Revealing the Secrets behind the Medal Table

Weixiang Wang, Yulong Ma, Panyin Tao, Yutong Yang, Qicai Liu
Shijiazhuang Tiedao University, Shijiazhuang, Hebei, China
**These authors contributed equally to this work*

Abstract: With the 2024 Paris Olympics concluded, this study predicts the 2028 Los Angeles Games' medal standings, examines the "great coach" effect, and derives model-based insights, using historical data (athletes, events, past three Games' medals) processed via Random Forest to forecast 2028 gold and total medals, assess post-2024 shifts, and estimate first-medal odds for non-winners, including host effects, while Bayesian Variable Point Detection (BVPD) quantifies the "great coach" effect on Brazilian soccer, British swimming, and Chinese volleyball, confirming its impact, revealing weak event-medal correlation, host influence, and coaching significance, guiding national Olympic committees in resource allocation and strategy.

Keywords: Random Forest; Bayesian Variable Point Detection; Logistic Regression; Great Coach Effect

1. Introduction

1.1 Problem Background

The Olympic Games have a long history, from the ancient Olympic Games to the modern Olympic Games, and have conveyed the concept of gender equality and promoted cultural exchanges. It is not only a competition between athletes, but also a platform to demonstrate the comprehensive strength of a country. For the hosts, organizing the Olympic Games reflects their economic strength and stimulates related industries. For the sponsor countries, it is an opportunity to show the level of science and technology and promote scientific and technological cooperation. At the same time, the Olympic Games promote diplomacy and enhance understanding and trust between countries. The number of Olympic medals represents the strength of national sports to a certain extent, and accurate prediction of the number of medals is an important issue in sports data analysis. Athletes' performance and other factors affect the number of medals, and the establishment of a suitable prediction model can provide data support for the Olympic Committee's strategic decision-making.[2]

1.2 Restatement of the Problem

Given the background, topic constraints, and dataset, this study addresses the following:

Question 1: We develop a medal counting model to predict gold and total medal counts for each country at the 2028 Los Angeles Summer Olympics, including prediction intervals and performance measures, while assessing potential improvements or declines from 2024. The model forecasts the number of countries winning their first medal, with probabilities, and examines how event types and numbers influence medal outcomes, identifying key events for countries and the impact of the host's event choices.

Question 2: Using the provided data, we construct a model to estimate the "great coaching" effect on medal counts, selecting three countries to analyze sports where this effect is evident and quantifying its impact.

Question 3: We elaborate on original insights from these models, explaining their value for National Olympic Committees in refining Olympic strategies.

2. Assumptions and Justifications

Assumption 1: Historical Data Availability

Justification: This report assumes reliable, accurate, and current historical data, ensuring an objective reflection of countries' past Olympic performance.

Assumption 2: Correlation Between Medals and Events

Justification: The model assumes a relationship between medal counts and event participation, where increased events and host-selected events significantly influence medal outcomes.

Assumption 3: Future Tournament Projections

Justification: The report assumes the 2028 Los Angeles Olympics will resemble the prior three,

with event number changes affecting medal distribution.
Assumption 4: Host Effect
Justification: It assumes hosts favor their strong events, altering event and medal counts to boost their performance compared to non-hosts.
Assumption 5: Model Error
Justification: The report assumes random, unbiased model errors following a normal distribution to align with reality.
Assumption 6: "Great Coach" Effect
Justification: It assumes a linear athlete-medal link and a non-linear coach-medal relationship, where coaching significantly impacts medal fluctuations.

3. Notations

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

Table 1. Notations Used in This Paper

Symbol	Description
NOC	National delegations
Gold	the number of gold medals won by each country in a given year
Total	the total number of medals won by each country in a given year
Gold_growth_rate	the annual growth rate of the number of gold medals
Total_growth_rate	the annual growth rate of the total number of medals
Whet_Metal	A binary variable for whether or not a medal was awarded, with 1 indicating that a medal was awarded and 0 indicating that a medal was not awarded
Whet_Host	A binary variable indicating whether or not it is the host country, with 1 indicating that it is the host country and 0 indicating that it is not the host country
Metal_Count	Number of medals won by each country in a given sport (may include gold, silver, bronze)
$Change_{2016-2024}$	the change in the number of programs in a given sport from the 2016 Olympics to the 2024 Olympics
$Change_{2020-2024}$	the change in the number of programs in a given sport from the 2020 Olympics to the 2024 Olympics
Gold_Prediction	predicting the number of gold medals
Gold_Prediction_Interval	Confidence intervals for predicting the number of gold medals
Total_Prediction	predicting the total number of medals
Total-Prediction_Internal	Confidence intervals for predicting the total number of medals, the
Prob_Medal	the predicted probability of winning a medal
Prob_Gold	the predicted probability of winning a gold medal
Sport_Cluster	the results of the clustering of countries in different sports
Metal_Change	The contribution of changes in the number of events, host or non-host country, to the number of medals between years
Contributions	The contribution of the "great coach" effect

4. Task One: Prediction of Medals

4.1 Data Processing and Visualization

4.1.1 Outlier Handling

The dataset is tested for missing values and outliers. Host data (Year, Host) and medal data (Rank, NOC, Gold, Silver, Bronze, Total, Year) show no missing values (dtype: int64). Athlete data (Name, Sex, Team, NOC, Year, City, Sport, Event, Medal) are also complete (dtype: int64). Project data (Sport, Code, Sports Governing Body) have no gaps, but Discipline and Years (2016, 2020, 2024) each have two missing values (dtype: int64). Discipline gaps are filled using sport data, while Year gaps lead to row deletion. No outliers are detected in medal or program data.

