Problem Chosen	2025	Team Control Number
C	MCM/ICM Summary Sheet	2515168

Medal Predictions and Coaching Impact: Building a Model for Los Angeles Olympic 2028

Summary

The quadrennial Olympic Games are a global platform where nations compete for medals and honor. This paper analyzes Olympic medal acquisition and explores the impact of exceptional coaches, aiming to predict medal outcomes.

Firstly, we develop an **XGBoost** prediction model to forecast the medal outcomes for the 2028 Los Angeles Olympic Games in the United States. This model employs the **TPE(Bayesian Optimization)** algorithm for hyperparameter tuning of XGBoost, selecting optimal parameters that resulted in an **RMSE of 3.268 for gold medal predictions and 4.853 for total medals**, indicating **high accuracy**. The model predicts the United States, with **43 gold and 123 total** medals. We established a scoring system and predicted that **Great Britain** is most likely to **improve**, while **South Korea** is most likely to **decline**.

Secondly, using the **bootstrap method**, we estimate with a **0.95 confidence interval**, the predicted gold medal range for U.S. is between **[36.99, 59.29]**. Meanwhile, we track the probability of countries that have yet to earn a medal, forecasting that **10 countries will earn their first medal**, with **Libya having a 0.69 probability** of winning its first medal.

Thirdly, To explore the relationship between the number of sports and the distribution of medals, we use the **Herfindahl Index**, finding that the **number and variety of events increase** leads to a more **dispersed medal distribution**. We calculate the **Spearman coefficient** for the host country effect, finding a **strong correlation between the host nation and medal success in specific events**.

Fourthly, to examine the potential impact of a "great coach" effect, we conduct **LMM (Linear Mixed-Effects Model)**. By validating the underlying assumptions, We have examined that a "great coach" effect has a **significant positive impact** on **medal attainment**. Meanwhile, its **fixed effect coefficient** is regressed to be **1.088**, indicating a "great coach" treatment can provide **one additional bronze medal**.

To choose the proper countries and sports, we get a combined weight of the **Analytic Hierarchy Process and Entropy Weight Method**. Then, we use **TOPSIS** method to rank alternatives, finding that the top 10 closeness is between **0.68** and **0.36**. We mainly consider **Japan, Great Britain** and **France**. Then, we determined the **top 3 Sports** for each country according to the **closeness** and made predictions with previously trained LMM. For the UK, this effect is expected to boost the Cycling Track score by **21.77% (0.671 units)**.

Finally, we give advise to the country Olympic committees with three other insights.

Keywords: XGBoost; TPE; Bootstrap; Herfindahl Index; TOPSIS; AHP; EWM; LMM

Contents

1 Introduction

1.1 Background

The Olympic Games are a global celebration of athletic excellence, with the medal table serving as a key indicator of national sporting success. The 2024 Paris Olympics saw the United States and China tied for the most gold medals, while the host nation, France, secured 5th place in gold medals but 4th in total medals. Notably, smaller nations like Dominica and Saint Lucia made history by winning their first Olympic medals, including gold, highlighting the Games' role in inspiring breakthroughs and fostering national pride. These achievements underscore the diverse and evolving landscape of international sports.

1.2 Restatement of the Problem

The dataset provides comprehensive information on the Summer Olympics from 1896 to 2024, including details on the competitions (such as sport, year, and medal outcomes), medal counts by country, host countries, and a breakdown of events by sport. The objective is to develop mathematical models to address the following questions.

Problem 1: Create a model to predict medal counts (focusing on gold and total medals) for each country. Calculate the estimation of the uncertainty/precision of the model predictions and measures of how well model performs.

Sub-tasks

- Investigate the relationship between the number and types of Olympic events and the medal distribution. Determine which sports are most significant for various countries and explain why. Analyze how the host country's choice of events influences the results.
- Use the model to provide projections for the medal table at the 2028 Los Angeles Summer Olympics, including prediction intervals for all results. Identify which countries are likely to improve their performance and which may perform worse compared to 2024.
- Include countries that have never won a medal and predict how many countries will earn their first medal in the next Olympics. Provide probability estimates for this prediction.

Problem 2: Analyze a "great coach" effect on medals and evaluate coach investment recommendations.

Sub-tasks

- Examine the data for evidence of the "great coach" effect on medal counts.
- Estimate the contribution of the “great coach” effect to medal counts.
- Select three countries and identify sports where investing in a “great coach” could have the most impact. Estimate the potential improvement in medal counts.

Problem 3: Use the model to uncover additional insights about Olympic medal counts.

Sub-tasks

- Explain how these insights can inform country Olympic committees.

2 Assumptions and Overviews

2.1 Assumptions

To simplify our problem, we make the following basic assumptions that each of assumption is made with sufficient reason.

- **Assumption 1:** Countries that did not participate in any Olympic Games after 2012 (including the 2012 Olympics) for any reason (such as national affairs or individual athlete circumstances) are assumed not to participate in the 2028 Los Angeles Olympics. For instance, Russia is considered not participate in the 2028 Olympics due to national affairs.
  Reason: The participating countries for the 2028 Los Angeles Olympics have not yet been confirmed. This assumption essentially covers all potential participants and simplifies the complexity of the prediction.
- **Assumption 2:** If there are multiple teams from the same country (e.g., France-1 and France-2) in a given Olympic year, the medal counts of these teams are aggregated into the total medal count for that country, without considering the specifics of individual teams.
  Reason: Since this study focuses on predicting medal counts at the national level, aggregating medal counts from multiple teams helps simplify data complexity, reduce noise, and more accurately reflect macro-level trends in the overall performance of countries.
- **Assumption 3:** The "great coach" effect is theorized to exert a homogeneous influence across diverse Olympic disciplines and among all competitors given identical parameters. Meanwhile, factors such as national economic conditions, policies, and individual athlete characteristics will be treated as normally distributed noise.
  Reason: The available dataset solely encompasses the participation of each athlete in various competitions and their corresponding outcomes, which can be utilized to modeling and quantify the "great coach" effect.

