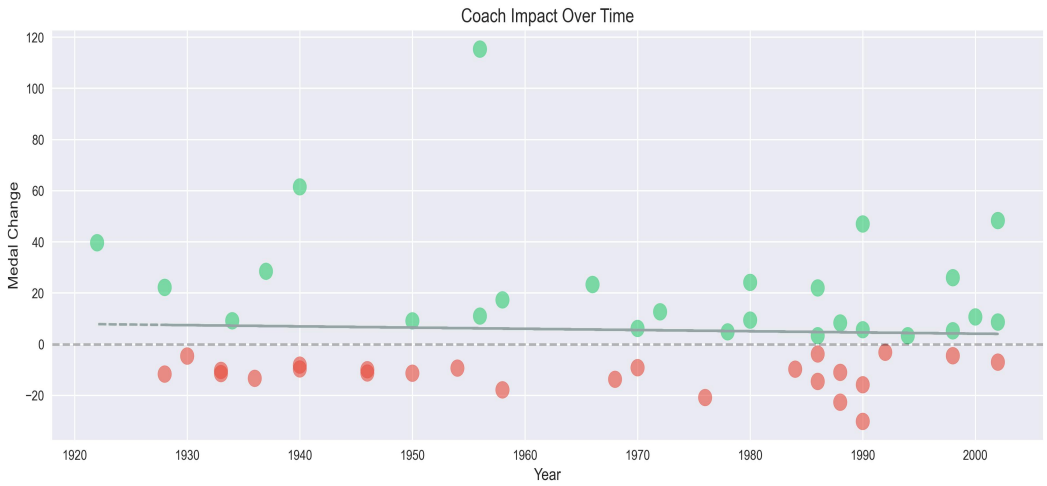

Figure 11: Coach Impact Over Time

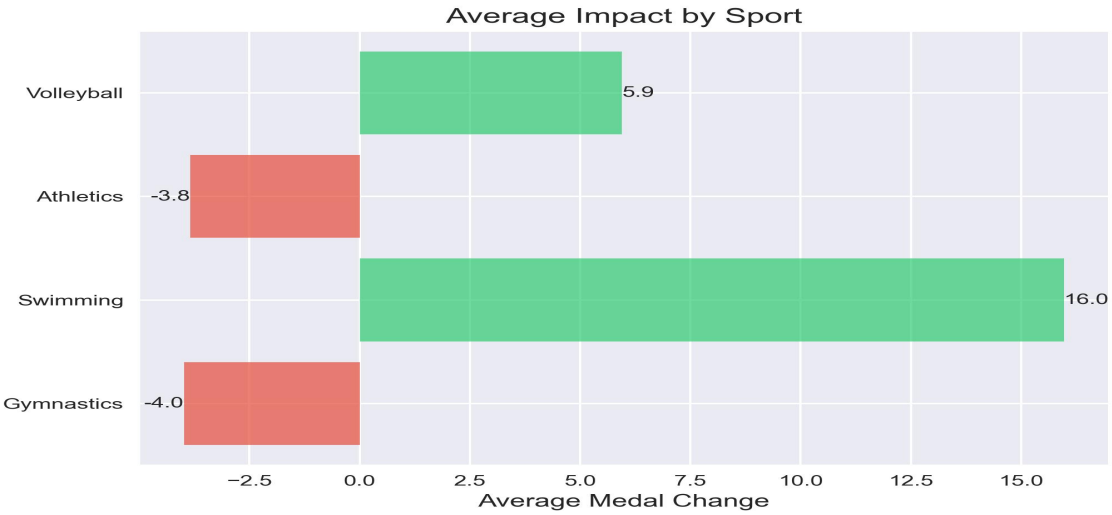

Figure 12: Average Impact Over Time

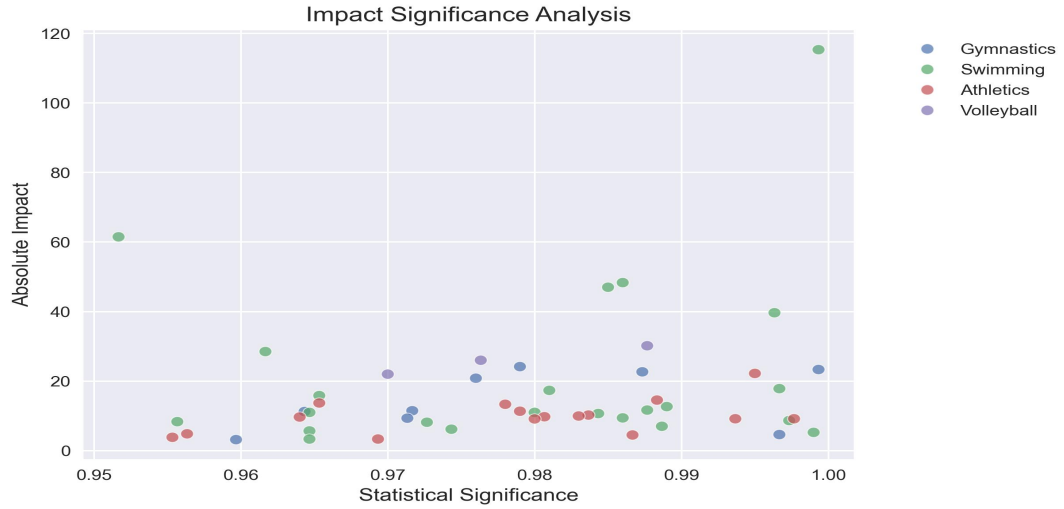

Figure 13: Impact Significance Analysis

Figure 11, "Coach Impact Over Time": The x-axis is from 1920-2000, y-axis is Medal Change. A scatter plot shows coaches' impact on medal number changes. Medal changes fluctuate around 0 in most years, with a significant positive change (110-120) around 1950-1960, indicating good coaches can boost athlete performance.

Figure 12, "Average Impact by Sport": It shows coaches' average impact on medal changes in Volleyball, Athletics, Swimming, and Gymnastics. Swimming has the greatest positive impact (16.0), while Athletics (-3.8) and Gymnastics (-4.0) have negative impacts, meaning coaches' impact is sport-related.