4.1.2 Establishment of mapping tables

Table 2. Mapping Table

NOC	Country/Region
AFG	Afghanistan
ALB	Albania
ALG	Algeria

Due to the inconsistency in the expression of NOC in the tables of the various annexes, and in order to harmonize the expression to facilitate the construction of the model, the following mapping table 2 is made in this report.

4.1.3 Data visualization

Before constructing the model, this report visualizes the given data, so that the overall data characteristics can be grasped as shown in Figure 1, 2, 3.

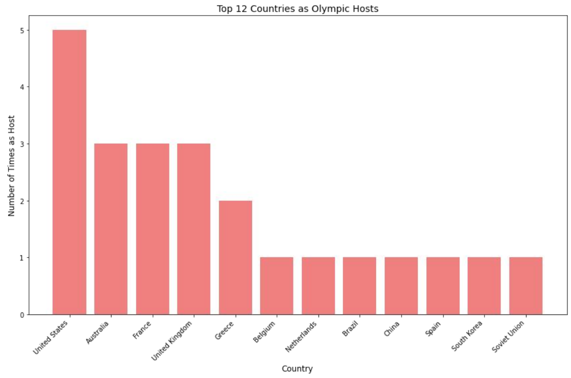

Figure 1. Top12 Countries as Olympic Hosts

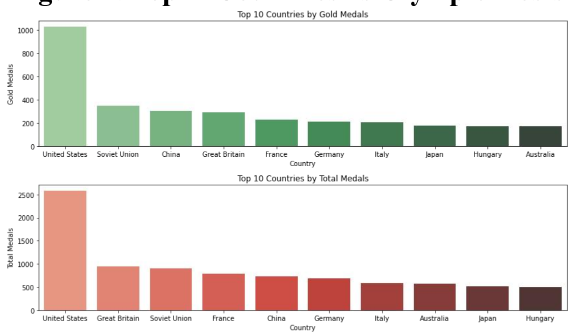

Figure 2. Top 10 Countries by Medals

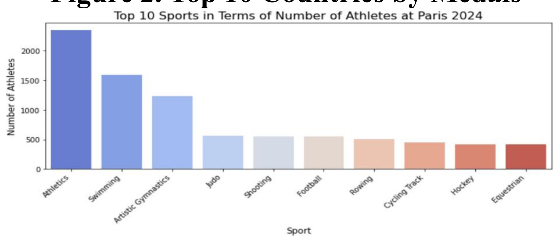

Figure 3. Top 10 Sports in Terms of Number of Athletes at Paris 2024

4.2 Modeling of Medal Predictions

4.2.1 Model building

Based on the forecasting task, two main models - regression model and classification model - are constructed in this report.

·Model-regression model for predicting the number of gold medals and the total number of medals

1) A machine learning regression model was selected for predicting gold and total medal counts in this report, as its continuous output quantifies prediction uncertainty effectively.

2) Quantile regression, estimating conditional quantiles without normality assumptions, was used to assess prediction uncertainty with selected quantiles.[3] It provided confidence intervals for predicted gold and total medal counts per country.

3) The formula:

$$\text{Gold Prediction Interval} = [\text{Gold}_{0.1}, \text{Gold}_{0.9}] \quad (1)$$

$$\text{Total Prediction Interval} = [\text{Total}_{0.1}, \text{Total}_{0.9}] \quad (2)$$

A classification model was selected to predict award outcomes due to its discrete, finite category outputs. It analyzes historical medal data, athlete numbers, and growth rates (medal and athlete) to estimate each country’s award probability. Based on these outputs, ranging from 0 to 1, the model assesses the likelihood of countries winning at the 2028 Olympic Games.

4) The formula:

$$P(\text{ Medal}) = \frac{\text{Number of countries predicted to win medals}}{\text{Total number of countries}} \quad (3)$$

4.2.2 model evaluation

Regression Modeling Assessment Criteria

MSE: Average squared difference between predicted and actual values; lower MSE shows better fit.

MAPE: Percentage error relative to actuals; lower MAPE indicates higher accuracy.

Classification Models

Accuracy: Ratio of correctly classified samples to total, reflecting performance.

AUC-ROC: Plots false vs. true positive rates; higher AUC signifies better model performance.