2.2 Overviews

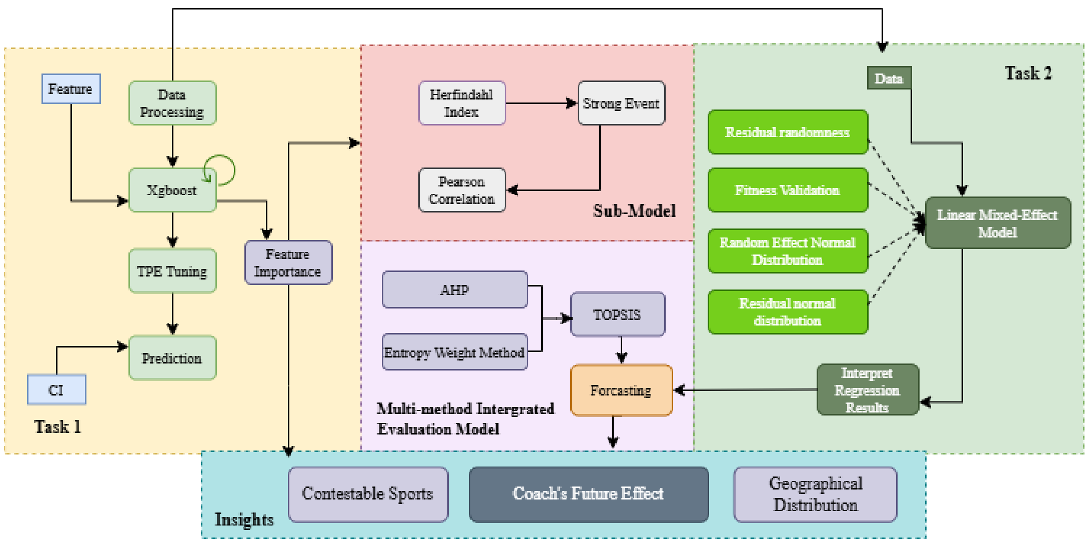

Figure 1: Flow Map

3 Model Preparation

3.1 Notations

Symbols	Description
D	The training dataset, where $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$
x_i	The feature vector of the i-th sample
A_i	The number of athletes in country i
G_i	The average number of gold medals for country i over the last 4 years
M_i	The average number of total medals for country i over the last 4 years
E_{ij}	The medal count for country i in event j
H_i	Host country effect for country i
w_k	The weight of the k-th tree in the XGBoost model
C_{nj}^{t}	Indicating if athlete m has been directed by a "great coach" in event j by year t
P_{nj}^{t-Y}	Performance score of the athlete m in event j in previous $t - year$
N_{nj}^{t}	Athlete m's accumulated participation in event j by year t
X_{ij}	The value of i-th alternative for the j-th criterion
W_i	The weight of i-th criterion
$\text{Tree}_k(x_i)$	The prediction of the k-th tree for the i-th sample
$\mathcal{L}(y_i, \hat{y}_i)$	The loss function for the i-th sample

Table 1: Notations

3.2 Data Pre-processing

The data we use include the provided files: `summerOly_athletes.csv`, `summerOly_hosts.csv`, `summerOly_medal_counts.csv`, and `summerOly_programs.csv`.

These data files provide the necessary information to solve the problems. However, we discovered that some of the data need to be pre-processed before use.

Data Cleaning and Standardization

First, we clean the raw data by removing garbled and invalid entries, filling in and replacing erroneous values, meanwhile removing abandoned sports under the instruction of `data_dictionary.csv`. This step ensured the completeness and accuracy of the data, laying a reliable foundation for subsequent analysis.

To ensure data consistency, we standardized the names of National Olympic Committees (National Olympic Committees [NOCs], 2004), such as merging entries that refer to the same country under different names. Additionally, we removed data from countries that no longer exist or participate in international competitions, ensuring the timeliness and relevance of the data.

Data to be processed	Adjustment	Sample Adjustment
Garbled text	Deletion	Italy? → Italy
Erroneous data	Replace	Missing value → Median
Abandoned Sports	Deletion	Figure Skating → Deletion
NOC	Replace	CHN → China
Not Participate Country	Deletion	Soviet Union → Deletion

Table 2: Data Cleaning and Standardization

3.3 Medal Scoring Mechanism

To comprehensively quantify athletes' performance and assess the trends of progress and regression for different countries in the Olympics, while analyzing the impact of a "great coach" effect on medal counts, we adopt a quantitative mechanism of medal substitution points, which is defined as follows.

Medal	Gold	Silver	Bronze	No Medal
Scoring	5	3	1	0

Table 3: Medal Scoring System

4 Model I: XGBoost Prediction Model

4.1 Feature Engineering

In the data processing phase, we addressed missing and erroneous data and standardized naming conventions. Following this, we performed feature engineering, where we extracted and processed key variables likely to influence Olympic performance, leveraging historical performance data and relevant features.

- A_i: The number of athletes in country i.
- G_i: The average number of gold medals for country i over the last 4 Olympics game.
- M_i: The average number of total medals for country i over the last 4 Olympics game.
- Y_{ik}: The year of Olympics k (Standardized) for country i.
- H_i: Host country effect for country i, where $H_i = 1$ if country i is the host, and $H_i = 0$ otherwise.
- E_{ij}: The medal count for country i in event j.

According to research, the typical career length of athletes is about 20 years (Sherry, 2016). Another study by SV Allen and WG Hopkins suggests that athletes are most competitive around the age of 30. Based on these findings, we have chosen $X = 4$ as the data foundation for the feature variables, considering the career development of athletes and potential impacts from illness and accidents.

4.2 Model Training and Hyperparameter Tuning

We chose the XGBoost algorithm (Chen & Guestrin, 2016) for its ability to handle complex datasets and model non-linear relationships, which is crucial for predicting Olympic medal counts. The model is defined as:

$$\hat{y}_i = f(x_i; \theta) = \sum_{k=1}^{K} w_k \cdot \text{Tree}_k(x_i)$$

where $\hat{y}_i$ is the predicted medal count for country i, x_i is the feature vector (including A_i, G_i, M_i, H_i, and E_{ij}), and $\text{Tree}_k(x_i)$ is the prediction from the k-th decision tree.

The model iteratively refines the predictions by minimizing the following objective function:

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

where $L(y_i, \hat{y}_i)$ is the loss function, and $\Omega(f_k)$ is the regularization term that controls the complexity of the trees:

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$$

In each iteration of the **XGBoost** training process, the model builds a tree by splitting the data at nodes, increasing the number of leaves, T, as the tree deepens. At each step, the leaf weights, w_j, are adjusted based on the gradients of the loss function L to minimize the prediction error. The regularization parameters γ and λ were optimized through grid search, with optimal values of $\gamma = 0.1$ and $\lambda = 2.0$ selected based on **RMSE (Root Mean Squared Error)** minimization using **5-fold cross-validation**. The heatmap shows the effect of different parameter combinations on the RMSE, highlighting the optimal values.

