Olympic Medal Prediction Based on Multi-Factor Analysis Model

LiJie Tao #, 1, *, JiKai Zhang #, 1, Shuang Feng #, 2, YiTong Luo 3, XinYang Shu 4

1 School of Science, Xi'an University of Posts and Telecommunications, Xi'an, China, 710121

2 School of Cyberspace Security, Xi'an University of Posts and Telecommunications, Xi'an, China, 710121

3 School of Communications and Information Engineering, Xi' an University of Posts and Telecommunications, Xi'an, China, 710054

4 School of Computer Science and Engineering, Xi'an University of Technology, Xi'an, China, 710121

* Corresponding Author Email: 18209142818@163.com

#These authors contributed equally.

Abstract. The Olympic Games, as a world-renowned sporting event have garnered considerable attention due to their immense influence. Predicting Olympic medals holds significant guidance for related industries and national development strategies. This paper first conducts correlation analysis using references and practical situations to identify factors influencing medal wins. It employs the entropy weight method to determine the weight of each factor and applies the TOPSIS method to weigh medal data from different years. Subsequently , multi-factor analysis and multiple regression equations are established to predict countries' medal-winning potential and probability scores. After ranking, initial models for medal rankings and estimated medal counts are established. Considering population, the number of participants from each country in various events, historical medal losses, and the probability of medal seizure, the binomial distribution is then used to estimate the probability of medal seizure and predict its variation range for each country, leading to the final medal rankings forecast. Combined with Pareto analysis, suggestions for potential advantageous events are provided. Mathematical models are used to predict countries' Olympic award situations as accurately as possible, offering effective guidance for their future competitions and related industries and promoting the rational allocation of resources and the correct implementation of strategies.

Keywords: Multifactor Analysis, Multiple Regression, Pareto Analysis Method.

1. Introduction

During the Paris Olympics, in addition to marveling at the wonderful moments of the sports competition, people also paid close attention to the medal table. In addition to the countries that are at the top of the medal table and the host countries that have received a lot of attention, the countries that won the first medal have also generated public interest. It is worth noting that at the moment, more than 60 countries have yet to win a medal. Analyzing various factors, such as the data of the participating athletes and the proximity of the host country to the Olympics, can lead to the expected medal results. These predictions enable spectators to watch events more strategically and enable athletes to better assess risk and plan their game strategies.

This paper establishes a medal prediction model by combining a medal ranking score model with an estimated medal count model using methods such as multi-factor analysis, Pareto analysis, and multiple regression. This combination allows us to obtain medal predictions for various countries, aiming to provide guidance for their future preparations for the Olympics and related industries.

Nicolas Scelles et al. predicted the total number of medals for 2016 and 2020 using the Tobit model and Hurdle model[1], with overall results aligning with 2020 despite some significant differences. Duncan Robert C.et al. introduced a new ranking mechanism by defining a ranking index U using a simple binomial sum[2]. Xu Yang et al. introduced a comprehensive indicator for ranking models[3], set weights, validated the obtained medal rankings, and presented the advantages and satisfaction of the new rankings. Domestic and foreign scholars have discussed various factors influencing the models, including participation experience, historical performance, national emphasis, GDP, number of Olympic participants, and home-court advantage[4-6]. They have also considered factors such as first-time medal-winning countries[7]and newly added events[8]for medal table predictions.

Traditional methods are generally characterized by a single research object, overly simplistic model construction, limited prediction scope, weak adaptability, or mere simple analysis. They summarize possible future development trends[9-11]without providing prediction results and quantitative analysis. Additionally, some constructed models have excessive data accuracy and quantity requirements that are difficult to collect. Therefore, this paper adopts multi-factor analysis, the analytic hierarchy process, the entropy weight method for Z weight setting, and Pareto analysis. It maintains comprehensive research factors and conducts prediction analysis of countries' medal counts from multiple levels and broad fields. While ensuring the rationality and accuracy of predictions, it simplifies computational complexity and reduces data requirements. At the same time, it attempts to answer the following questions: What are the predicted medal counts and ranking results for various countries? How do predictions for dark horse countries look? What is the specific impact of potential advantageous events on the future sports industry and Olympics of various countries? By addressing these questions and analyzing the results, hope to provide practical and effective guidance for the future sports development and resource allocation of various countries.