Figure 13, "Impact Significance Analysis": The x-axis is statistical significance, y-axis is absolute impact. Points of different colors represent different sports. The scatter of points shows coaches' influence significance and absolute impact vary greatly across sports. This helps determine the reliability of coaches' influence. Concentrated data points in high-significance areas mean more credible influence, while in low-significance areas, there's greater uncertainty. It helps sports management and coaching teams clarify resource investment focus.

6.3.2 Analysis of the Investment Results of the Great Coaches Project

Figure 14: Expected Medal Returns

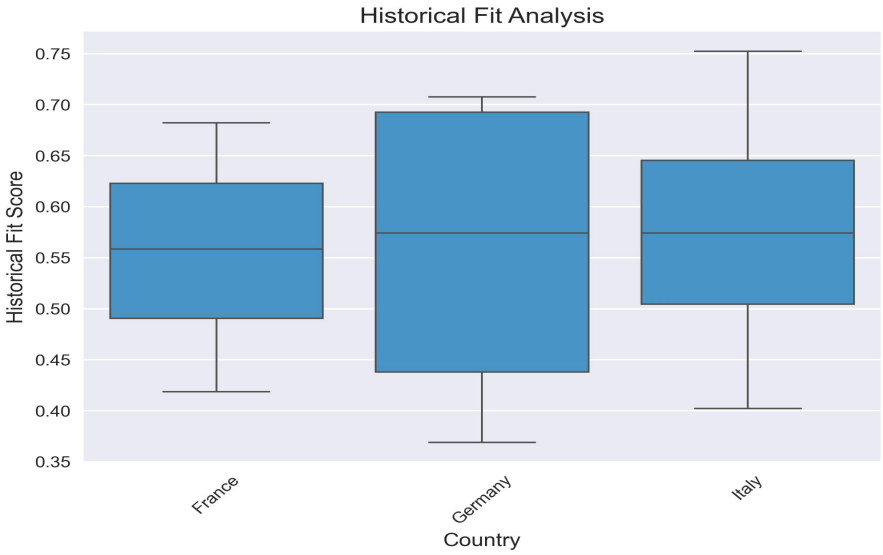

Figure 15: Historical Fit Analysis

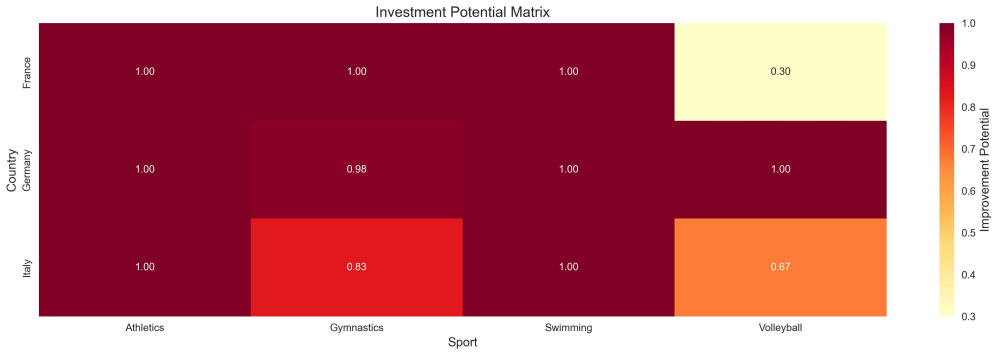

Figure 16: Historical Fit Analysis

We analyzed France, Germany, and Italy. Figures 15-17 show the results.

