Prediction for future events based on dual-mechanism, ARIMA and Monte Carlo simulation

Yifan Li#

Department of Mathematics

Southern University of Science and TechnologyShenzhen, China

Yibin Lou#,\*

Department of Computer Science and Engineering

Southern University of Science and Technology

Shenzhen, China

12310513@mail.sustech.edu.cn

Yuhan Rui#

Department of Computer Science and Engineering

Southern University of Science and Technology

Shenzhen, China

#These authors contributed equally.

Abstract—Our study proposes a novel multi-objective optimization framework for addressing complex decision-making problems under uncertainty, leveraging historical data and predictive modeling techniques. The research focuses on developing a robust methodology that integrates data preprocessing, clustering analysis, and advanced predictive models to optimize resource allocation and performance forecasting across multiple objectives. Initially, the framework preprocesses heterogeneous datasets, accounting for missing data and reallocating variables based on proportional distributions. Subsequently, K-means clustering is employed to categorize variables into six distinct groups based on correlation patterns, enhancing the granularity of the analysis. Predictive modeling is achieved through a hybrid approach combining linear regression with mean-based adjustments for stable trend predictions, and the ARIMA model for capturing temporal dependencies in non-stationary data. Additionally, Monte Carlo simulations and random forest algorithms are utilized to handle stochasticity and non-linear relationships, respectively. The integration of these models is optimized using a Poisson distribution to establish confidence intervals, ensuring a comprehensive assessment of uncertainty. This multi-model synergy significantly improves computational efficiency, solution feasibility, and the reliability of the optimization outcomes, providing a versatile tool for tackling multifaceted optimization challenges in various domains.

Keywords—dual-mechanism, linear regression, ARIMA, random forest, Monte Carlo simulation.

# Introduction

This study presents an advanced predictive model designed to forecast competition outcomes in complex, evolving environments. The model addresses the critical challenge of making accurate predictions when the composition of competition categories undergoes frequent changes, with new categories being introduced and existing ones modified over time. Such dynamic conditions significantly increase the complexity of outcome forecasting.

Previous approaches to this problem [1] have typically incorporated macroeconomic indicators and resource allocation factors as primary predictors. However, these variables often prove unreliable due to their susceptibility to external influences and geopolitical factors, making them poor predictors for future events. Our methodology addresses these limitations by focusing instead on historical performance data and participant metrics, which demonstrate greater stability and predictive power.

Existing prediction models [2] have primarily concentrated on aggregate outcome totals, resulting in substantial prediction variance. Through our analysis, we identified significant performance correlations within related competition categories. This finding led to our innovative categorical approach, where we classify all competition types into six major groups based on the pattern which is generated by historical performance. This classification is based on K-means clustering method. By predicting outcomes for each category separately and aggregating the results, our model achieves substantially reduced variance compared to conventional methods.

Traditional prediction techniques [3] have predominantly employed linear regression models, including high-dimensional variants [4]. While these methods can be effective for categories with clear performance trends, they often produce unstable predictions for categories with more volatile outcomes. Our model overcomes this limitation through a novel dual-mechanism approach: applying regression analysis for categories with strong correlation of historical performances, while utilizing average for categories exhibiting greater fluctuation. This dual-mechanism approach can significantly enhance the stability and the reliability of our prediction model.

The results demonstrate that our model outperforms previous approaches by comprehensively accounting for historical performance patterns and the impact of category changes. By employing this categorical framework and dual-mechanism prediction system, we provide a more robust and accurate forecasting tool for competitive environments with evolving category structures. This advancement offers significant improvements in prediction reliability and has important implications for strategic planning and resource allocation in competitive fields.

# Model

## Clustering Analysis

Considering the correlation between different sports events and the types of sports that each country excels in (as can be seen from the database, some countries only win medals in specific events), it is necessary to classify the more than seventy events [5]. Clustering analysis [6] helps identify these strengths and groups similar events together.

We construct a matrix with countries as rows and events as columns, where the values represent medal-winning rates. K-means clustering [7] uses these rates to group similar events. According to some literature insights [8], the optimal number of clusters,. Misclassified points are manually corrected for accuracy.