2. Medal Prediction Model

2.1. Medal Points ranking Model

In the medal points ranking model, to quantify competition experience, statistically analyzed the number of times each country has participated in various events over the past seven Olympic Games. The data are based on the medalists of the modern Summer Olympics from 1896 to 2016[12]and the official website of the International Olympic Committee. Some of the indicators are shown in Figure 1.

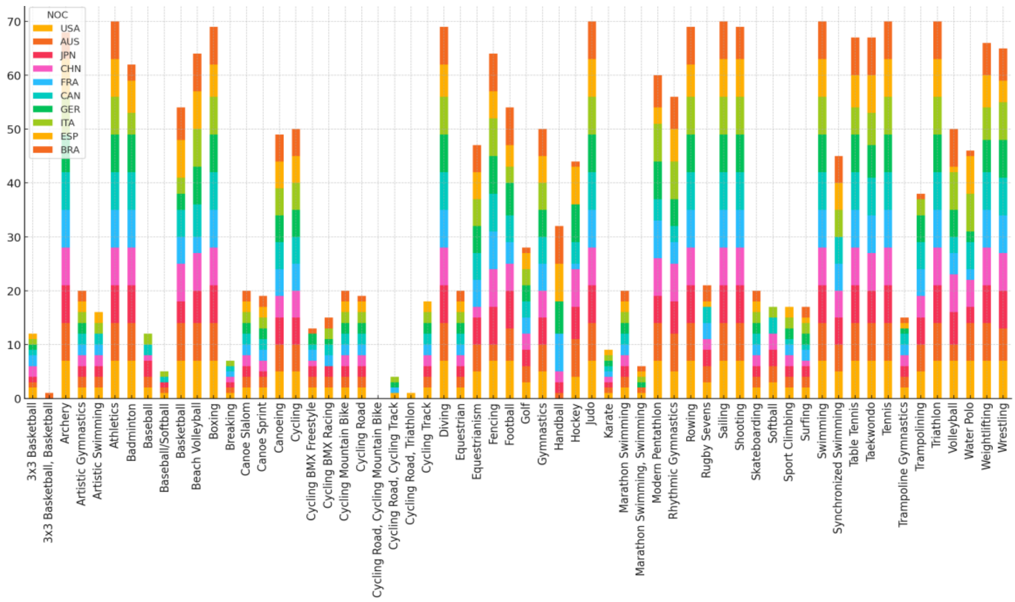

Figure 1: Stacked Bar Chart of Event Participation Frequency

For historical performance, directly using medal rates to calculate actual performance can lead to significant errors. Therefore, introduce a scoring system to provide a more accurate and reasonable evaluation of historical achievements, as follows.

$$S_{ij} = 3 \cdot gold_{ij} + 2 \cdot silver_{ij} + 1 \cdot bronze_{ij} \tag{1}$$

Where S_{ij} represents the total score of the i-th country in the j-th event, and $gold_{ij}$, $silver_{ij}$,and $bronze_{ij}$ represent the number of medals of all kinds won by the i-th country in the j-th event.