Figure 14 ("Expected Medal Returns"): In swimming, athletics, gymnastics, and volleyball, all three countries are expected to get over 10 medals in swimming, and their medal expectations in different events are similar.

Figure 15 ("History Fit Analysis"): Germany's historical fit scores are high and concentrated, while France and Italy's distributions differ. Based on this, the three countries can formulate strategies.

Figure 16 ("Investment Potential Matrix"): In athletics and swimming, all three countries have a potential value of 1.0. In gymnastics, Germany has 0.98, Italy 0.83; in volleyball, Italy has 0.67, France 0.30. So, France should invest in athletics, gymnastics, and swimming; Germany can consider all four sports; Italy should invest in athletics and swimming.

7 Model III: Olympic Strategic Insight Model

Through the integration of these three models, we can analyze the trends and patterns of medals, including calculating medal concentration, identifying emerging and declining countries. We can also analyze the performance of medals in different regions, including calculating regional development indices, regional stability, and dominant sports. Finally, we calculate each country's medal trend score, stability score, and diversity score to generate

insights for each country, providing decision-making recommendations for each National Olympic Committee.

7.1 Establishment of Olympic Strategic Insight Model

7.1.1 KMeans clustering model

To gain insights into the Olympic strategy, we conducted an in-depth analysis of the Olympic medal data. First, using the KMeans clustering model, we assume there are $n$countries, and x_i is the medal-related feature vector of the i-th country, where the feature vector is $[G_{i,t}, S_{i,t}, B_{i,t}, T_{i,t}]$.$\mu_j$is the center of the j-th cluster, representing the central features of countries with similar medal characteristics. $C = \{\mu_1, \mu_2, \ldots, \mu_k\}$ is the set of cluster centers, and k is the number of clusters. $\|x_i - \mu_j\|^2$ is the squared Euclidean distance from the i-th sample to the center of the i-th cluster. The objective is to minimize the sum of squared distances of all countries to the center of the cluster with similar medal characteristics, which is denoted as J.

$$J = \sum_{i=1}^{n} \min_{\mu_j \in C} \|x_i - \mu_j\|^2, \quad (i = 1,2,3, \ldots, n) \tag{1}$$

Through the result J of clustering, we can determine the state of sports development and its types in different countries, and classify the levels of sports in various countries.

7.1.2 Theil-Sen regression model

After using the Theil-Sen regression model, given a set of data points as $(t_i, T_{c,t_{c,i}})$, where $i = 1,2,3, \ldots, n$, calculate the slopes for all pairs of data points $(t_j, T_{c,t_{c,j}})$ and $(t_k, T_{c,t_{c,k}}), (j \neq k)$as follows:

$$s_{jk} = \frac{T_{c,t_{c,j}} - T_{c,t_{c,k}}}{t_j - t_k}, \quad (j = 1,2,3, \ldots, n, k = 1,2,3, \ldots, n) \tag{1}$$

Take the median of these slopes as the regression slope $\widehat{\beta}$ .Consider $T_{c,t_{c,i}} = \widehat{\beta} t_i + \hat{b}$, where $\hat{b}$ can be calculated using the mean of the sample $T_{c,t_{c,i}}$ and t_i and the estimated slope $\widehat{\beta}$ The slope $\widehat{\beta}$ obtained represents the trend of the country's total medals changing with the years, a positive slope indicates an upward trend in the number of medals, and a negative slope indicates a downward trend. Similarly, we can obtain the trends for gold, silver, and bronze medals separately.

7.1.3 Establish the Gini coefficient G

Finally, establish the Gini coefficient G , which is used to measure the degree of imbalance in the distribution of medals. The closer the value of G is to 1, the more imbalanced the distribution of medals; the closer it is to 0, the more balanced the distribution of medals.

$$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\left|T_{i,t} - T_{j,t}\right|}{2n\sum_{i=1}^{n}T_{i,t}}, \quad (i = 1,2,3,\ldots,n, j = 1,2,3,\ldots,n) \tag{1}$$