Also, we separate the analysis into two parts: the old events before 2024 and the new events added in 2028. For the historical events, we fully utilize the data from [https://www.contest.comap.com/undergraduate/contests/mcm/contests/2025/problems/](https://www.contest.comap.com/undergraduate/contests/mcm/contests/2025/problems/" \t "_new). We apply linear regression with means to predict the outcomes. While for the 5 new events added in 2028, we calculate predictions based on authoritative sources from the website https://www.olympics.com/en/sports/).

## Linear Regression Model

Given that athletes or teams in one specific country often retain their medals in consecutive games, our strategy is to use the results from the previous games to predict the performances of athletes in the upcoming games.

Let i represent the categories of sports, then

(1)

where Water Sports, Ball, Sports, Track and Field, Technical Events, Combat Sports, and Racing Events correspond to numbers 1 to 6.

Let j represent the outcomes in the next Olympic Games. Then

. (2)

Let k represent the k-th Olympic games since 2000. Then

. (3)

Symbols a, b, c, d, e respectively represents the medal types (or situation) for the athletes in the previous Games: a represents gold medals, b represents silver medals, c represents bronze medals, d represents non-medal athletes, e represents athletes who did not participate in the previous Games.

To explore whether the number of medals in consecutive Olympic Games is related to the year, we propose different prediction algorithms:

1. If the medal count shows a strong correlation with the year, we use linear regression to predict.

2. If the correlation is weak, we use the average of the data as the prediction.

We use 0.5 as the threshold for determining strong or weak correlation.

Take golden medal as an example, the predicted number of golden medals in 2028 is defined as:

(4)

where is the predicted number of gold medals in category i for the 2024 Games, and r is the correlation coefficient for , where k = 1, 2, ...6. The linear regression value is:

Since this value may be negative, we take the maximum between the linear regression value and 0.

Similarly, we apply the same methodology for , , and , as well as their total predicted values , , .

For athletes who did not participate in the previous Olympic Games but are expected to win medals in the upcoming Games, we define as the total number of athletes in category i who did not participate in the previous Games but win medal j (where 0 ≤ j ≤ 3). The prediction for the number of such athletes is defined as:

(6)

where r is the correlation coefficient for and the linear regression value is:

(7)

Since we have completed discussing the prediction of old events for medal-country, we then move on to predict the probability of non-medal countries winning their first medal of old events in the 2028 Los Angeles Olympics.

### The number of events participated in the old event of all the non-medal country in 2028

Previous study shows that the number of events participated of a country has a strong correlation to the number of its medals, so we shall use the data on the number of events participated by non-medal countries in the past few Olympic Games. Similarly, based on the number of participated events of each country from 2000 to 2024, we select whether to use a specific model which obtains the average value or linear regression by examining the correlation, and we obtain the predicted participation numbers for 2028.

where represents the number of events that non-medal country i participated in at the 1996 + 4j year Olympics. r is the correlation coefficient for  with k = 1~7. The linear regression value is calculated as:

(9)

Next, we calculate the total number of events of non-medal country participated in the next Olympics:

### The number of new award-winning countries

For non-medal countries, we consider the number of countries that win their first medal in each Olympic Games for old events: from 1984 to 2024, the data shows 4, 5, 8, 16, 6, 5, 6, 8, 3, 0, and 4. The average value is 5.91, and the standard deviation is 3.85. It is clear that the difference between the number of first-time medal winners in 1996 (which was 16) and the average is significantly higher than twice the standard deviation, so we need to explain why so many non-medal countries won their first medals in that year, thus reconstruct our model.

According to dataset, there is an anomalous increase in the number of first-time positive occurrences in a specific period. This anomaly corresponds to a structural shift in the dataset, where several new categories emerged simultaneously.

To address this, we segment the entities into two categories based on their appearance timeline in the dataset:

(1) Emerging entities that were introduced after a certain cut-off point, here as Class A countries.

(2) Established entities that have historical representation across earlier periods, as Class B countries.

By modeling these two groups separately, we reduce the variance introduced by structural noise and ensure better alignment with the assumptions of time-series forecasting models. This stratification also allows us to apply tailored predictive strategies—such as exponential smoothing for rare emerging entities and ARIMA-based regression for well-established classes—thereby improving both short-term prediction accuracy and long-term trend estimation [9].



Let i represent the i-th order of Olympic since , when emerging entities appear.

(12)

For established countries, based on historical data, we apply a combination of the ARIMA model and Random Forest model [10], using time series data and feature engineering to generate prediction results and their confidence intervals.