The extent of a country's emphasis on an event is reflected in pre-event investments in athletes and the construction of training facilities. Abstract the country's emphasis by using recent medal performance in specific events. This can be represented as follows.

$$\begin{cases} R_{ij} = \beta_1 \frac{M_{ij}}{M_{ti}} + \beta_2 \frac{M_{ijg}}{M_{tig}} \\ \beta_1 = \frac{3}{5} \\ \beta_2 = \frac{2}{5} \end{cases} \tag{2}$$

$$\begin{cases} bf = \frac{K_{ij}}{5} \cdot \theta \cdot \rho \\ \rho = 0 \quad , Rij \leqslant 0.2 \\ \rho = 1 \quad , Rij > 0.2 \end{cases} \tag{3}$$

As shown in Equations (2) and (3),obtain the national emphasis R_{ij} ,where β_1 is the coefficient indicating the number of medals won in the event, and β_2 is the coefficient indicating the number of gold medals won in the event. M_{ti} represents the total number of medals won by the i-th country in a particular Olympic Games, and M_{tig} represents the total number of gold medals won by the i-th country in that Olympic Games. The values of β_1 and β_2 are determined using the entropy weight method. bf is the enhancement ratio of emphasis on scores, and k_{ij} is the number of times the i-th country has won medals in the j-th event over the past five Olympic Games.

For the home advantage factor, introduce a variable H_i to account for the influence of the host country as follows.

$$H_i = \frac{a+b}{n-1} \tag{4}$$

where n is the sum of the edition in which the country was the host and the editions immediately before and after that, which are considered to have reference value. Typically, two editions before and two editions after the host edition are included. For the calculation factors a and b they satisfy the following formulas.

$$\begin{cases} a = \frac{M_{thi} - M_{thi-1}}{M_{thi-1}} + \frac{M_{thi} - M_{thi-2}}{M_{thi-2}} + \ldots + \frac{M_{thi} - M_{thi-\frac{n-1}{2}}}{M_{thi-\frac{n-1}{2}}} \\ b = \frac{M_{thi} - M_{thi+1}}{M_{thi+1}} + \frac{M_{thi} - M_{thi+2}}{M_{thi+2}} + \ldots + \frac{M_{thi} - M_{thi+\frac{n-1}{2}}}{M_{thi+\frac{n-1}{2}}} \end{cases} \tag{5}$$

where M_{thi} represents the number of medals won by the i-th country when it was the host nation. The coefficients for addition and subtraction correspond to the editions before and after the host edition.

The number of participants in each event, as determined by logistic analysis, shows that when the number of participants has not reached a limiting peak, there is a positive correlation between the number of participants and the number of medals won. This method provides higher accuracy compared to simply counting the number of events. A stacked bar chart showing the frequency of participation by athletes from selected countries in various events is shown in Figure 2.

In summary, the medal-single event scoring ranking model can be expressed as following formula.

$$\begin{cases} M_{ti} = \varphi_1 E_i + \varphi_2 \sum_k S_{ij} + \phi_3 r_{ki} + \varphi_4 GDP_i + \varphi_5 T_{ij} \\ N_{ti} = M_{ti} \cdot (1 + H_i) \\ O_{ti} = M_{ti} \cdot (1 + \frac{K_{ij}}{5} \cdot \theta) \\ H_{ti} = M_{ti} \cdot (1 + H_i) \cdot (1 + \frac{K_{ij}}{5} \cdot \theta) \qquad (\theta = 0.2) \end{cases} \tag{6}$$