7.2 Results

7.2.1 Analysis results of medal trends and patterns

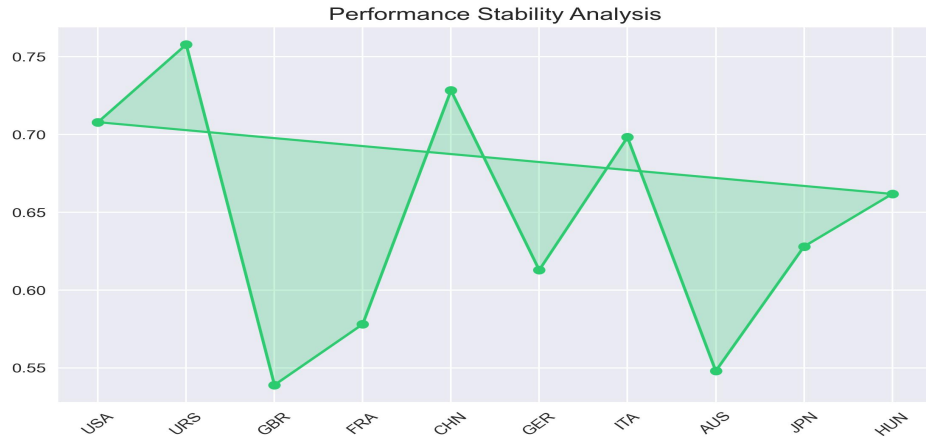

Figure 17: Performance Stability Analysis

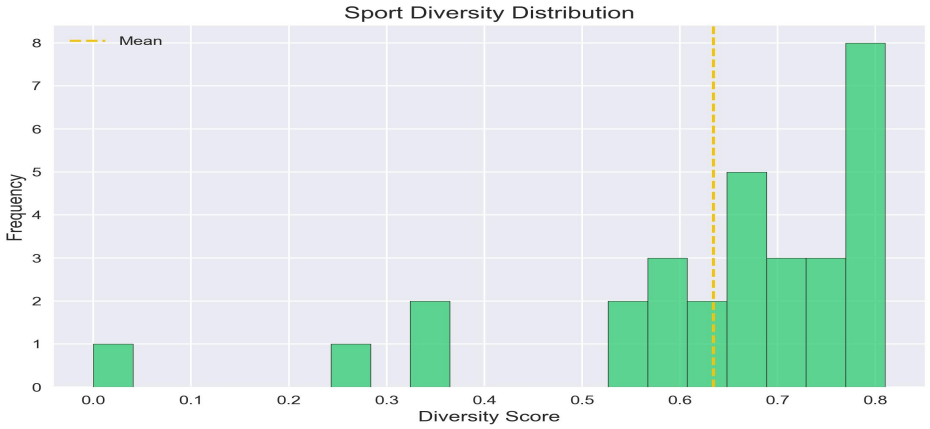

Figure 18: Sport Diversity Distribution

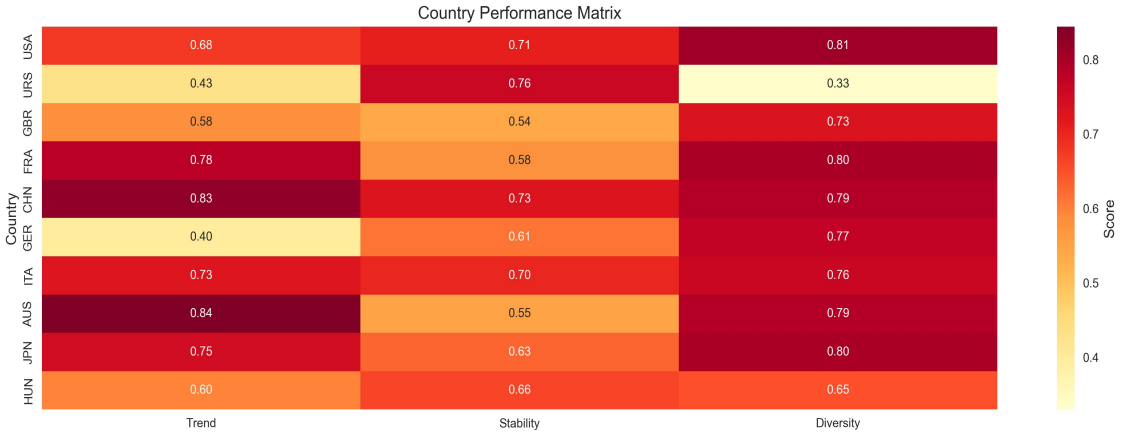

Figure 19: Country Performance Matrix

"Performance Stability Analysis" (Figure 17): This line chart displays the performance stability scores of the aforementioned countries, with lower scores indicating greater instability.

"Sport Diversity Distribution" (Figure 18): A bar chart showing the frequency distribution of diversity scores for various sports, with most countries clustered between 0.6-0.8 points.