The reason for applying the model to address this issue is that the number of countries winning their first medal at the Olympics constitutes typical time series data with an underlying trend (the number of participating countries increases, the number of old events grows, etc.), occurring in an interval which will be determined.

Compared to other models, a key advantage of ARIMA is its ability to handle non-stationary time series by applying differencing to transform them into stationary sequences for forecasting. For Olympic data, external factors may introduce fluctuations. ARIMA’s flexibility allows it to adapt to such uncertainties, delivering reliable predictions. The ARIMA model effectively captures these characteristics and is utilized to predict the data for 2028.

Based on the dataset, we split the data from 1896-2004 and 2008-2024 into training and testing sets, with each Olympic year as a time point. The target variable is the number of countries winning their first medal. After performing first and second-order differencing on the original target variable, we conduct an ADF test to verify the stationarity of the data. The ADF test results indicate that the original series is the most stationary.

Next, we plot the ACF and PACF of the original series. The tailing patterns in the ACF and PACF indicate long-term dependencies, leading us to select the ARIMA model for modeling. To determine the optimal parameters for the ARIMA model, we use the AIC criterion.

The AIC is calculated as

where K is the number of model parameters and L is the maximum likelihood value. By plotting the AIC heatmap, we find that when p = 0 and q = 2, the AIC value is minimized, so we select ARIMA (0,0,2) as the optimal model.

We validated the ARIMA model’s fit using residual checks: ACF/PACF plots and the Ljung-Box test showed no autocorrelation (p > 0.05), while a histogram, QQ plot, and Shapiro-Wilk test confirmed normality (p > 0.05). This indicates a good fit, satisfying model assumptions.

For the countries emerging from disintegration after 1996, the number of new countries winning their first medal is as follows: 11, 2, 1, 1, 1, 0, 0, 0, forming an apparently monotonically decreasing sequence approaching 0. The original data, first-order difference, and second order difference fail to pass the stationarity test. Considering that 18 of the 21 disintegrated countries have already won medals, we use an exponential regression model to predict a 95% confidence interval of [0, 0.1], which is shown in Figure 1. Therefore, we conclude that the number of countries from disintegrated nations winning their first medal in 2028 will be 0, namely u = 0.

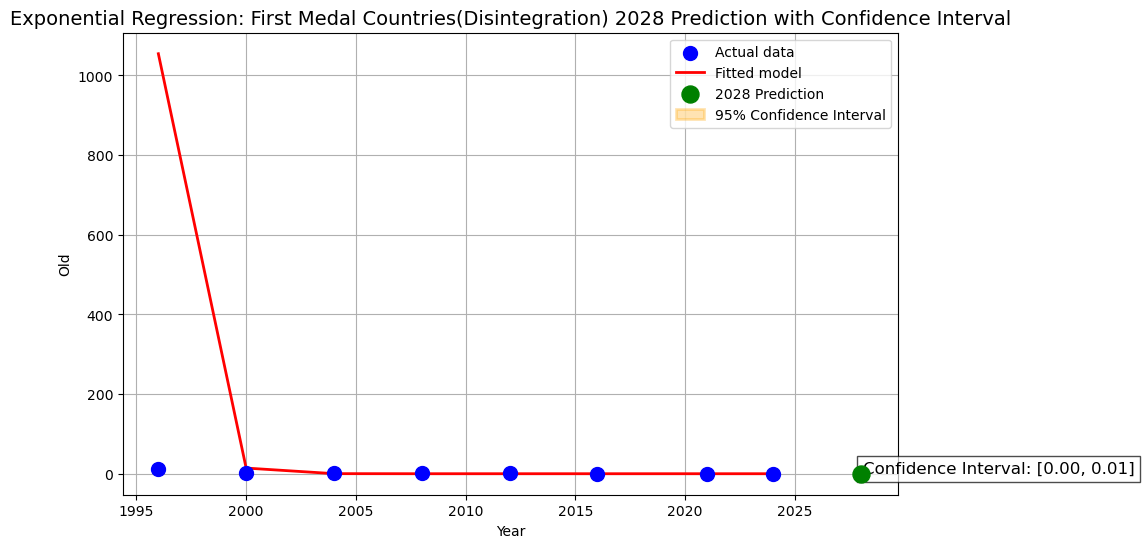


Figure 1. Using AMIRA to predict the number of countries winning their first medal in the next Olympics

Thus, we can calculate the probability of country i (which has not won a medal before) winning its first medal in the 2028 Los Angeles Olympics in old events as

Since the number of events is large and the probability of winning in each event is sufficient low approaching zero, the distribution of Li can be approximated by a accumulation of Poisson distribution: Poisson (Li). From the Poisson distribution formula, the winning probability  is the complement of the probability of zero medals.