4.2.3 Analysis of results

Predicted Number of Gold Medals and Total Medals: Using the regression model, this report predicts gold and total medal counts for each country in the 2028 Los Angeles Olympics.

Trends in the Performance of Individual Countries: Projections and historical data suggest that countries like the United States may improve in 2024, while others, such as Austria, might see a decline post-2024.

4.3 Concrete Realization of the Model

4.3.1 Actual feature selection

This report extracts data from the last three Olympic Games, as frequent changes in hosts, programs, countries, and athletes make older data less reflective of current trends, while recent data ensures timeliness and captures performance and shifts, avoiding over-reliance on distant history; seven tables—Historical Performance, Program Participation, Program Survival Years, Athlete Participation, Athlete Participation Years, Host Country Effect, and HHI Index—are derived via Python, from which features like Athlete_count, Athlete_growth_rate, Gold, Total, Gold_growth_rate, Total_growth_rate, Whet_Host, and Whet_Metal are selected to predict 2028 medal and gold totals and win probabilities, reflecting participation scale, growth, past performance, and hosting status.

4.3.2 The classification model predicts that

Model meritocracy

This report compares Random Forest and Logistic Regression for stability and accuracy. Random Forest uses Bagging and random subspace, selecting samples and features to grow trees and vote on classifications[1]. Logistic Regression maps linear outputs to probabilities via a logistic function for prediction. Accuracy and AUC-ROC simulations are provided. As shown in Figure 4.

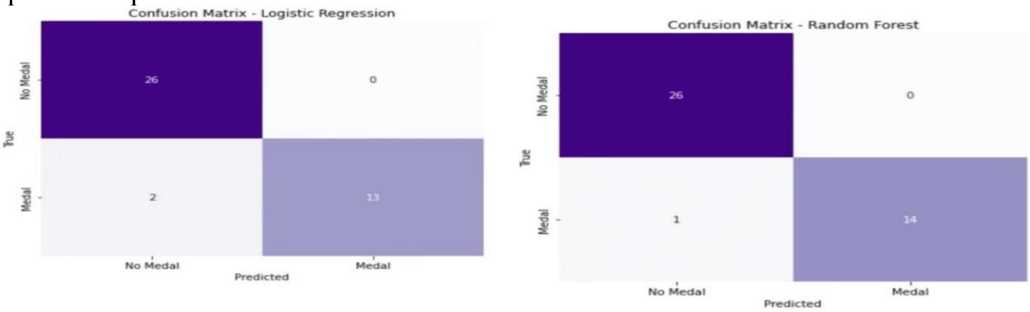

Figure 4. Two Model Training Simulations

Random Forest (Classification) achieves 0.9756 accuracy and 0.9667 AUC, outperforming Logistic Regression's 0.9512 accuracy and 0.9333 AUC, leading to its selection for subsequent analysis due to superior comprehensive performance.

4.3.3 The regression model predicts that

This report employs Random Forest Regression to predict 2028 medal counts, capturing non-linear relationships with decision trees and "out-of-bag" samples, evaluated by MSE (23.1455), RMSE (4.8110), MAE (1.4605), and R^2 (0.8361)[5]. Its strong fit yields accurate predictions, as shown in Figure 5.

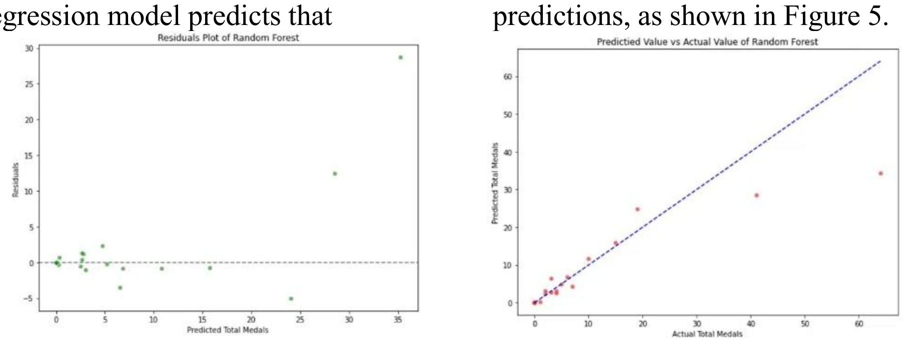

Figure 5. Scatter Plot Vs. Residual Plot

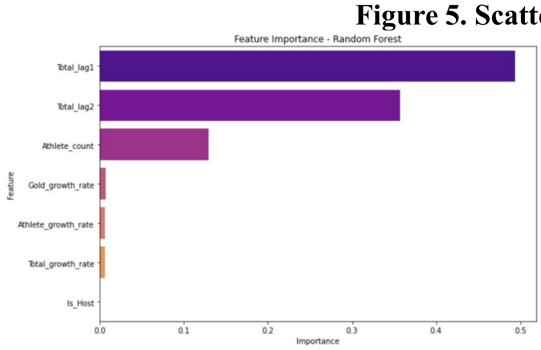

Figure 6. Importance of the Characteristics of the Random Forest Model

The figure 6 shows that athlete count, gold medal growth rate, and athlete growth rate are relatively significant features.