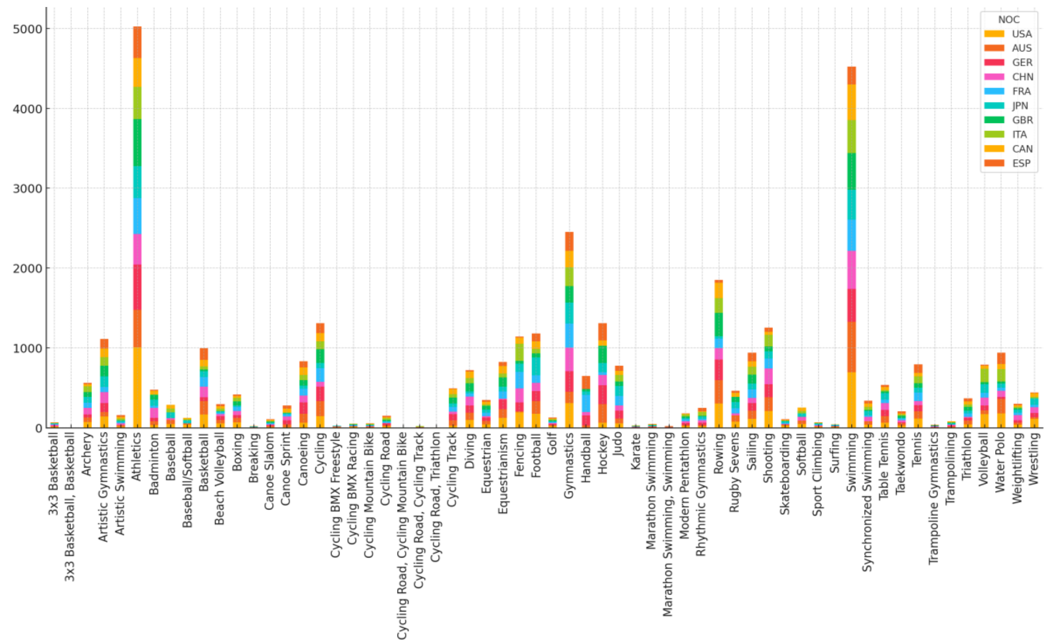

Figure 2: Participation Frequency Chart by Country and Event

where M_{ti} represents the performance score obtained by the i-th country when it is not the host nation and the emphasis on the event is considered average. φ is the weight coefficient, r_{ki} is the number of participants from the i-th country in the k-th Olympic Games, GDP represents the economic strength of the country, N_{ti} is the performance score when the country is the host nation, E_i is the participation experience of the i-th event in the 21st century, O_{ti} is the performance score when the country is not the host but places high importance on a particular event, R_{ij} is the emphasis level on the j-th event. When $R_{ij} \geq 0.2$, it indicates that the country places significant emphasis on the event. H_{ti} is the performance score when the country is both the host and places high importance on the event, T_{ij} is the medal count for the j-th event of the i-th country.The specific weights are shown in Table 1.

Based on the above, can derive a scoring model for medals in individual events. By summing up the total scores of all events for each country and then ranking them,can obtain a medal score ranking model.

$$\begin{cases} W_i = M_{ti} \cdot (1 + H_i \cdot P) \cdot (1 + \frac{K_{ij}}{5} \cdot \theta \cdot \rho) \\ W_z = \sum W_i \\ P、Q \in \{0,1\} \end{cases} \tag{7}$$

where W_i is the total score for the i-th country in a specific event.

2.2. Estimated medal count Model

When summing up the gold, silver, and bronze medals for each corresponding rank on the medal table of each edition, since the influencing factors decrease with the passage of time, the medal tables of different years are weighted using the TOPSIS method. The weight for the year 2000 is set to 1,for 2004 it is set to 2,and so on until 2024,which is set to 7. After weighting and summing the data, it is divided by the total weight value to normalize the weights, initially obtaining the theoretical number of medals and gold medals for the corresponding rank.

The preliminary prediction of the medal table is made according to the ranking based on scores. Next, the impact of being the host country on obtaining medals is added. According to the analysis of the host country factor in the aforementioned medal scoring ranking model, the number of medals and gold medals for the host country will be 30% higher than during non-host periods.

2.3. First Medal Win Prediction Model for Countries That Have Never Won Medals

For countries that have not yet won any medals, the criteria for predicting their first medal win are still relatively rough. Therefore, statistically analyze the countries that won their first Olympic medals in the 2028 Summer Olympics and the four preceding editions. Meet by year:

$$\begin{cases} \omega_1 < \omega_2 < \omega_3 < \omega_4 < \omega_5 \\ \sum_{i=1}^{5} \omega_i = 1 \end{cases} \tag{8}$$

By calculating the weighted number of participants N_i for each country winning its first medal in each edition, determine the theoretical minimum number of participants N expected to win their first medal in 2028, which satisfies the following formula.

$$\begin{cases} N_i = \sum_{j \in Si} \omega_i n_{ij} \\ N = \dfrac{\sum_{i=1}^{4} N_i}{\sum_{i=1}^{4} | S_i |} \end{cases} \tag{9}$$

where S_i is the set of countries winning their first medal in the i-th edition, and n_{ij} is the number of participants from country $j \in S_i$.