## Monte Carlo Simulations

After discussing the prediction of old events, we are about considering the new events. Since it lacks data from previous Olympics, we search some relevant global events. By gathering authoritative data on rankings and points for each country in these events, we compute the winning probability for each country:

where and are the ratings of countries A and B, respectively. We simulate the events according to the Olympic competition rules, where the top 16 athletes are selected and paired for a knockout tournament. After two rounds, the top 4 athletes (or teams) will compete for the medals.

These 5 new events are split into male and female categories, making a total of 10 events. To obtain more accurate probabilities, we simulate each event for 1 million times through Monte Carlo Simulations. The results are then used to estimate each countries expected medals for these new events, denoted as , and .

For some events, due to the lack off official data, so we refer to the available world rankings. We use logistic regression substituting the previous formula to assess the performance gaps between athletes:

(17)

where and are the rankings of countries A and B respectively. Thus the probabilities of each athlete (or team) winning gold, silver, and bronze can be calculated. Finally, for the new events, the probability of each non-medal country earning can be estimated as follows.

### Predictions of new events for non-medal countries

There are some countries who have a chance of winning therir first medal in the new events. Based on the previous Monte Carlo simulation, the probability of these countries winning medals of each new events is derived, denoted . The case of the other countries is simple as followed.

(18)

# Results and analysis

## Data Processing

Countries are divided into two categories: countries have not won any medal in history (we call them non-medal- countries) and countries has won medal(s) in history (we call them medal-country). For non-medal-countries, we predict their odds of winning their first medal in the next Olympics. For medal-country, we construct a model to predict the number of medals of each country in the next Olympics.

When predicting the number of medals for a disintegrated country, the data for these new countries before disintegrated is allocated based on the proportion of the total number of medals after the disintegration of the old country. In addition, we optimized the model to handle missing data from countries that skipped certain competitions (such as Russia's absence from the 2024 Paris Olympics).

Events are divided into deleted events, old events, and new events. According to the IOC (International Olympic Committee), five events which will not appear in the 2028 Los Angeles Olympic will be removed in our prediction model. For old events, we use medal data from previous Olympics to make predictions. Due to the lack of historical data for the new Olympic events, we use the current country rankings or cumulative scores for our predictions.

## Sports Distribution Classification (K-means)

The clustering visualization is shown in Figure 2:

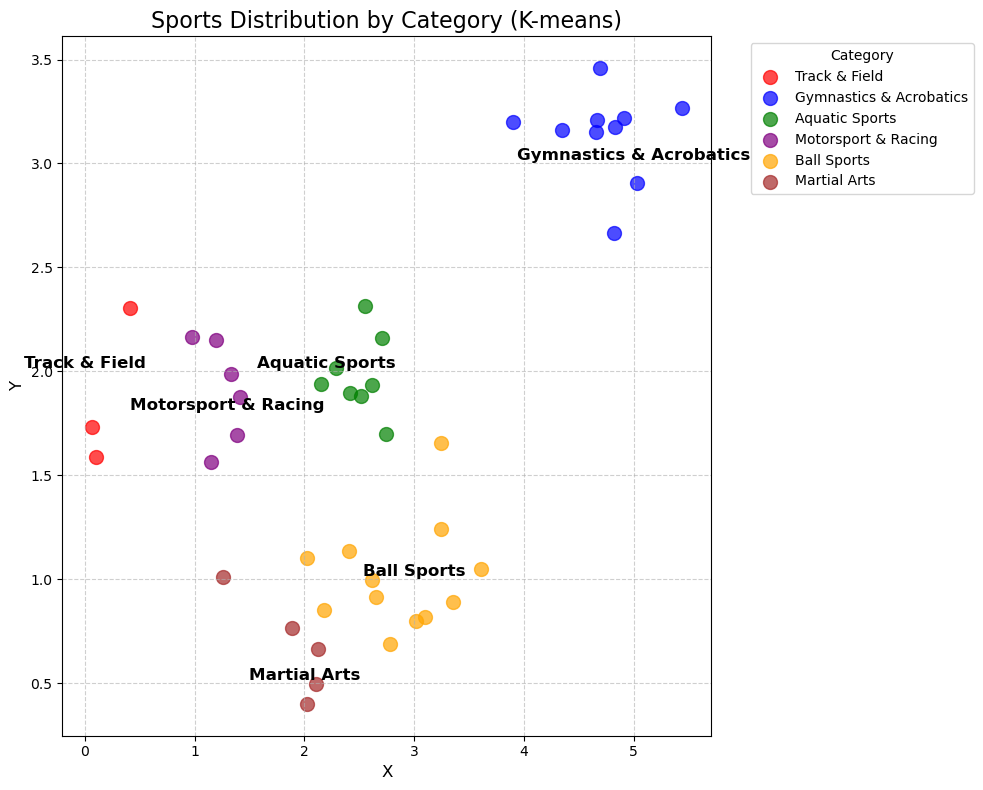


Figure 2. The 6 major categories of events grouped by K-means algorithm

## Old events prediction

For events that have appeared in previous Olympic Games, we use linear regression combined with mean method to predict the number of gold, silver, and bronze medals. The top 10 forecast results are shown in the Table I:

Table I. Predicted results of the top 10 countries with the highest total number of medals won in old events

| Country | Medal | | | |
| --- | --- | --- | --- | --- |
| Gold | Silver | Bronze | Total |
| USA  CHN  GBR  RUS  FRA  JPN  AUS  GER  NED  ITA | 44.9  33.5  24.1  21.0  21.8  14.8  16.0  12.9  14.6  11.9 | 36.5  26.3  11.5  21.9  22.2  21.8  16.1  16.5  9.6  12.7 | 40.3  27.5  19.2  25.9  16.2  21.8  16.0  14.7  14.2  13.8 | 121.7  87.3  54.8  68.8  60.2  58.4  48.1  44.1  38.4  38.4 |

## New events prediction

For the events that will appear for the first time in 2028, the result of the top 10 countries that have won gold, silver, and bronze medals in the new events using Monte Carlo simulations are in Table II:

Table II. Predicted results of the top 10 countries with the highest total number of medals won in new events

| Country | Medal | | | |
| --- | --- | --- | --- | --- |
| Gold | Silver | Bronze | Total |
| EGY  GBR  USA  AUS  JPN  CAN  NZL  MEX  PUR  IND | 1.4040.8680.9370.5930.57  0.5770.3830.3150.2040.207 | 1.309  0.952  0.735  0.628  0.578  0.528  0.381  0.308  0.216  0.198 | 1.304  0.962  0.706  0.632  0.584  0.51  0.381  0.308  0.217  0.198 | 4.017  2.782  2.378  1.853  1.732  1.615  1.145  0.931  0.637  0.603 |

Combining old events with new events, we finally calculate the total medals court for each country. The top 10 forecast results are shown in the Table III:

Table III. Predicted results of the top 10 countries with the highest total number of medals won

| Country | Medal | | | |
| --- | --- | --- | --- | --- |
| Gold | Silver | Bronze | Total |
| USA  CHN  GBR  RUS  FRA  JPN  AUS  GER  ITA  NED | 45.8  33.6  21.9  21.8  14.9  24.6  16.5  13.1  12.0  14.7 | 37.2  26.4  22.8  22.2  22.0  12.1  16.7  16.7  12.8  9.7 | 41.0  27.5  26.9  16.2  22.0  19.8  16.7  14.8  14.0  14.3 | 124.0  87.5  71.6  60.2  58.9  56.5  49.9  44.6  38.8  387 |

## The Host effect

The host country factor in the Olympics cannot be ignored. Table IV shows the medal counts for new events by host countries from 1984 onwards.

Table IV. Medal counts for the new events by the host countries

| Medal | Year | | | | |  | |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2004 | 2008 | 2012 | 2016 | 2020 | | 2024 | | |
| Gold  Silver  Bronze | 0  0  0 | 2  1  0 | 2  0  1 | 1  0  0 | 7  2  1 | | 1  4  2 | | |

By analyzing the data from 1896 to 2024, it is clear that the host country’s medal count, both total and gold, has generally increased.