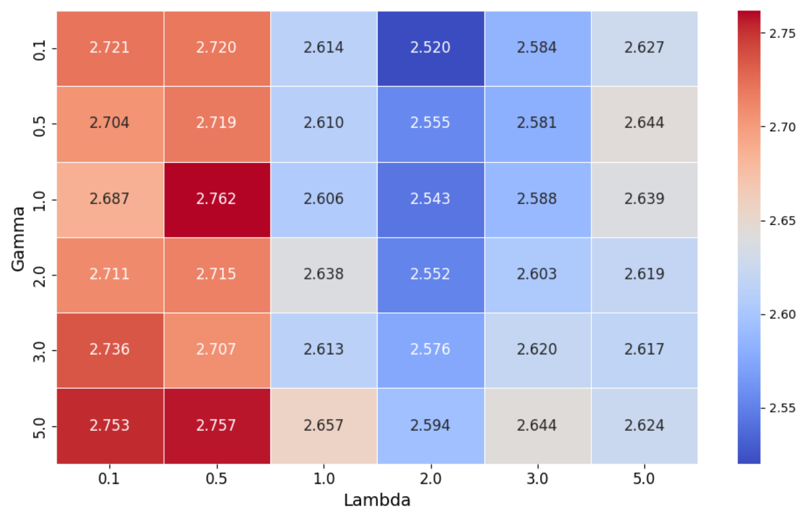

Figure 2: Effect of Regularization Parameters on Cross-Validation RMSE

To optimize the hyperparameters of the XGBoost model, we applied **TPE (Bayesian Optimization)** with the goal of minimizing the RMSE. The objective function for hyperparameter tuning is defined as:

$$L(\theta) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \Omega(\theta)$$

During the optimization process, TPE used the RMSE on the validation set to predict which hyperparameters were likely to yield better performance. This approach allowed us to efficiently narrow down the search space and identify the optimal hyperparameters. After applying TPE, we arrived at the following settings: **learning rate = 0.225**, **max depth = 3**, with RMSE values of **3.268** for gold medal prediction and **4.853** for total medal prediction on the test set.

We compared the models trained with and without TPE optimization. The scatter plots below show the predictions for the gold and total medal counts, illustrating the difference in model performance. As shown in the figures, the **TPE-optimized model** provides more accurate predictions, as reflected in the higher R^2 scores.

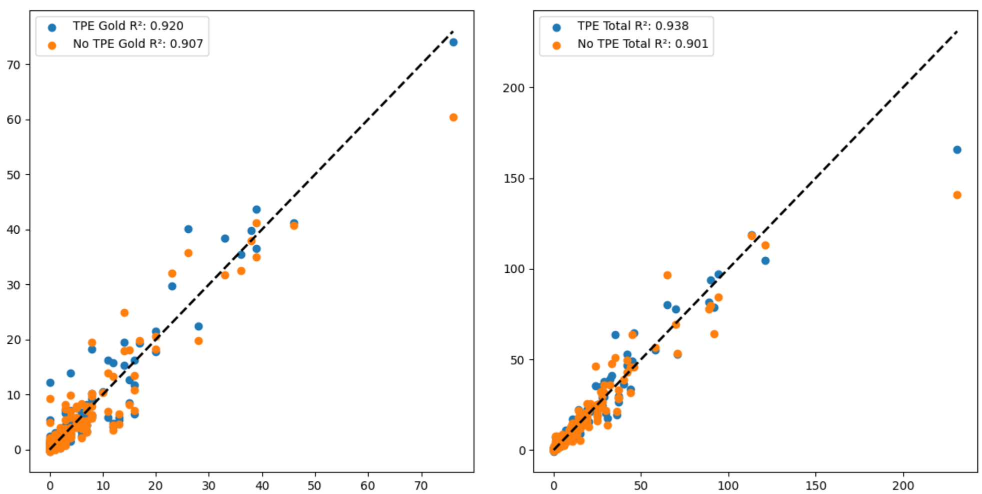

Figure 3: Comparison of predictions

4.3 Forecasting Results

4.3.1 Prediction of the Medal Table

After fine-tuning the model, we predicted the medal counts for the 2028 Los Angeles Olympics. We focused on the top ten countries based on their predicted performance. The results were visualized in a medal table, showing the expected number of gold and total medals for each country.

Rank	Country	Gold	Total
1	United States	43	123
2	China	38	90
3	Great Britain	27	69
4	France	17	62
5	Japan	21	53
6	Australia	17	50
7	Germany	14	41
8	Italy	12	39
9	Netherlands	14	36
10	Canada	8	27

Figure 4: Predicted Medal Table for the 2028 Los Angeles Olympics (Top 10 Countries)

4.3.2 Prediction of the Improvement and Decline

Utilizing our medal-score mechanism to predict scores, the three countries most likely to improve are:

- **Great Britain**: Scoring of 30.64
- **Germany**: Scoring of 12.13
- **Japan**: Scoring of 10.21

On the other hand, the three countries most likely to experience a decline are:

- **South Korea**: Scoring of -13.99
- **New Zealand**: Scoring of -8.10
- **Australia**: Scoring of -5.91

This analysis highlights the countries with the most significant expected improvements and those that may face challenges in the 2028 Olympics.

4.4 Bootstrap Method for Confidence Interval Estimation

The **Bootstrap method** is a statistical approach for assessing prediction uncertainty and deriving confidence intervals. It involves resampling the original dataset multiple times. In our analysis, we used forecasts from a weighted combination model for the upcoming 60 periods as the base data, generating 800 bootstrap samples by resampling 800 times with replacement, denoted as $Y_B = (Y_1, Y_2, \ldots, Y_{800})$.

Each sample is presumed to reflect the original forecast distribution. We calculated the standard deviation of these samples to estimate the variability of our predictions.

According to the **CLT (central limit theorem)**, we calculated the **95%**, **90%**, and **80%** confidence intervals for the predicted medal counts using the **z-statistic**, with the following formulas:

$$\hat{y}_i^{lower} = \mu - z_{\alpha/2} \times \frac{SD}{\sqrt{n}}$$

$$\hat{y}_i^{upper} = \mu + z_{\alpha/2} \times \frac{SD}{\sqrt{n}}$$

The tables below show the forecast intervals for the predicted number of gold medals and total medals for the United States at different confidence levels:

Confidence Level	Gold Prediction (Lower Limit)	Gold Prediction (Upper Limit)
95%	36.99	59.29
90%	39.34	59.17
85%	40.58	57.93

Table 4: Predicted Gold Medal Counts at Different Confidence Levels

Confidence Level	Total Prediction (Lower Limit)	Total Prediction (Upper Limit)
95%	91.40	155.85
90%	94.82	144.80
85%	97.93	141.69

Table 5: Predicted Total Medal Counts at Different Confidence Levels

From a statistical perspective, the predicted number of medals falls within these calculated confidence intervals, providing a range of values that represent the uncertainty in the model's predictions. The intervals for both gold and total medals offer valuable insights into the possible outcomes, helping us to assess model performance under varying confidence levels.