4.3.4Practical forecasting applications

Step 1: Using a categorical predictive model to determine whether or not a prize was awarded, with Medal as the target variable

Formula:

$$P(\text{ Medal }) = \text{classification} \setminus \_\text{model}\ (X) \text{ where } P(\text{ Medal }) \in \{0,1\} \quad (4)$$

The output of the classification prediction model is the probability of each country winning the award, which is determined by setting a threshold of 0.5. The prediction results are as shown in Table 3:

Table 3. The Prediction Results

	NOC	Predicted_Prob
1	AFG	0.0
2	ALB	0.03
3	ALG	0.26
4	AND	0.0
5	ANG	0.0
6	ANT	0.0

Step 2: Predict the number of gold medals and the total number of medals won by the winning country using a regression prediction model with the objective variables Gold and Total

Gold Prediction $= f$ (Features) where f is the regression model (5)

Total Prediction $= g($ Features $)$ where g is the regression model (6)

The projections are as shown in Figure 7:

	NOC	Gold_pred_2028
1	ARG	1.3
2	AUS	9.44
3	AZE	1.79
4	BEL	2.07
5	BRA	6.53
6	CAN	5.17

	NOC	Total_pred_2028
1	ARG	2.07
2	ARM	3.2
3	AUS	30.73
4	AUT	1.53
5	AZE	8.07
6	BEL	4.35

Figure 7. Total Medals/Gold

$$\text{Gold Prediction Interval} = [\text{Gold}_{0.1}, \text{Gold}_{0.9}] \quad (7)$$

$$\text{Total Prediction Interval} = [\text{Total}_{0.1}, \text{Total}_{0.9}] \quad (8)$$

Step 3: Determination of prediction intervals

We used quantile regression to divide the prediction intervals, thus making it easier to quantify the uncertainty in the predictions. For each country, we predicted the number of gold medals and the total number of medals for the 10%, 50% and 90% quantiles.

Here,$\text{Total}_{0.1}$ and $\text{Total}_{0.9}$ are predicted by 10% and 90% quantiles of the total medal count ,respectively. $\text{Gold}_{0.1}$ and $\text{Gold}_{0.9}$ are predicted by 10% and 90% quantiles of the gold medal count, respectively.

The forecast range is as shown in Figure 8:

	NOC	Gold_Prediction_Inte...	Total_Prediction_Inte...
1	ARG	1 - 2	2 - 3
2	ARM	0 - 0	3 - 4
3	AUS	9 - 10	30 - 31
4	AUT	0 - 0	1 - 2
5	AZE	1 - 2	8 - 9
6	BEL	2 - 3	4 - 5

Figure 8. Forecast Interval Plot

4.3.5 Analysis of results

Based on the regression model, we assess countries' potential advance or retreat in 2024 by calculating growth in gold and total medal counts.

Growth in the Number of Gold Medals:

$$\Delta \text{ Gold} = \text{Gold}_{2028} - \text{Gold}_{2024} \quad (9)$$

IfΔ Gold > 0, then the number of gold medals won by the country will increase after 2024 and may perform better.

IfΔ Gold < 0, the country will win fewer gold medals after 2024 and may perform worse.

Total medals growth:

$$\Delta \text{ Total} = \text{Total}_{2028} - \text{Total}_{2024} \quad (10)$$

IfΔ Total > 0, then the country's total number of medals won after 2024 will increase and may perform better.

IfΔ Total < 0, the country will receive fewer total medals after 2024 and may perform worse.

The number of countries projected to improve their performance is 30, as shown in Figure 9:

	NOC	Gold_Change	Total_Change	Gold_pred_2028	Total_pred_2028
1	AUS	9.44	27.53	9.44	30.73
2	AZE	1.79	6.54	1.79	8.07
3	BEL	0.2799999999999998	-3.7200000000000006	2.07	4.35
4	BRA	4.460000000000001	14.799999999999999	6.53	19.15
5	CAN	5.17	16.709999999999997	5.17	19.24
6	CHN	16.090000000000003	48.84	21.26	68.08

Figure 9. Countries with Improved Performance

	NOC	Gold_Change	Total_Change	Gold_pred_2028	Total_pred_2028
1	ARM	-1.3	1.1300000000000003	0.0	3.2
2	AUT	-9.44	-29.2	0.0	1.53
3	BRN	-6.53	-17.91	0.0	1.24
4	CIV	-21.26	-66.55	0.0	1.53
5	CUB	-0.1299999999999999	1.9100000000000001	1.96	8.68
6	DEN	-0.75	3.5600000000000005	2.03	9.08

Figure 10. Countries with Declining Performance

Twenty-nine countries are projected to experience a decline in performance, as shown in Figure 10:

From the above projections, it is easy to see that China (CHN), the United States (USA) and Australia (AUS) are likely to make significant progress in the post-2024 Olympics, while Austria (AUT), the Philippines (PHI) and Uzbekistan (UZB) are likely to take a step back in the future Olympics and perform poorly.