"The Country Performance Matrix" (Figure 19) displays the scores of countries such as the United States (USA), the United Kingdom (GBR), France (FRA), China (CHN), Germany (GER), Italy (ITA), Australia (AUS), Japan (JPN), Hungary (HUN) in three dimensions: Trend, Stability, and Diversity. The darker the color, the higher the score. Based on the analysis results of development trends, stability, and diversity, the National Olympic Committees of each country can deduce the potential for sports development in their own countries, analyze their strengths and weaknesses, such as whether development is stable, and whether there is diversity in sports.

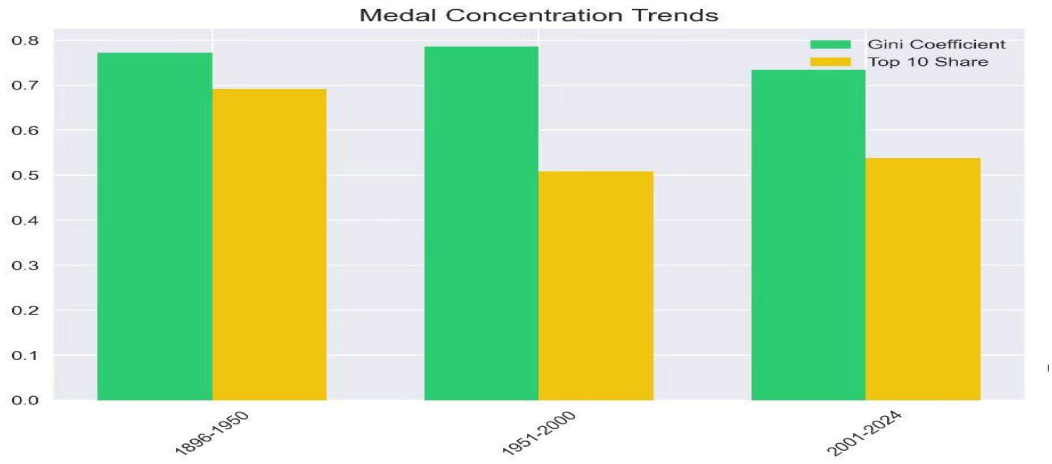

Figure 20: Medal Concentration Trends

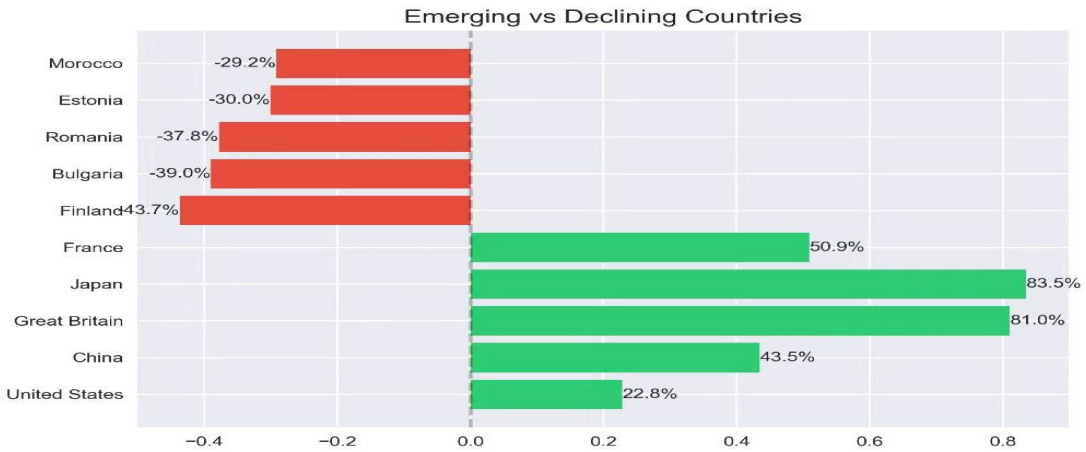

Figure 21: Emerging vs Declining Countries

Figure 20, "Medal Concentration Trends", shows medal concentration changes from

1981-2024. A higher Gini coefficient means more concentrated medal distribution. Medal concentration rose in 1991-2000, with a significant drop in the top ten countries' share, helping the Olympic Committee analyze national sports competitiveness for decision-making.

Figure 21, "Emerging vs Declining Countries", lists countries' medal percentage changes. Morocco, Estonia, Romania, Bulgaria, and Finland have negative changes, showing a downward trend. France, Japan, the UK, China, and the US have positive ones, with Japan's 83.5% increase being the largest. Analyzing these countries helps understand international Olympic competition, and national Olympic committees can use it to plan improvement strategies.

7.2.2 Unique insights on Olympic medal counts

1. Medal Trends and Pattern Analysis

Analyze medal trends and patterns to help the Olympic Committee identify national medal variation patterns, and adjust training strategies and resource allocation accordingly.