Therefore, we predict the United States’ medals in the 2028 Los Angeles Olympics separately for old and new events. First, we calculate the increase in medals for traditional events based on the host country’s performance in previous Olympics. This is done by comparing the number of medals won in the host countries traditional events with the average number of medals won in the two previous Olympics, then calculating the rate of increase.

Old events for countries with missing data from the previous two Olympic Games, we use the two closest Olympic Games in time (e.g. for the 1984 Los Angeles Olympics, since the U.S. did not participate in the 1980 Moscow Olympics, we use the data from 1976 and 1988 for the U.S.).

### : An increase in the number of medals of the host

We use to denote the rate of increase in the number of different types of medals of the host country (i = 1,2,3 respectively represents gold, silver, bronze medal, and the case without i means the total medal)

,,respectively represents the number of medals won in the previous, current, and next Olympics of the current host country respectively, i = 1,2,3 respectively represents gold, silver, bronze medal, and the case without i means the total medal. j means the year of Olympics. So we can conduct the rate of increase:

(19)

Next, referring to the prediction medals of the U.S. (the prediction results of old events), multiplying by the increase rate, adding the prediction results of new events, we finally get the prediction for the expected gold, silver, bronze and total medals in the 2028 Los Angeles Olympics considering the ‘host effect’, which are which are 59, 48, 53 and 160 respectively (after rounding off to the nearest integer)

Considering a reduction in the advantage associated with the host country's performance will correspondingly lead to an increase in the performance metrics of other participating countries. Considering the enhancement in competitive outcomes typically observed for the host nation due to host-related factors, the relative performance indicators of non-host nations may consequently decline. Specifically, in the upcoming international sporting event scheduled for 2028 in Los Angeles, the United States is expected to achieve a significant additional number of top-tier results, which will substantially impact the relative standing and success rates of other prominent participant countries. Therefore, it is necessary to introduce an adjustment parameter to recalibrate the performance indicators of other nations to account for this host-related effect.

We use to denote the rate of decrease in the number of different types of medals of the non-host country, and  denotes the number of medals of the U.S. we predicted before ( i = 1,2,3 respectively represents gold, silver, bronze medal, and the case without i means the total medal). From IOC, we have calculated the number of events and medals for the next Olympics, which is approximately 300. Thus, we can calculate in terms of :

(20)

,

,

For a specific country, the new prediction ( i = 1,2,3 respectively represents gold, silver, bronze medal, and the case without i means the total medal) considering host effect is:

(21)

After adding the new predictions of old events to the predictions of new events, we conclude our final prediction of the top 10 countries with the highest total number of medals won in old events concerning the host effect, which is shown in Table V.

Table V. The final prediction of top 10 countries with the highest total number medals won in the old events concerning the host effect

| Country | Medal | | | |
| --- | --- | --- | --- | --- |
| Gold | Silver | Bronze | Total |
| USA  CHN  GBR  RUS  FRA  JPN  AUS  GER  ITA  NED | 59  32  21  21  14  23  15  12  11  14 | 48  25  22  21  21  12  16.  16  12  9 | 53  26  26  15  21  19  16  14  13  13 | 160  83  69  57  56  54  47  42  36  36 |

## Prediction for Countries Winning Medals for the First Time

### Result of Arima

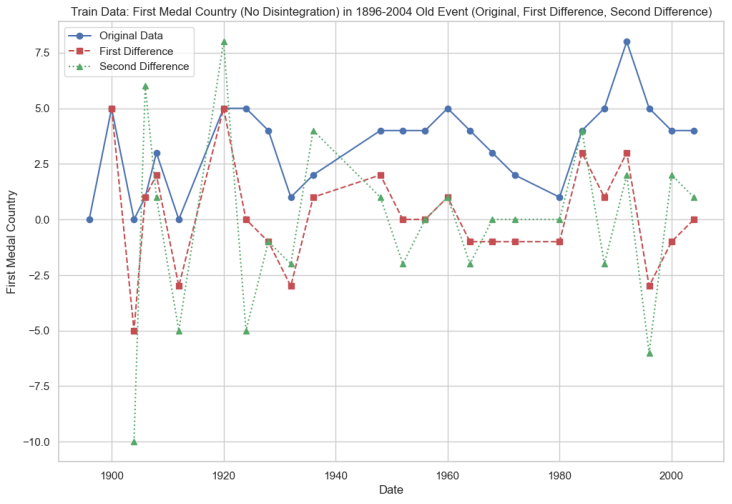


Figure 3. Trend chart of original data, first-order difference and second-order difference. The original data exhibits the best stationarity

Figure 3 demonstrates the trend chart of original data, first-order difference, and second-order difference. The original data exhibits the best stationarity.