4.4 Projections of Countries Receiving Awards for the First Time

4.4.1 Volume projections

The classification model predicts likely 2028 winners, identifying first-time medalists among countries with no prior wins, using Predicted_Prob to estimate their chances.

$$\text{First Time Medal} = \begin{cases} 1 & \text{if Predicted\_Medal} = 1 \text{ and Gold} = 0 \text{ and Total} = 0 \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

The number and likelihood of first-time winners can be estimated using the following formula.

$$P(\text{ First Time Medal }) = \frac{\text{Number of First Time Medal Countries}}{\text{Total Number of Predicted Medal Countries}} \quad (12)$$

From the results, it is easy to see that the number of countries that could win their first medal at the next Olympics is zero.

4.4.2Likelihood estimates

The predicted probability distribution of medals for all countries is plotted:

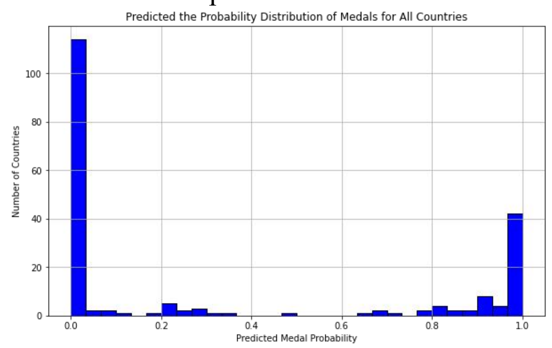

Figure11. Probability of Winning the Award Distribution

The figure 11 shows most countries' winning probabilities range from 0 to 0.1, indicating low chances, while a few approach 1, suggesting high likelihoods.

4.5 Race analysis

In predicting Olympic medals, in addition to athlete data and host effects, it is also necessary to take into account changes in the type and number of Olympic events in the current edition.

4.5.1Project and medal count analysis

This report examines the correlation between event entry changes and medal counts across years, calculating Pearson's correlation coefficient for these variations.

$$\text{Correlation}(X,Y) = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2\sum_{i=1}^{n}(Y_i-\bar{Y})^2}} \quad (13)$$

found X_i is the change in quantity of item i. Y_i is the number of medals corresponding to the program. $\bar{X}$ cap (a poem) $\bar{Y}$ are the mean values of the number of programs and the number of medals, respectively.

Calculating correlations yields a correlation matrix that measures which events contribute most to the medal count as shown in Figure 12.

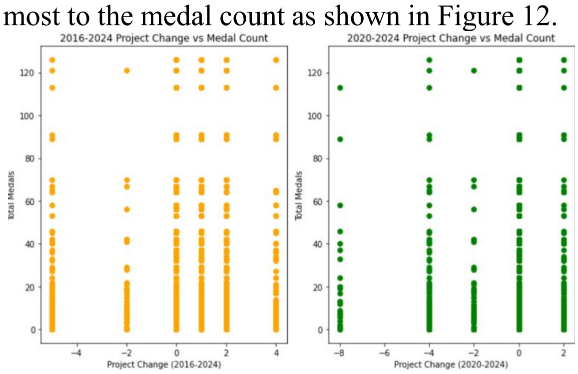

Figure 12. Scatter Plot

4.5.2 Analysis of results

The correlation for 2016-2024 is 0.03,and for 2020-2024 is 0.02.The correlation between the change in the number of events and the number of medals was shown to be very low and the change in the number of events did not have a significant effect on the number of medals.

4.6 The host effect

4.6.1 Based on an analysis of program and medal counts, the

Based on 4.5 event changes and medal counts, we compare host vs. non-host medal differences across events, using mean averages to evaluate if hosts gain more medals by adding events.

$$\Delta M_{\text{Host}} = \frac{1}{N_{\text{Host}}}\sum_{i\in\text{Host}} \mathrm{M}_i - \frac{1}{N_{\text{Non-Host}}}\sum_{i\in\text{Non-Host}} \mathrm{M}_i \quad (14)$$

M_i is the number of medals for each country. N_{Host} cap (a poem) $N_{Non-Host}$ host and non-host country sizes, respectively. The regression model yields predicted total medal counts when hosting: 0 (non-host) = 13.67 medals, 1 (host) = 49.67 medals (Name: Total, dtype: float64).