2. Analysis of medal concentration

Calculate the concentration of trends, understand the distribution of medals, help the Olympic Committee assess their competitiveness, and think about breakthroughs or opportunities.

3.Emerging Country Identification

Use composite indicators to identify emerging and declining countries, help the Olympic Committee to learn from their successful experiences, and also avoid similar problems in their own country by analyzing the situations of declining countries.

4.Regional Analysis

The enhanced regional analysis includes historical evolution and leading projects, helping the Olympic Committee to develop strategies that align with regional characteristics.

5.Calculation of National Development Trend Scores

Calculate the national development trend score to assist the Olympic Committee in assessing its own development situation and adjusting strategies.

6.Calculation of Stability and Diversity Scores

Calculate stability and diversity scores to assist the Olympic Committee in assessing its stability and consider expanding the scope of projects.

These insights analyze Olympic medal data from multiple dimensions, providing the Olympic Committee with comprehensive information to help them develop scientific and reasonable strategies, enhancing sports competitiveness.

8 Model Analysis and Evaluation

8.1 Sensitivity Analysis

To assess the robustness of the model, we manipulate the training data by randomly

sampling 70%-90% of the original dataset. The figures 22 and 23 presented depict the estimation of the uncertainty range under different training data sampling scenarios. Evidently, when data sampling is employed, the prediction results exhibit a closer alignment with the actual situation. The box plot at the base of the graph represents the uncertainty range of the predictions. From the graph, it can be observed that predicting the total number of medals may entail greater volatility and uncertainty. Nevertheless, overall, the uncertainty of our model remains low, ensuring relatively accurate predictions.

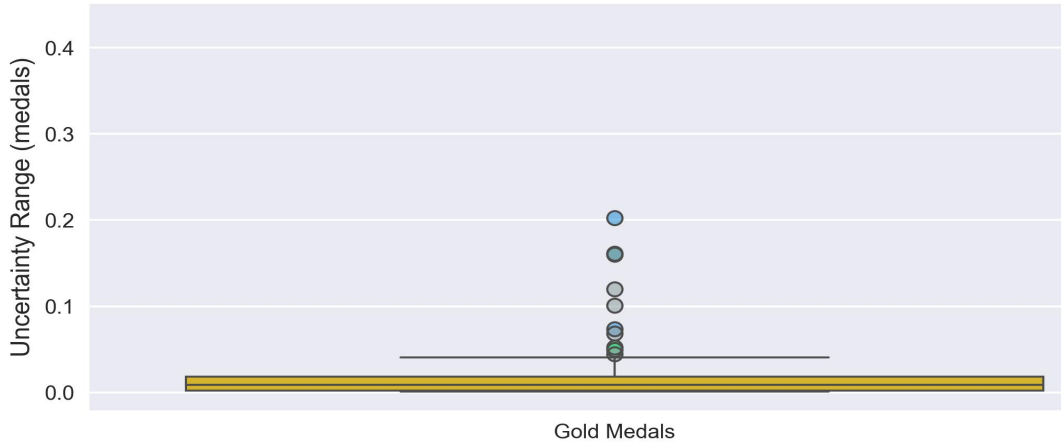

Figure 22: Uncertainty Estimation of the Number of Gold Medals

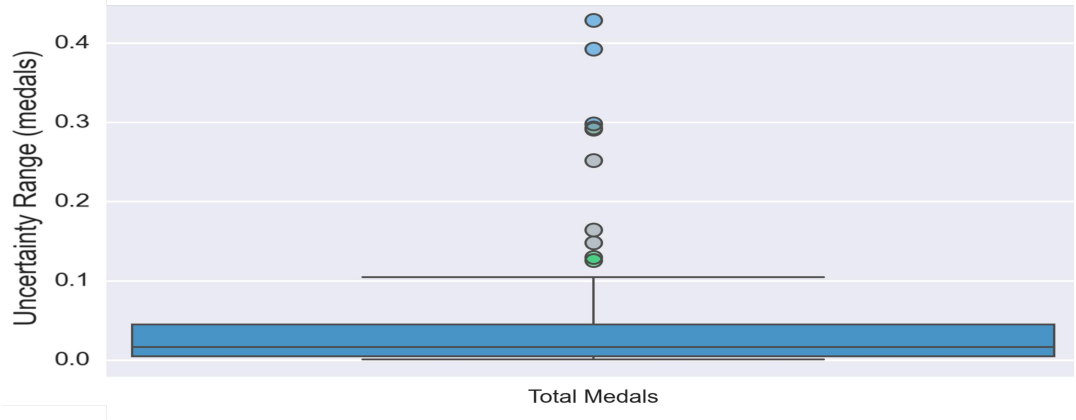

Figure 23: Uncertainty Estimation of the Number of Total Medals

8.2 Robustness Analysis

For the basic medal-prediction model, we utilize the coefficient of determination R^2 score to quantify the model's goodness-of-fit to the medal counts. The formula is as follows:

$$R^2score = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (1)$$

A value of R^2 score close to 1 indicates that the model can effectively fit the data, while an R^2 score value close to 0 implies a poor fit, with the predicted medal counts diverging significantly from the actual values. Figures 24 and 25 illustrate the R^2 scores of gold medals and total medals for the four models within the ensemble model across different validation folds.