4.5 First Medal Prediction

The probability of a country winning at least one medal in the 2028 Olympics is computed by using the Bootstrap method. For each bootstrap iteration, a model is trained and predictions are made. The probability is the proportion of iterations where the predicted total medal count is greater than zero:

$$P(\text{At Least One Medal}) = \frac{1}{n_{\text{bootstrap}}} \sum_{i=1}^{n_{\text{bootstrap}}} \mathbb{I}(y_{\text{predicted}}^{(i)} > 0.3)$$

where $\mathbb{I}(\cdot)$ is the indicator function, and $y_{\text{predicted}}^{(i)}$ is the predicted total medal count in the i-th iteration. Since the predicted medal counts for those countries that have never won a medal could lead to severe skewness in the results, we determine the threshold to be 0.3 by comparing data results from different thresholds. This can effectively reduce the number of **false positives**.

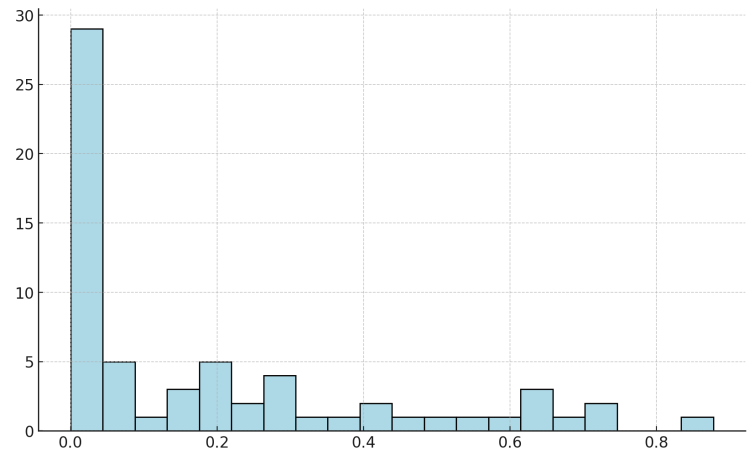

Figure 5: Distribution of Predicted Medals for No Medal Countries

According to the results, approximately eight countries that have yet to earn medals are most likely to achieve their first medal. The country names and the probabilities are listed in the figure below.

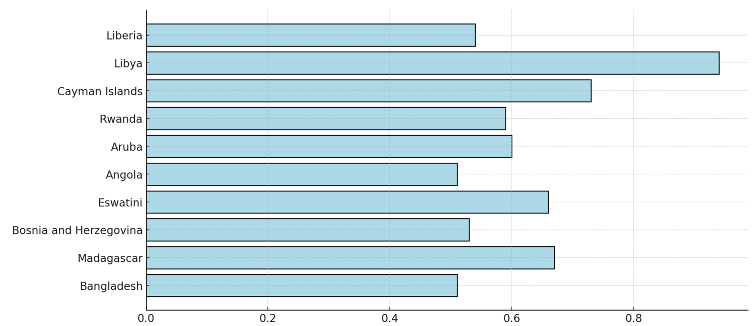

Figure 6: Probability of Achieving First Medal

4.6 Sub-Model: Relationship between events and countries medal earn

4.6.1 Feature Importance

We calculate the importance of each feature to identify those that are most critical for predicting gold and total medal counts. This analysis helps us to better explain the model's predictions and understand the role of each feature.

Feature	Importance
isHost	0.014923
G_i	0.237422
M_i	0.002371
Event	0.745284

Table 6: Gold Medal Prediction

Feature	Importance
isHost	0.003768
G_i	0.042192
M_i	0.394452
Event	0.559588

Table 7: Total Medal Prediction

We found that the importance of the host country variable was relatively low, while the event feature had an extremely high importance in our model. In response to this, we developed a sub-model to explore how these factors influence medal opportunities.

4.6.2 Using the Herfindahl Index to Analyze Medal Distribution

We applied the **HHI (Herfindahl Index)** to assess how the number of events impacts medal concentration. The formula for HHI is:

$$HHI = \sum_{i=1}^{n} s_{ij}^2$$

where s_{ij} is the proportion of medals won by country i in event j.

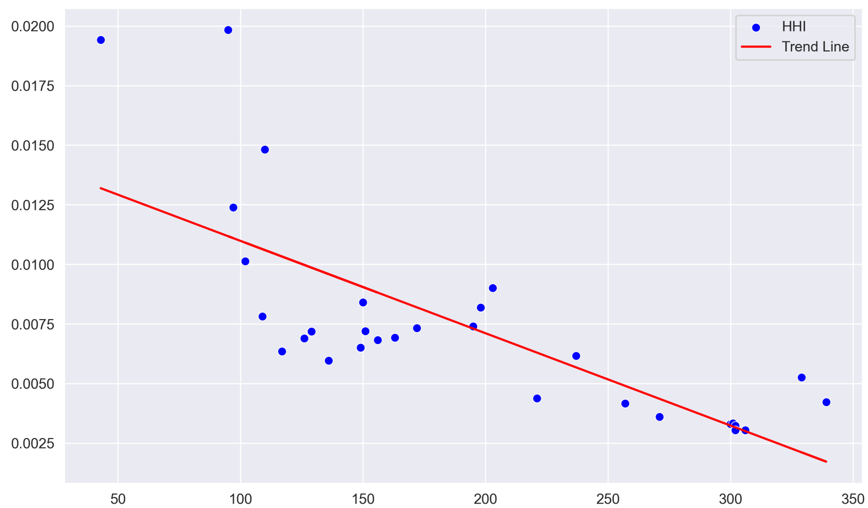

Figure 7: HHI Trend

The results of HHI showed that as the number of events increases, the distribution of medals becomes less concentrated, suggesting that more events create broader opportunities for various countries to win medals.

4.6.3 Identifying Strong Countries in Specific Sports

To identify dominant countries in particular sport, we designed the following formula:

$$\text{Strength}_{i,j} = \left(\frac{\text{Medals}_{i,j}}{\text{Total Medals}_i}\right) \times \left(\frac{\text{Medals}_j}{\text{Total Olympic Medals}}\right)$$

where $\text{Medals}_{i,j}$ is the number of medals country i won in event j, and Medals_j is the total number of medals awarded in event j. This allowed us to pinpoint the strongest sports for each country, highlighting areas where nations are more likely to succeed. For example, the top two strongest events for the United States are swimming (0.419) and athletics (0.34).

4.6.4 Examining Host Country Effects Using Spearman Correlation

To understand the host country effect, we calculated **the Spearnman correlation coefficient** between host countries' event selections and their medal outcomes:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

Here, d_i represents the difference in ranks between the event and medal count for the host country, respectively. And n represents the total number of events considered. In our analysis, only the USA and France were able to successfully calculate the correlation coefficients between the event selections and medal counts for each sport, along with their corresponding p-values. Despite Australia having hosted two Olympic Games, the quality of its data was too poor to support the calculation of Spearman correlation coefficients. This allows us to explore how the host country's event choices might influence its medal tally and impact other countries' chances in those events.