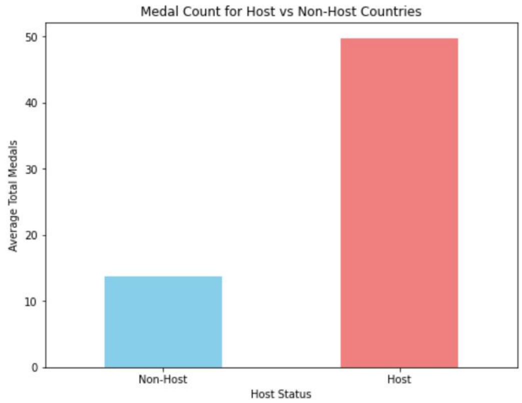

Figure 13. Impact of Adding a Dominant Program on Hosts vs. Non-hosts

This study evaluates the host effect by analyzing if hosts boost dominant events, assessing their impact via annual event and medal count changes at the Olympics.

4.6.2 Changes in the host country's choice of events and number of medals in different years

Host countries' higher average medal counts compared to non-hosts indicate a tendency to select advantageous events to boost their medal totals.

Results, as shown in Figure 14:

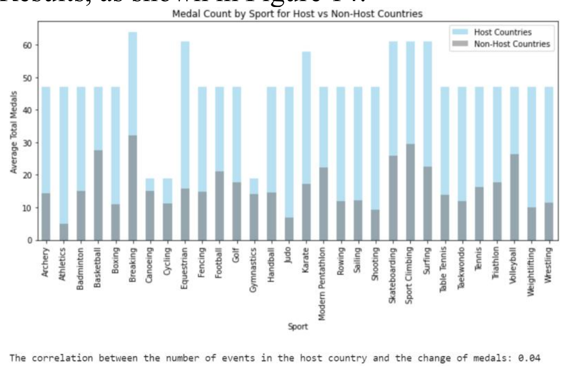

Figure 14. Host Country vs. Non-host Country Programs vs. Number of Medals

$$\Delta P_{\text{Host}} = \text{Change in Projects}_{\text{Host}} \text{ and } \Delta M_{\text{Host}} = \text{Change in Medals}_{\text{Host}} \tag{15}$$

4.6.3 Analysis of results

Correlation Analysis: The correlation between Olympic events and medal counts is negligible (cycling: 6.938894e-18; basketball, canoeing: 0.000000e+00). Host event changes and medals show a weak correlation of 0.04, indicating a minimal relationship.

Medal Count Comparison and Correlation Analysis: Hosts outperform non-hosts in medals, especially in boxing, surfing, and rock climbing, suggesting event selection boosts host tallies. Event-medal correlations are near zero (cycling: 6.938894e-18; basketball, canoeing: 0.000000e+00), with host event changes and medals at 0.04, implying a slight link.

5. Task Two: The Coaching Effect

5.1 Data Description

Set as the weighted number of MEDALS won by a country in a sport in year t:

·Gold Medal: Weighting of 5

·Silver medal: weighting of 4

·Bronze medal: weighting of 3

Then the weighted medals in year t, the$W_t$can be expressed as follows:

$$W_t = 5 \cdot \text{Gold}_t + 4 \cdot \text{Silver}_t + 3 \cdot \text{Bronze}_t \tag{16}$$

5.2 Bayesian Change-Point Detection

5.2.1 Modeling

This report assumes the existence of a time series of data on the number of medals $W_1, W_2, \ldots, W_T$ (where T represents the year of the Olympic Games). It is assumed that the data have different distributions before and after the change point, and the data in each interval obey normal distribution, and the mean and variance of the normal distribution in each interval are different from each other.

Let μ_k and σ_k^2 be the mean and variance of the KTH interval respectively, then the weighted number of MEDALS in year t can be expressed as:

$$W_t \sim \mathrm{N}(\mu_k, \sigma_k^2), t \in [t_{k-1}, t_k] \tag{17}$$

Among them, μ_k and σ_k^2 change with the occurrence of change points.

5.2.2 Bayesian inference

According to Bayes' formula, the posterior distribution can be expressed as follows:

$$p(t_k|W) = \frac{p(W|t_k)p(t_k)}{p(W)} \tag{18}$$

Among them. $p(W|t_k)$ for a given change point$t_k$the likelihood function of the data at the time; the $p(t_k)$ as a change point t_k of the prior distribution, it is usually assumed that the change points occur with uniform probability, i.e., that the $p(t_k) = \frac{1}{T}$; $p(W)$ The marginal likelihood of the data is generally estimated by numerical methods.[4]By calculating the posterior probability, we can identify the change points in the data.