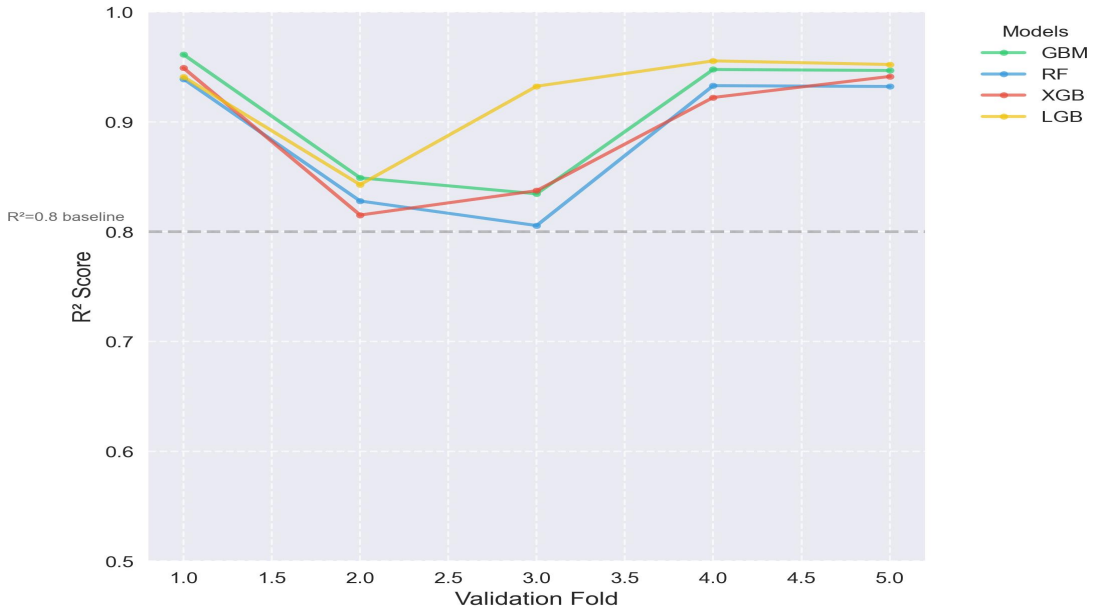

Figure 24: Gold Medal Score R^2 Trends

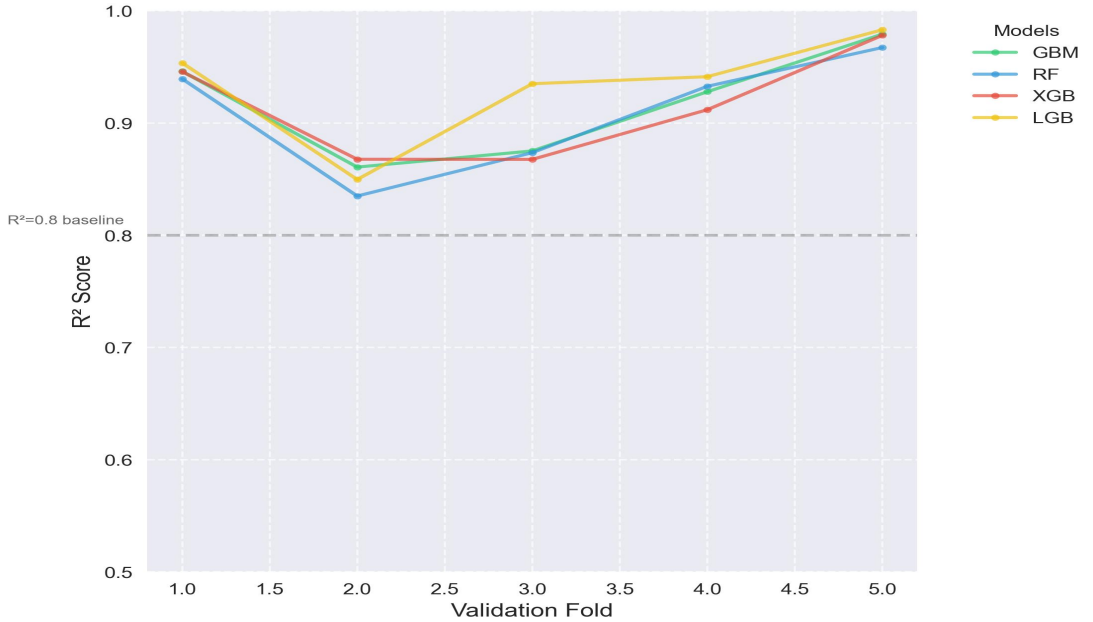

Figure 25: Total Medal Score R^2 Trends

Despite different models showing varied prediction performances and fluctuating fitting capabilities, the overall R^2 score remains above 0.8, indicating high stability. The model also has a stronger ability to fit the total medal count, validating the accuracy of the results and strengthening the predictive framework.

8.3 Strengths

- The medal prediction model using multi-source data and integrated algorithms is scientifically valid. It withstands tests such as cross-validation, and its predictions have reliable statistics.

- The multi-model framework, with “great coach” and Olympic strategy models, proves effective in different cases, ensuring system robustness.
- Large-sample data enables scientific assessment of Olympic strategies. This helps decision-makers make rational choices on event settings, coach selection, etc., to enhance medal competitiveness.

8.4 Possible Improvements

- Existing data can't fully cover medal-affecting factors like emerging events' potential and new training methods' impact. This limits prediction accuracy in special cases, and more comprehensive, real-time data is needed for in-depth analysis.
- The model's simplified assumptions and approximate algorithms may cause deviations when estimating countries' event-specific improvement potential, especially during extreme situations like major sports reforms or a surge of outstanding athletes.

8.5 Conclusion

This paper develops a basic integrated model to predict the 2028 Olympics medal count, highlighting the host-country advantage and sport-country links. Then, it verifies the "great coach" effect to assess coaches' influence and athletes' improvement potential, guiding countries' coaching investments. Finally, a detailed Olympic strategic insight model is built to analyze medal trends and offer strategic guidance, helping national Olympic committees enhance their medal-winning competitiveness.

References

[1] Luo Yubo, Cheng Yanfang, Li Mengyao, et al. Forecast of China's medal count and overall strength at the Beijing Winter Olympics-Based on the host effect and grey prediction model [J]. Contemporary Sports Technology, 2022, 12(21): 183-186. DOI: 10.16655/j.cnki.2095-2813.2112-1579-2956.