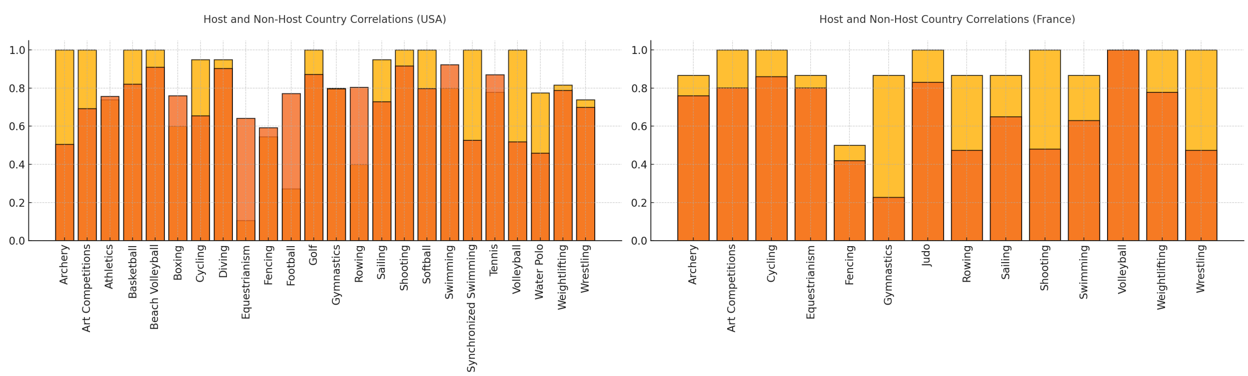

Figure 8: Host Country Correlation

From these two charts, it is evident that the host country effect is significantly manifested in most sports, with the host country's event selection having a notable impact on medal outcomes, making it an important factor influencing the results of the events.

5 Model II: "Great Coach" Effect Estimation Model

5.1 Feature Engineering

Using this scoring system, we can perform feature engineering to quantify the impact of a "great coach". Specifically, we define the following features:

- C_{nj}^{t}: Dummy variable. Indicating Whether athlete m has been directed by a "great coach" in event j by year t
- P_{nj}^{t-4}: Performance score of the athlete m in event j in previous $t-4$ year.
- P_{nj}^{t-8}: Performance score of the athlete m in event j in previous $t-8$ year.

- P_{nj}^{t-12}: Performance score of the athlete m in event j in previous $t-12$ year.
- N_{nj}^{t}: Athlete m's accumulated participation in event j by year t.

5.2 Linear Mixed-Effects Model

5.2.1 Model Building

In our dataset, the award outcomes of different athletes across various years and events exhibit a hierarchical structure and potential correlations, which is complex to analysis (Gałecki & Burzykowski, 2013). Therefore, we utilize **LMM (Linear Mixed-Effects Model)** to evaluate and estimate the impact of the "Great Coach" effect. The model is defined as:

$$\hat{h}_{ij}^{t} = \beta_0 + \beta_1 C_{nj}^{t} + \beta_2 P_{nj}^{t-4} + \beta_3 P_{nj}^{t-8} + \beta_4 P_{nj}^{t-12} + \beta_5 N_{nj}^{t} + u_i + \epsilon_{ij}^{t}$$

where $\hat{h}_{ij}^{t}$ is the predicted medal score for athlete i in event j, year t; β_0 is the intercept, β_1-β_5 are the fixed-effects coefficients corresponding to defined feature variables, u_i is the random effect, and ϵ_{ij}^{t} is the residual.

Based on the fact that Lang Ping and Béla Károlyi brought immense success to their respective teams after switching teams, the original dataset consists of the participation and medal results of Chinese and American women's volleyball players from 1984 to 2024, as well as American and Romanian gymnasts from 1960 to 1980, including the specific events and medal achievements in the Olympic Games.

5.2.2 Model Validating and Training

We initially fitted a LMM to the dataset and assessed whether its underlying assumptions were satisfied. By constructing box-plots and identifying outliers based on quartiles and the **IQR (Inter-Quartile Range)**, we subsequently removed these outliers.

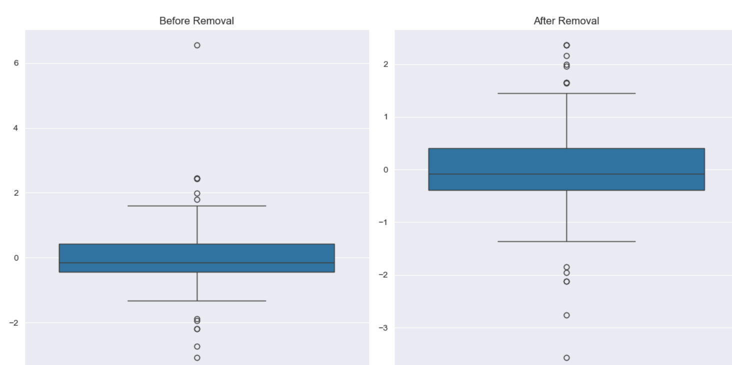

Figure 9: Residuals Box-plot

We then generated diagnostic plots involving residuals, normality, random effects, and fitted vs. observed values. In the plot, the residuals do not exhibit any specific pattern with the fitted values, and in the normal Q-Q plot, the data points are close to the diagonal, indicating that they approximately follow a normal distribution.