5.2.3 Practical inferences

The results of the solution are as shown in

Figure 15：

Detected change points (years): [2016 2020 2024]

Corresponding change point scores: [14.50116017 11.28665804 8.7927194]

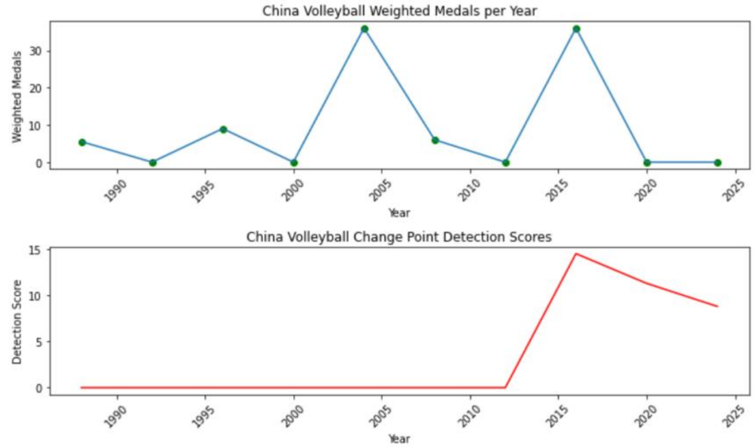

Figure 15. China Volleyball Team

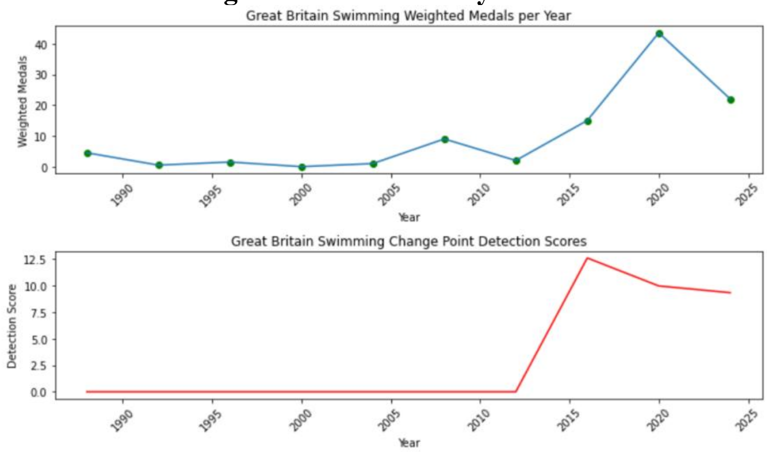

Figure 16. British Swimming

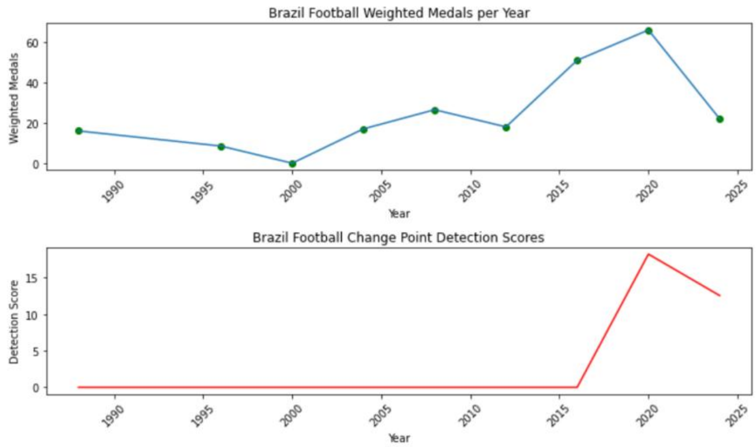

Figure 17.Brazilian Soccer Team

Figure 16 Detected change points (years): [2016 2020 2024]

Corresponding change point scores: [12.62203441 9.98566096 9.3494459]

Figure 17 Detected change points (years): [2020 2024]

Corresponding change point scores: [18.21142249 12.54203792]

5.2.4 Analysis of results

In a comparison of the number of medals before and after the change points for Chinese volleyball, British swimming, and Brazilian soccer, it can be seen that there is a significant increase in the number of medals due to the effects of coaching turnover or strategic adjustments.

5.3 The Contribution of the Coaching Effect

After identifying the change points according to 5.2, this report quantifies the contribution of the coaching effect by weighting the number of medals.

5.3.1 The average of the number of medals

1) before the point of change: W_t

$$W_{\text{before}} = \frac{1}{T_{\text{before}}}\sum_{t\in\text{before}} W_t \quad (19)$$

2) after the change point: W_t

$$W_{\text{after}} = \frac{1}{T_{\text{after}}}\sum_{t\in\text{after}} W_t \quad (20)$$

3) contribution calculation

Coach's contribution is calculated by calculating the ratio of the average number of medals after the change point to the average number of medals before the change point.

$$\text{Contribution} = \frac{\bar{W}_{\text{after}}-\bar{W}_{\text{before}}}{\bar{W}_{\text{before}}} \quad (21)$$

If the contribution is greater than 0, it means that the new coach had a positive impact on the number of medals and thus improved the country's performance in the event: if the contribution is less than 0, it means that the new coach had no effect on the number of medals.

5.3.2 The contribution of the three projects

British Swimming: A contribution of 10.29 reflects a substantial medal increase, driven significantly by highly effective coaching.

Brazilian Soccer: A contribution of 4.05 shows a strong coaching impact on medal counts, though less than British swimming.

China Volleyball: A contribution of 0.77 indicates a notable positive coaching effect on the team's medal performance.