Considering that the relationship between the number of participants and the number of medal winners is not a simple linear relationship, a logistic regression model is used for prediction.

$$P(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}} \tag{10}$$

where x is the number of participants from a country in a given edition, $P(x)$ is the probability of winning a medal, and β is calculated using the maximum likelihood estimation method based on the performance and number of participants from the past four editions.

Assuming the number of participating countries in 2028 is M ,and the number of participants from each country is x_1 to x_M , then, the number of countries that won the first medal is as follows.

$$\hat{Y} = \sum_{m=1}^{M} p(x_m) \tag{11}$$

Logical analysis and historical data indicate that while the total number of participating countries has been increasing since 2000, it has recently stabilized. The number of countries that have never won a medal is decreasing overall, and the number of countries winning their first medal will also show a decreasing trend. Therefore, an adjustment factor r_k related to the edition of the games is introduced.

$$P(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x + \gamma r_k)}} \tag{12}$$

2.4. Specific Event Selection and Its Impact on Countries

For different countries, their strengths and focuses in various events differ. Therefore, based on the country's emphasis and historical performance, discuss the impact on events again. Taking China's table tennis event as an example, using equation (2), the corresponding R_{ij} from 2000 to 2024 are 0.1399, 0.0943, 0.0790, 0.0995, 0.1265, 0.0949 and 0.1013, with an average of 0.1051. Compared to many other sports, the emphasis on table tennis exceeds 10%,and the medal-winning rate and gold medal-winning rate for table tennis over the years have been significantly higher than other sports. Since a country may focus on multiple events, an emphasis above 10% is considered significant. Similarly, determine the events that each country focuses on: China: Diving, Table Tennis; USA: Basketball, Swimming; Brazil: Football, Volleyball; Japan: Table Tennis, Football.

3. Results

3.1. Weight factor

Since the weight coefficients for each event are inconsistent, only a portion of the event weights are shown. The weight of the partial events is shown in Table 1.

Table 1: Event Weight Table

sports	aspect1	aspect2	aspect3	aspect4	aspect5
3x3 Basketball	0.1482	0.1829	0.2042	0.2561	0.2086
Archery	0.1855	0.2288	0.1048	0.3167	0.1642
Artistic Gymnastics	0.1828	0.2255	0.1174	0.2931	0.1813
Artistic Swimming	0.1452	0.1792	0.1799	0.2935	0.2022
Athletics	0.2587	0.3191	0.0045	0.2911	0.1266

3.2. Results of the Medal Prediction Model

By combining the medal points ranking model with the estimated number of medals model, arrange the gold medal leaderboard. Additionally, considering the number of athletes each country has in various events and the number of events they participate in, perform a secondary ranking of the total number of medals. The resulting table with the calculated points, total medals and gold medals .

Since the actual situation may differ from the medal and gold medal numbers given in the proposed table, it is necessary to discuss the magnitude of this deviation to ensure the rigor and accuracy of the results. From the integrated data provided, know that in recent years, approximately three countries that have never won medals manage to win medals each year. Using this information, can estimate the probability of countries' medals being "competed for" to predict the range of changes.

For example, in 2024, the total number of medals is 1039, with the top 12 countries having the following medal counts:126,91,45,53,64,34,65,32, 40, 33, 20, and 27 medals respectively, accounting for 60.6% of the total medals. Therefore, discussing the probability of these countries' medals being "competed for" by previously non-medaling countries can help determine the range.