图表, 瀑布图

AI 生成的内容可能不正确。

Figure 4. ADF Statistic and p-value. The ADF statistic is -4.01, and the p-value is 0.0014, verifying the stationarity of the original series

Figure 4 illustrates the ADF statistic and p-value for three series, with the original data showing the lowest ADF statistic (-4.01) and p-value (0.0014), confirming the stationarity of the original series.

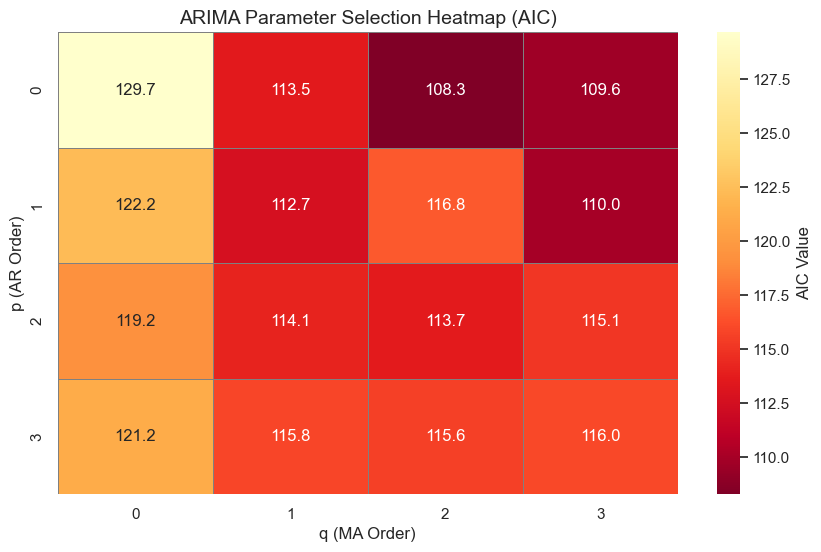


Figure 5. AIC heatmap. The figure shows that the AIC criterion selects the ARIMA(0,0,2) model, where p = 0 and q = 2 minimize the AIC value

The heatmap in Figure 5 shows that the AIC criterion selects the ARIMA(0,0,2) model, where p = 0 and q = 2 minimize the AIC value.

日历

AI 生成的内容可能不正确。

Figure 6. ACF and PACF of residuals. The residuals show no significant autocorrelation, indicating a good fit of the model

图表, 折线图

AI 生成的内容可能不正确。

Figure 7. Ljung-Box test results. The Ljung-Box test shows no significant autocorrelation in the residuals (p > 0.05)

图表, 折线图

AI 生成的内容可能不正确。

Figure 8. Shapiro-Wilk test results. The test indicates that the residuals are normally distributed (p > 0.05)

图表, 散点图

AI 生成的内容可能不正确。

Figure 9. QQ plot of residuals. The residuals align with the fitted line, supporting their normality

The ACF and PACF plots in Figure 5 and Figure 6, along with the Ljung-Box test in Figure 7, confirm no significant autocorrelation in the residuals (p > 0.05), indicating a good ARIMA model fit. The Shapiro-Wilk test in Figure 8 and QQ plot in Figure 9 further show the residuals are normally distributed (p > 0.05), with a bell-shaped histogram and alignment along the fitted line, supporting the model’s validity and normality assumptions.

图表, 折线图

AI 生成的内容可能不正确。

Figure 10. Comparison of actual vs predicted number of first-time medal-winning countries (2008-2028). The prediction aligns well with the actual results, indicating high accuracy of the model

Figure 10 indicates the comparison of the actual vs predicted number of first-time medal-winning countries (2008-2028). The prediction aligns well with the actual results, indicating high accuracy of the model.