5.3.3 Analysis of results

The "great coaches" effect: Because of the differences in strategies and management styles of individual coaches, coach turnover leads to significant changes in the number of medals won.

6. Task Three: Original Insights and Implications

The host effect increases medal counts via strategic event selection, aiding Olympic Committee planning; the “great coach” effect underscores coaches’ vital role in event optimization and performance, driving quick gains with skilled hires; the medal prediction model identifies first-time winners, promoting diversity and informing IOC support.

7. Sensitivity Analysis

In this thesis, the sensitivity analysis of the model is realized by changing the parameter values of certain features as well as certain combinations of features and observing the degree of change in the predictive model. The weights of the features are as shown in Fingure 18:

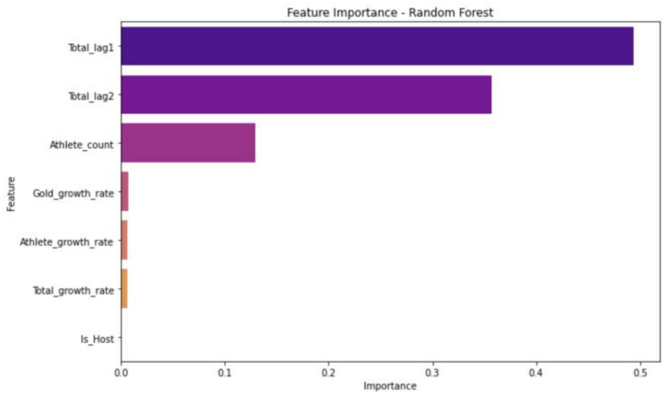

Figure 18. Weights of Each Feature

Critical profiling

Athlete_count: high importance and the model is sensitive to its changes. By modeling the change in athlete count, its impact on medal prediction is assessed.

Secondary characterization

Gold_growth_rate, Athlete_growth_rate, Total_growth_rate: Low model impact; reduce reliance in feature selection. Whet_Host: Minimally significant, little effect on forecasts; consider excluding in practice.

Aggregate analysis

Feature Combinations: Assess the effect of combining features like Athlete_count and Gold_growth_rate on predictions.

Importance vs. Significance: Integrate feature importance with significance evaluation; e.g., Athlete_count is key but may reflect past performance, aiding feature optimization and model decisions.

8. Model Evaluation and Further Discussion

8.1 Strengths

This paper offers a straightforward, logical model using Random Forest for accuracy and high-dimensional data processing, and Bayesian Variable Point Detection for precision and flexibility, ensuring reliable predictions with robust stability and strong extensibility across tests.

8.2Weaknesses

The Random Forest model requires lengthy training and high memory, while Bayesian Variable Point Detection involves complex parameter selection and elevated computational demands, potentially over-relying on specific traits, sidelining other factors.

8.3 Promotion

In preprocessing the Random Forest model, feature selection reduces low-impact features and data dimensions, accelerating training, while for Bayesian Variable Point Detection, approximate inference like Markov Chain Monte Carlo (MCMC) lowers computational complexity and boosts efficiency.

9. Conclusion

This paper's model, built on provided data, offers rational, accurate predictions, training a Random Forest classification model on processed data to preliminarily identify winning countries, followed by a regression model forecasting 2028 medal and gold totals, predicting progress for China (CHN) and the USA, regression for Austria (AUT) and the Philippines (PHI), with uncertainty measured for reliability, while noting no first-medal wins due to limited historical data for non-winners, a non-linear event-medal link, and a positive host-event effect; Bayesian change-point detection quantifies the "great coach" effect, showing its significant medal impact—especially during coaching shifts—via weighted averages, with deeper analysis yielding original insights.

References

[1] Xiao Chonghui, Ming Yang. A left-turn signal loss time prediction model based on random forest regression[J]. Municipal Technology, 2024, 42 (11): 99-104.

[2] Shi Huimin, Zhang Dongying, Zhang Yonghui. Can Olympic medals be predicted? --Based on Interpretable Machine Learning Perspective [J]. Journal of Shanghai University of Physical Education and Sports, 2024, 48 (04): 26-36.

[3] Wei Xueyong, Jin Liangqiong, Shen Ting,et al. An ALD-based Bayesian quantile regression estimation [J]. Journal of Hunan College of Arts and Sciences(Natural Science Edition), 2022, 34 (02): 1-7.

[4] Fang Yin, Chen Qiuju, Pan Jifei,et al. Bayesian-based multifunctional radar pulse train change point detection [J]. Journal of Command and Control, 2019, 5 (04): 308-315.

[5] Yang Jinghan. Design of Olympic medal ranking prediction based on univariate and multiple regression models [J]. Information System Engineering, 2018, (02): 149-152.

[6] Zhu Zihan. Analyzing and predicting Olympic medal list through data modeling method[J]. Electronic Technology and Software Engineering, 2018, (02): 167.