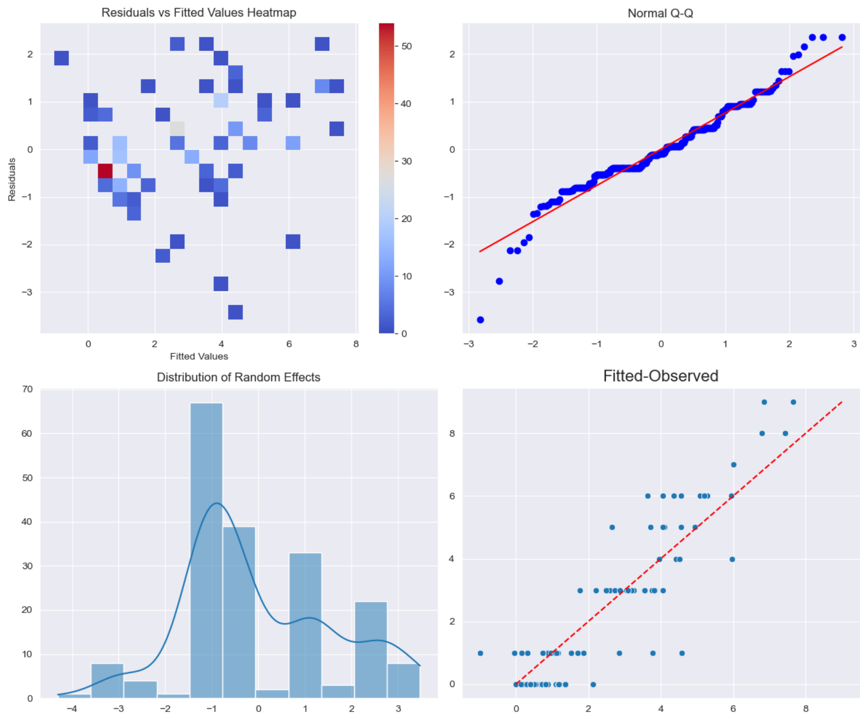

Figure 10: Feasibility Validating

Although the random effects distribution exhibited skewness (left-skewed), we further validated the significance of the random effects by fitting an additional LMM without random effects and performing a **LRT (likelihood ratio test)**. The test yielded a **LRT Statistic** of 51.36 with P-value $= 7.6676130e - 13 << 0.001$, which provides strong evidence to reject the null hypothesis of no random effects, thus confirming their significance. Hence, we have demonstrated the feasibility of applying the LMM to the dataset.

5.2.3 Regression Results Interpretation

The mixed linear regression results from the training indicate that the model has successfully converged to a linear form. At the same time, the variable C_{nj}^t was found to be statistically significant, with a P-value of $0.000(z = 6.720)$. This result strongly suggests that coaching status has a notable impact on the outcome, as the null hypothesis is rejected at the usual significance levels. Therefore, we successfully examined the positive effect of a "great coach" effect. Specifically, the model estimates that the effect of a "great coach" on performance can increase an athlete's score by approximately 1.088 units, assuming other parameters remain constant. The medal score conversion indicates that a "great coach" could approximately provide the athlete with one additional bronze medal in an event, equivalent to 0.3627 bronze medals, or 0.2176 gold medals.

5.2.4 Model Evaluation

In order to calculate the fitness for the mixed effects model, Conditional R^2 is chosen, and its formula is as below.

$$R^2_{\text{conditional}} = \frac{\text{Var}(Y_{\text{fixed}} + Y_{\text{random}})}{\text{Var}(Y_{\text{total}})}$$

The model performs well, with a Conditional R^2 of 0.8950, indicating that it explains 0.895 of the variance in the response variable. Meanwhile, the MSE of 0.5967 and RMSE of 0.7725 suggest that the model's predictions are accurate, with small errors on average. Overall, the model effectively captures the data's variation and provides reliable predictions.

Metric	Value
R^2	0.8950
MSE	0.5967
RMSE	0.7725

Table 8: Model Evaluation Metrics

5.3 Multi-Method Integrated Evaluation Model

To ensure enough data, we choose these 3 countries: Japan, Great Britain and France. To choose alternatives among various sports, we choose the **combined weight** of the **Analytic Hierarchy Process** (De FSM Russo & Camanho, 2015) and **Entropy Weight Method** to evaluate and apply **TOPSIS** method to choose the best 3 sports of each country. This evaluation model balances subjective judgments and objective relationships between data.

5.3.1 Judgment Matrix Construction and AHP Weight

This method implements subjective weights between criteria. A judgment matrix is constructed to represent the pairwise relative importance of criteria. We choose 4 **benefit** criteria for the AHP method that may enhance the "great coach effect": Participants, Athletes' experience, Behavior of the last game, and Behavior of the penult game. Reasons: The effect can be multiplied if many athletes receive the guidance. If the athletes are experienced and competitive, the coach can apply targeted training, which will be effective.

The criteria in the matrix follow the order: Participants, Behavior of the last game, Behavior of the penult game, Athletes' experience.

$$A = \begin{bmatrix} 1 & \frac{1}{9} & \frac{1}{5} & \frac{1}{5} \\ 9 & 1 & 3 & 3 \\ 5 & \frac{1}{3} & 1 & 1 \\ 5 & \frac{1}{3} & 1 & 1 \end{bmatrix}$$

In constructing the matrix for the AHP, the criterion "Behavior of the last game" is considered more important than "Behavior of the penult game" because the most recent behavioral data reflects the athletes' current condition, which is crucial for prediction. "Behavior of the penult game" and "Athletes' experience" are both reflections of athletes' experience and ability, so they are considered of the same importance. For the "Participants", we do not want events of team competition to take the dominant place, so we put it as the lowest.

The AHP weight is defined as follow:

$$W = \{v/|v| \mid Av = \lambda_{max} v\}$$

Compute the largest eigenvalue and corresponding normalized eigenvector. The consistency check value **CR = 0.0123**, so the matrix is valid, and we can use the eigenvector as **AHP weights**:

$$W^{AHP} = \begin{bmatrix} 0.0564799 & 0.53746027 & 0.22844587 & 0.22844587 \end{bmatrix}$$

5.3.2 Data Normalization and Entropy Weight

Entropy weights are computed to reflect the degree of information diversity in each criterion. High entropy indicates less importance to decision.

$$X_{ij}^{norm} = \frac{X_{ij} - X_{j_{min}}}{X_{j_{max}} - X_{j_{min}}} \longrightarrow P_{nj} = \frac{X_{ij}^{norm}}{\sum_{i=1}^{m} X_{ij}^{norm}}$$

$$\downarrow$$

$$W_j = \frac{1 - E_j}{\sum_{j=1}^{n}(1 - E_j)} \longleftarrow E_j = -k \sum_{i=1}^{m} P_{nj} \ln(P_{nj})$$

Figure 11: Algorithm of Entropy Weight Method

Convert measurement data into a common scale $X_{ij}^{norm} \in [0,1]$ to ensure **unbiased comparisons**. Then, compute the probability distribution P_{nj} and calculate entropy E_j. Finally, get the weight of EWM.

where m is the total number of alternatives. n is the total number of criteria.

$k = \frac{1}{\ln(m)}$ is a constant to normalize the entropy to the range $[0, 1]$.

When $P_{nj} = 0$, define $P_{nj} \cdot \ln(P_{nj}) = 0$.

$1 - E_j$ represents the information utility of the j-th criterion.