The weights x_i for the top 12 countries are as follows.

$$x_i = \frac{M_i}{\sum_{i=0}^{206} M_i} \tag{13}$$

where M_i is the number of medals for the country ranked i, each country has the possibility of having 0, 1, 2, or 3 medals taken by countries that have never won medals before. Since all countries are potentially subject to having their medals "competed for," discard the possibility of 0 (even though the probability for some countries is extremely low). For the competition probability, use the binomial distribution, which satisfies the following formula.

$$P_y = 3 \cdot x_i^{k} \cdot x_i^{3-k} \tag{14}$$

where P_y is the probability of having y medals taken. From this, can obtain the following table.

From Table 2,it is known that there is at least a 2.92% chance that the top two countries on the medal table will have two medals contested. For the other countries, there is at least a 98.77% probability that the number of medals contested will be less than 2. Therefore, can predict the medal and gold medal table intervals for the 2028 Summer Olympics in Los Angeles, USA is shown in Table 2.

Table 2: Probability of Medal Loss Table

number	total	Xi	P1	P2	P3	number	total	Xi	P1	P2	P3
1	126	0.1213	0.2166	0.0389	0.0018	7	65	0.0626	0.156	0.0123	0.0002
2	91	0.0876	0.2016	0.0292	0.0007	8	32	0.0308	0.0899	0.0031	0
3	45	0.0433	0.1168	0.0064	0.0001	9	40	0.0385	0.1078	0.0045	0
4	53	0.051	0.1334	0.0087	0.0001	10	33	0.0318	0.0918	0.0033	0
5	64	0.0616	0.1538	0.0121	0.0002	11	20	0.0192	0.0555	0.0007	0

Based on the predicted data in Table for the top 5 countries on the medal table, analyzed the changes in ranking and medal counts between the 2024 and 2028 Olympic medal tables. Our analysis revealed that although the United States maintained its ranking, its medal count would increase. China, on the other hand, retained its ranking but would see a decrease in its medal count. Australia experienced an increase in both its ranking and medal count. The predicted data for the top 5 countries on the medal table is shown in Table 3.

Table 3: Predicted Medal Table

Number	country	gold	total	Number	country	gold	total
1	United States	[44,46]	[128,130]	6	Germany	[13,14]	[29,30]
2	China	[34,36]	[95,97]	7	United Kingdom	[11,12]	[58,59]
3	Australia	[25,26]	[60,61]	8	Italy	[10,11]	[34,35]
4	Japan	[18,19]	[48,49]	9	Canada	[9,10]	[32,33]
5	France	[15,16]	[52,53]	10	Spain	[9,10]	[16,17]

3.3. Prediction Results of the Probability of a Country Winning a Medal for the First Time

Logical analysis and historical data indicate that while the total number of participating countries has been increasing since 2000,it has recently stabilized. The number of countries that have never won a medal is decreasing overall, and the number of countries winning their first medal will also show a decreasing trend. Therefore, an adjustment factor r_k related to the edition of the games is introduced:

$$P(x) = \frac{1}{1+e^{-(\beta_0+\beta_1 x+\gamma r_k)}} \tag{15}$$

where k indicates the number of editions counted backward from the current edition, and $r_1 > r_2 > r_3 > r_4$. γ is a constant.

Substitute the number of participants from each country into equation (12).If $P(x) > 0.8$,it is considered that the country can win a medal. Summing up the number of countries where $P(x) > 0.8$ gives the predicted number of countries winning their first medal in the next edition. Some examples of the predicted probabilities for countries winning their medal is shown in Table 4.

Table 4: Predicted Probability Table for First Medal Wins of Selected Countries

Number	NOC	Probability	Number	NOC	Probability	Number	NOC	Probability
1	ANG	69.80%	12	RWA	44.40%	23	BEN	42.11%
2	HON	63.44%	13	BOL	43.85%	24	ARU	41.50%
3	SAM	56.10%	14	CAY	43.85%	25	COD	40.87%
4	MLI	50.52%	15	MAD	43.85%	26	LBA	40.87%
5	GUI	50.10%	16	MLT	43.85%	27	SKN	40.87%
6	ESA	49.24%	17	NCA	43.85%	28	LBR	40.22%

To ensure its accuracy, the actual data is compared with the theoretical data for the years 16, 20and 24. It was found that the accuracy rate is between [35%, 43%].