[2] Zhu Mengnan, Peng Tao, Chen Ke. Mathematical analysis of factors influencing the Olympic medal standings [J]. Contemporary Sports Technology, 2017, 7(27): 239-243. DOI: 10.16655/j.cnki.2095-2813.2017.27.239.

[3] Xiang Jun, Li Hongbing. Analysis of the regional distribution characteristics of medals at the 31st Rio Olympic Games [J]. Hubei Sports Science and Technology, 2017, 36(05): 433-436.

[4] Guo Aimin, Zhao Mingfa. Forecasting the order of the gold medal standings at the 2016 Summer Olympics based on grey theory [J]. China Science and Technology Information, 2013, (09): 173-174.

[5] Yang Jia. Analysis of factors influencing the number of gold medals won by countries at the Olympic Games [J]. Journal of Jiangxi University of Science and Technology, 2012, 33(04): 116-118. DOI: 10.13265/j.cnki.jxlgdxxb.2012.04.028.

Memorandum

To: Decision-making Teams of National Olympic Committees
From: Team #2516010
Subject: Strategic Suggestions for Sports Development Based on Olympic Medal Prediction Models
Date: January 28, 2025

Dear Decision-making Teams of National Olympic Committees,

We have developed a system of Olympic medal prediction models to provide scientific basis for enhancing the medal - winning competitiveness of countries in the Olympics. The Basic Medal Prediction Integrated Model forecasts the medal counts for the 2028 Olympics, clarifies the changing trends of medal counts and project advantages of various countries, and predicts countries that may achieve a zero-medal breakthrough. The "Great Coach" Effect Validation Model shows that the impact of coaches varies by sport and analyzes the investment potential of certain countries in different sports. The Olympic Strategy Insight Model reveals the trends of medal distribution and can evaluate the sports development status of various countries.

Based on these research findings, we recommend that national Olympic committees: allocate resources precisely, increasing investment in advantageous sports and paying attention to emerging ones; optimize the coaching team, emphasizing the role of coaches and conducting regular evaluations; learn from the experiences of emerging countries to promote domestic sports development; and improve stability and diversity by analyzing the reasons for instability and expanding the scope of participating sports. We hope that countries can use these findings to develop strategies and achieve better results in the Olympics. If you have any questions, please feel free to contact us.

Best regards!

Yours Sincerely,
Team #2516010