### Result of random forest

图表, 折线图

AI 生成的内容可能不正确。

Figure 11. Training, testing, and 2028 prediction results for first-time medal-winning countries. The 95% confidence interval is [2.58, 4.87], indicating stable and reliable predictions

Figure 11 illustrates the training, testing, and 2028 prediction results for first-time medal-winning countries, showing a 95% confidence interval of [2.58, 4.87], which indicates stable and reliable predictions.

# Conclusions and Outlooks

## Conclusions

This study proposes a comprehensive predictive framework that integrates linear regression, ARIMA, random forest, and Monte Carlo simulation to address multi-objective forecasting tasks involving heterogeneous time-series data. The combination of statistical and machine learning methods enables the model to capture both temporal dependencies and nonlinear patterns. Confidence intervals are derived using Poisson and probabilistic distributions, enhancing the interpretability and robustness of the predictions.

A hybrid modeling approach is adopted, integrating ARIMA (p=0, d=0, q=2) and random forest models to account for temporal dependencies and nonlinear patterns, particularly for low-frequency or previously underrepresented entities. Furthermore, a combination of regression analysis and participation-based feature engineering enables more robust outcome estimation across multiple dimensions. To quantify predictive uncertainty, the framework applies Poisson-based confidence interval estimation, yielding interpretable bounds for all predicted outcomes. The results indicate a significant likelihood of performance increase for dominant entities under host-related conditions, while others may experience marginal declines due to redistribution effects.

However, the current approach remains dependent on high-quality historical data, which may limit its applicability in data-scarce scenarios. The simulation process is also sensitive to ranking-based assumptions, and the ensemble model could benefit from automated tuning and selection strategies.

Future work will focus on integrating transfer learning to improve prediction in low-resource categories, and exploring graph-based representations to capture inter-category and inter-entity dependencies more effectively. Furthermore, incorporating Bayesian techniques or ensemble uncertainty quantification could further improve confidence calibration under high variance conditions. These enhancements aim to broaden the applicability of the proposed framework across a wider range of sequential decision-making and resource allocation problems.

# References

[1] Fazlollahi P, Afarineshkhaki A, Nikbakhsh R. Predicting the medals of the countries participating in the Tokyo 2020 Olympic Games using the test of networks of multilayer perceptron[J]. Expert Systems with Applications, 2023, 220: 119320.

[2] Prince Nagpal, Kartikey Gupta, Yashaswa Verma, Jyoti Singh Kirar. Paris Olympic (2024) Medal Tally Prediction[C]. International Conference on Data Management, Analytics & Innovation, 2023, 249–267.

[3] Li S, Cai T T, Li H. Transfer learning for high-dimensional linear regression: Prediction, estimation and minimax optimality[J]. Journal of the Royal Statistical Society Series B: Statistical Methodology, 2022, 84(1): 149-173.

[4] Cetinkaya A, Peker S, Kuvvetli Ü. Analysis of countries' performances in individual Olympic Games using cluster analysis and decision trees: the case of Tokyo 2020[J]. Sport, Business and Management: An International Journal, 2024, 14(5 - 6): 648 - 666.

[5] Loland S. Classification in sport: A question of fairness[J]. European Journal of Sport Science, 2021, 21(11): 1477 - 1484.

[6] Zhaoyu Zhu, Sheng Jiajin, Yu Minghao, Liang Chengtian. Application of Data-Driven Prediction and Strategic Optimization in Olympic Medal Distribution[J]. International Journal of Advanced Science. 2025, 1(1):1-17.

[7] Justen O’Connor, Laura Alfrey, Dawn Penney. Rethinking the classification of games and sports in physical education: a response to changes in sport and participation. Physical Education and Sport ‘Pedagogy, 2021, 315-328.

[8] Zukerman I, Albrecht D W. Predictive statistical models for user modeling[J]. User Modeling and User-Adapted Interaction, 2001, 11: 5-18.

[9] Wang Yiquan, Wang Jiaying, Huang Tin-Yeh, Yang Jingyi, Xu Zihao . STGCN-LSTM for Olympic Medal Prediction: Dynamic Power Modeling and Causal Policy Optimization[J]. AeXiv:2501.17711, 2025.

[10] Christoph Schlembach, Sascha L. Schmidt, Dominik Schreyer, Linus Wunderlich. Forecasting the Olympic medal distribution–a socioeconomic machine learning model[J]. Technological Forecasting and Social Change, 2022, 175: 121-314.