3.4. Analysis of Event Influence and Correlation

Regarding the number of events (the number of major events and sub-events within a particular event) and the number of athletes participating in each edition for seven editions, the correlation heatmap for the gold medal-winning rate and medal-winning rate of the top seven countries in each edition is shown in Figure 3.

	year	Number of large items	Number of athletes	Number of small items	1st Number of Gold Medals	Second number of gold medals	Third, the number of gold medals	Fourth number of gold medals	Fifth gold medal number	Sixth gold medal number	Seventh number of gold medals	Number of medals in the first place	Second medal count	Third medal count	Fourth medal count	Fifth medal count	Sixth medal count	Seventh medal count
year	1	0.75051993	0.65754878	0.65354948	0.22864154	0.4556029	-0.5613414	0.6488857	0.57616596	0.3904344	-0.096225	0.8983223	0.2866802	-0.514388	0.247582	0.473898	-0.3222011	0.5028654
Number of large items	0.75051993	1	0.96923573	0.80029259	-0.3203201	0.5661214	-0.5361973	0.5831204	0.67265961	0.8492694	0.10642773	0.6245302	0.2033164	-0.5532876	0.1327675	0.8886045	-0.2997982	0.6509879
Number of athletes	0.65754878	0.96923573	1	0.77551454	-0.3514576	0.4134962	-0.5554625	0.5254063	0.6716144	0.8584013	0.16416888	0.5881672	0.2289086	-0.5817895	0.0365482	0.9565472	-0.3774187	0.5786719
Number of small items	0.65354948	0.80029259	0.77551454	1	0.13832815	0.609382	-0.5698309	0.7663712	0.71604833	0.8077896	0.15272758	0.4325355	0.4902	-0.5787807	0.0411873	0.6655798	-0.002865	0.4253209
1st Number of Gold Medals	0.22864154	-0.3203201	-0.3514576	0.13832815	1	0.0154326	-0.0917868	0.2818882	-0.0741622	-0.447538	-0.0586694	0.1091521	0.5826845	-0.2399092	0.1681072	-0.434654	0.07390243	-0.276291
Second number of gold medals	0.45560293	0.56612136	0.41349618	0.60938203	0.01543256	1	-0.3530539	0.2846848	0.05617346	0.43219	0.39618743	0.1078416	0.4617407	-0.5712612	0.3030764	0.3525401	-0.1097732	0.7401619
Third, the number of gold medals	-0.5613414	-0.5361973	-0.5554625	-0.5698309	-0.0917868	-0.353054	1	-0.074505	-0.2548203	-0.398485	-0.6285394	-0.640483	-0.235637	0.51146429	0.5467235	-0.390577	0.53085001	-0.716196
Fourth number of gold medals	0.64888568	0.58312038	0.5254063	0.76637115	0.28188822	0.2846848	-0.0745049	1	0.85158331	0.5700314	-0.4995124	0.4309359	0.2935892	-0.2288279	0.3961543	0.4157939	0.34347492	-0.088991
Fifth gold medal number	0.57616596	0.67265961	0.6716144	0.71604833	-0.0741622	0.0561735	-0.2548203	0.8515833	1	0.759848	-0.4435328	0.5537313	-0.08574	-0.0142585	-0.020172	0.5330391	0.29466875	-0.027803
Sixth gold medal number	0.3904344	0.84926944	0.85840134	0.80778962	-0.4475379	0.43219	-0.3984854	0.5700314	0.75984805	1	0.07513914	0.2845971	0.0324682	-0.2666755	-0.110725	0.8045651	0.08146944	0.4069508
Seventh number of gold medals	-0.096225	0.10642773	0.16416888	0.15272758	-0.0586694	0.3961874	-0.6285394	-0.499512	-0.4435328	0.0751391	1	-0.09879	0.3937823	-0.5500802	-0.439654	0.2020167	-0.6023599	0.670837
Number of medals in the first place	0.89832225	0.62453022	0.58816722	0.43253549	0.10915212	0.1078416	-0.6404826	0.4309359	0.55373134	0.2845971	-0.0987898	1	0.0009549	-0.3029761	-0.064701	0.3868693	-0.4186057	0.4197617
Second medal count	0.28668022	0.20331639	0.22890865	0.49020001	0.58268453	0.4617407	-0.2356369	0.2935892	-0.0857399	0.0324682	0.39378227	0.0009549	1	-0.8649911	0.3971946	0.2961805	-0.359939	0.2096895
Third medal count	-0.514388	-0.5532876	-0.5817895	-0.5787807	-0.2399092	-0.571261	0.51146429	-0.228828	-0.0142585	-0.266675	-0.5500802	-0.302976	-0.864991	1	-0.290752	-0.623457	0.68344907	-0.592245
Fourth medal count	0.24758203	0.13276751	0.03654824	0.04118726	0.16810722	0.3030764	0.54672353	0.3961543	-0.0201725	-0.110725	-0.4396536	-0.064701	0.3971946	-0.2907521	1	0.1348763	0.04351157	-0.117041
Fifth medal count	0.47389804	0.8886045	0.95654716	0.66557976	-0.4346539	0.3525401	-0.3905767	0.4157939	0.53303914	0.8045651	0.20201671	0.3868693	0.2961805	-0.6234567	0.1348763	1	-0.4250002	0.500724
Sixth medal count	-0.3222011	-0.2997982	-0.3774187	-0.002865	0.07390243	-0.109773	0.53085001	0.3434749	0.29466875	0.0814694	-0.6023599	-0.418606	-0.359939	0.68344907	0.0435116	-0.425	1	-0.60566
Seventh medal count	0.50286535	0.65098792	0.57867192	0.42532094	-0.2762906	0.7401619	-0.7161958	-0.088991	-0.0278027	0.4069508	0.67083698	0.4197617	0.2096895	-0.5922453	-0.117041	0.500724	-0.6056603	1