We get the **EWM weights**:

$$W^{EWM} = \begin{bmatrix} 0.14199852 & 0.23448226 & 0.28461418 & 0.33890504 \end{bmatrix}$$

5.3.3 Weights Combination

The subjective weights (from AHP) and objective weights (from EWM) are combined as follows:

$$W_j^{combined} = \frac{W_j^{AHP} \cdot W_j^{EWM}}{\sum_k W_k^{AHP} \cdot W_k^{EWM}}$$

We get the **combined weights**:

$$W^{combined} = \begin{bmatrix} 0.00297847 & 0.46802898 & 0.24146614 & 0.28752641 \end{bmatrix}$$

5.3.4 Alternatives Selection by TOPSIS

This method ranks alternatives based on their distance to the ideal solution. Alternatives closer to the positive ideal solution and further from the negative ideal solution will receive a higher score.

$$V_{ij} = W_j^{combined} \cdot X_{ij}^{norm} \longrightarrow \begin{array}{l} A^+ = \{\max(V_{ij}) \mid j \in J^+; \min(V_{ij}) \mid j \in J^-\} \\ A^- = \{\min(V_{ij}) \mid j \in J^+; \max(V_{ij}) \mid j \in J^-\} \end{array}$$

$$\downarrow$$

$$C_i = \frac{D_i^-}{D_i^+ + D_i^-} \longleftarrow \begin{array}{l} D_i^+ = \sqrt{\sum_{j=1}^{n} (V_{ij} - A_j^+)^2} \\ D_i^- = \sqrt{\sum_{j=1}^{n} (V_{ij} - A_j^-)^2} \end{array}$$

Figure 12: Algorithm of TOPSIS

Prepare the weighted normalized decision matrix V_{ij} for future calculations. Then, find the positive A^+ and negative ideal solution A^- and calculate the distance for each alternative to the two solutions. Finally, compute the relative closeness to the ideal solution.

5.3.5 Evaluation Results

Finishing the above process, we get the C_i of each alternative. we rank them according to the C_i and choose the top 3 sports of each country.

Country	Sport1	Sport2	Sport3
France	Judo	Basketball	Volleyball
Great Britain	Equestrian	Cycling Track	Artistic Gymnastics
Japan	Judo	Wrestling	Artistic Gymnastics

Table 9: Country-Sport

5.4 Forecasting Results

Based on the filtered results from the integrated evaluation model, we set the corresponding C_{nj}^t in the dataset to 0 and 1, respectively, and made predictions for each event under each sport. By calculating the difference between the treated and untreated predicted results, we eliminated the effect of fixed effects. Thereby, focusing on the impact of a "great coach" effect. Finally, we determined the impact of the "great coach" effect on three countries and the sports in which they should consider investing in terms of scores.

Country	Sport	Predicted Coach Effect (per Event)	Predicted Coach Effect (Sport)
France	Basketball	+0.6697	+1.3395
France	Judo	+0.5217	+7.8262
France	Volleyball	+0.6850	+1.3699
Great Britain	Artistic Gymnastics	+0.7282	+11.6508
Great Britain	Cycling Track	+0.6711	+8.0526
Great Britain	Equestrian	+0.5052	+3.0309
Japan	Artistic Gymnastics	+0.6882	+11.0108
Japan	Judo	+0.3104	+4.6566
Japan	Wrestling	+0.4016	+5.2208

Table 10: Effect of Coached Training on Medal Prediction by Country and Sport

6 Original Insights

6.1 Medal Distribution for Events

We analyzed medal distribution across sports in past Olympics and created a word cloud. Understanding medal dispersion helps Olympic committees reallocate resources to enhance medal potential. For host nations, recognizing sports with high medal dispersion informs strategic decisions in event addition and resource allocation, boosting competitiveness in key areas.

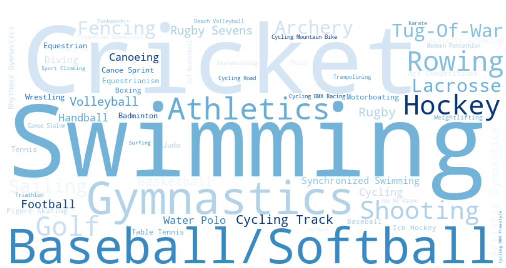

Figure 13: Sports Word Cloud

6.2 Great Coach's Future Effect

In the second model, we estimate the impact of the "Great Coach Effect" in the next game. When we take a longer-term perspective, we find that after these great coaches finish their tenure, the medal performances remain significantly better compared to before their tenure. The effect of great coaches continues to persist after their departure. This suggests that they may have established profound systems, cultures or methodologies during their tenure, which have a lasting positive impact on subsequent development.

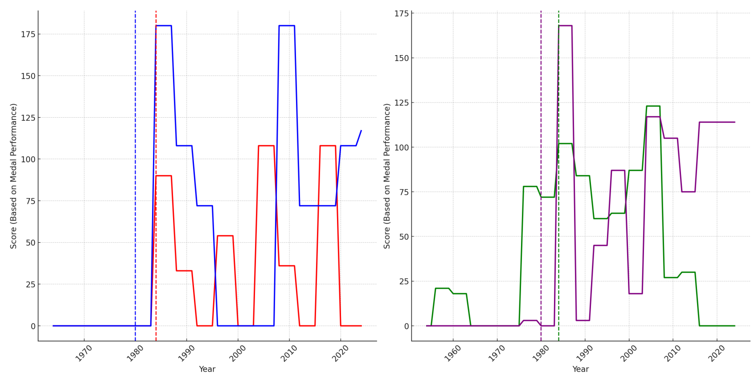

Figure 14: Long-term effect

6.3 Geographical Medal Prediction for the 2028 Olympics

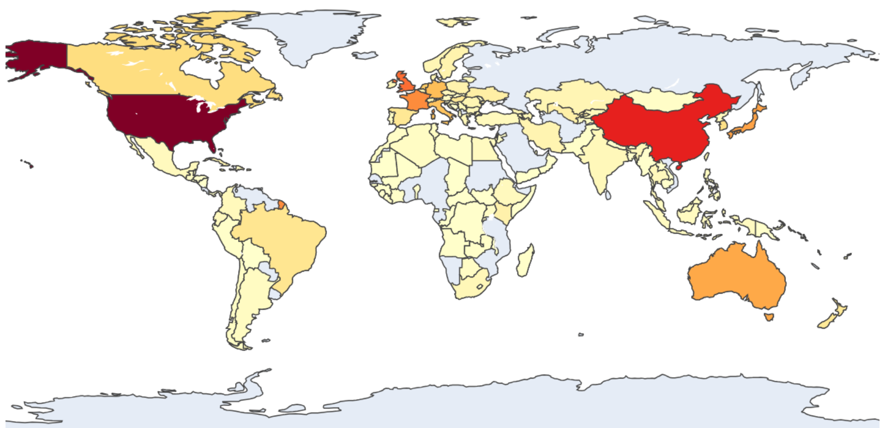

Figure 15: Geographical Distribution of Medal

We perform a geographical heatmap of predicted medal counts for the 2028 Olympics. From the figure, we find out that the United States and China are expected to dominate the 2028 Olympics medal table, with strong performances across various events. European countries like the UK, Germany, and France will also maintain high medal counts. In contrast, South America, Africa, and Oceania show weaker overall competitiveness, with countries like Brazil and those in Africa excelling in specific events but struggling to compete with the top nations.

6.4 Insights for country Olympic committees.

To enhance the diversity of medal distribution, the International Olympic Committee should introduce a wider variety of events. As the number and types of events increase, the medal distribution will become more dispersed, allowing more countries to compete for medals. For the host country's Olympic Committee, there should also be diversity in event selection. Due to the home advantage and other objective factors, the likelihood of winning in many events will naturally increase. For Olympic Committees of non-host countries, those with fewer medals should consider diversifying and developing unique events, while also inviting renowned coaches to improve the performance level. For countries that have already won a significant number of medals, in addition to the previous suggestions, they should continue to strengthen training in their dominant events to maintain their competitive edge.

7 Sensitivity Analysis

In our analysis of model I, we examined the feature importance of different variables. We found that the features G_i (gold medals) and M_i (total medals) demonstrated significant importance. To test the reasonableness and generalizability of our XGBoost prediction model, we conducted a sensitivity analysis on these two important feature variables. Specifically, we adjusted G_i by ± 0.25 gold medal and M_i by ± 0.5 total medals, observing the impact of these adjustments on the prediction results. This approach allows us to assess the model's sensitivity to these key features and verify its stability and predictive ability under varying data conditions.

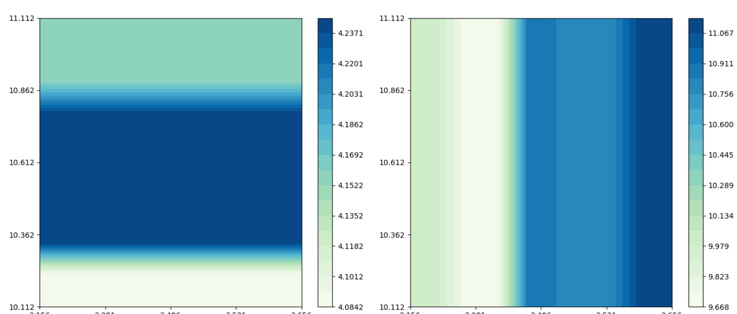

Figure 16: Sensitivity Analysis

From the graph, it is evident that the two selected features, G_i and M_i, do indeed influence the predicted medal count when they vary. However, during data collection, the average fluctuations of gold medals and total medals did not exceed the limits set in the sensitivity analysis. This indicates that the model maintains a high level of accuracy within the fluctuations of these features. Therefore, it can be concluded that our model successfully passed the sensitivity test and demonstrates stable predictive power within the defined parameter variation range.

8 Strengths and weaknesses

8.1 Strengths

- By combining our XGBoost model with Bayesian Optimization (TPE) for hyperparameter tuning, we improve the accuracy of Olympic medal predictions. TPE efficiently identifies optimal hyperparameters, allowing the model to better capture non-linear relationships in the data. This results in more precise predictions, as evidenced by the lower RMSE values compared to traditional tuning methods.

- Our model development and evaluation process is thorough, ensuring no feature, even those deemed insignificant, is ignored. For instance, the host country effect, though insignificant in the main model, was found to have a meaningful impact in a sub-model. This approach helps us better understand each variable's role.

- By quantifying medals into scores, it becomes possible to more evenly capture the overall scores of athletes, thereby more accurately reflecting their specific performance levels. At the same time, this can help balance differences between various sports to some extent. The organizing committee can make adjustments based on the characteristics of each sport, making comparisons between different sports fairer.

8.2 Weakness

- Due to time and dataset constraints, our model lacks some comprehensiveness. For example, external factors like geographical and political influences were not considered. In future optimizations, we plan to build a more comprehensive model that incorporates these factors to achieve more accurate results.

- Despite the fixed-effect coefficient of C_{nj}^{t} being regressed to be 1.088 with high significance, the final predicted result for the target data shows that the average coach effect at the project level is approximately 0.575673, which differs by about 0.5123 from the predicted fixed effect coefficient. Considering that the model has successfully passed all linear assumptions and the variance inflation factors (VIF) for each variable are no greater than 10, indicating no multicollinearity issues, the noticeable discrepancy may stem from the omission of important variables that influence the coach effect, such as the coach's years of experience and professional qualifications. This issue could be addressed with ease, given the dataset at hand.

Variable	VIF
isCoached_cumulative	1.063421
ExperienceScoring	1.546109
Scoring_Last4	1.355976
Scoring_Last8	1.112936
Scoring_Last12	1.038160

Table 11: Variance Inflation Factors (VIF) for the Variables

References

[1] Allen, S. V., & Hopkins, W. G. (2015). Age of peak competitive performance of elite athletes: A systematic review. *Sports Medicine*, *45*(10), 1431–1441. https://doi.org/10.1007/s40279-015-0354-3

[2] Knights, S., Sherry, E., & Ruddock-Hudson, M. (2016). Investigating elite end-of-athletic-career transition: A systematic review. *Journal of Applied Sport Psychology*, *28*(3), 291–308. https://doi.org/10.1080/10413200.2016.1162222

[3] Wikipedia contributors. (2025, January 24). National Olympic Committee. In *Wikipedia*. https://en.wikipedia.org/wiki/National_Olympic_Committee

[4] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). https://doi.org/10.1145/2939672.2939785

[5] Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., & De Freitas, N. (2015). Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, *104*(1), 148–175. https://doi.org/10.1109/JPROC.2015.2494218

[6] De F. S. M. Russo, R., & Camanho, R. (2015). Criteria in AHP: A systematic review of literature. *Procedia Computer Science*, *55*, 1123–1132. https://doi.org/10.1016/j.procs.2015.07.081

[7] Gałecki, A., & Burzykowski, T. (2013). *Linear mixed-effects models using R: A step-by-step approach*. Springer Science & Business Media.

AI Generated Content

- OpenAI ChatGPT (Jan 25, 2023 version, ChatGPT-4)
  Query1: Here is a file containing complete country medal count tables for all summer Olympics from 1896 to 2024. Upload summerOly medal counts.csv. Can you standardized the NOC for me. You can check the wiki for NOC references.
  Output: Provided the file with standardized NOC.

- OpenAI ChatGPT (Jan 25, 2023 version, ChatGPT-4)
  Query2: Can you merge the data for me. Upload the data file of our model prediction of gold and total medal.
  Output: Provided the file with the request.

- OpenAI ChatGPT (Jan 25 - Jan 28, ChatGPT-4)
  Queries for Latex code debugging

- OpenAI ChatGPT (Jan 25 - Jan 28, ChatGPT-4)
  Queries for Python code debugging