Figure 3: Heatmap of Correlation between Gold Medal Winning Rate and Medal Winning Rate

From Figure 3, it can be seen that there is a strong correlation between the number of major and minor events and the number of medals won by countries, generally above 0.85. The relationship of medal counts between countries is negatively correlated, which also reflects the competitive pressure among countries. The more major and minor events a country focuses on in the Olympics, the higher the number of medals they win, and the better their ranking.

3.5. Analysis of Model Accuracy, Stability, and Robustness

To test the accuracy of the model, the actual number of medals was compared to the theoretical projected medal count of the top 10 countries in the medal table for nine consecutive events from 2008 to 2024. The annual mean error is then calculated by normalization, and the accuracy is analyzed with a line chart visualization over time, as shown in Figure 4.

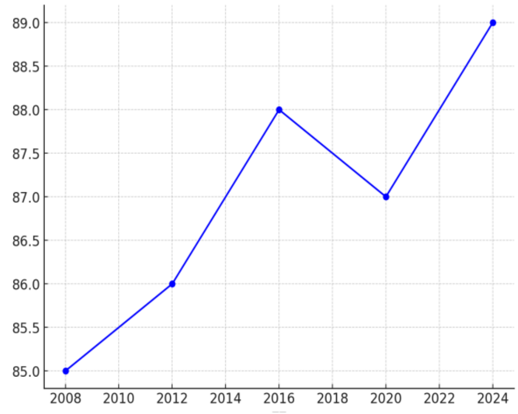

Figure 4: Accuracy Change Graph

From Figure 4, we can observe that, although the model accuracy occasionally decreases as the years progress from 2008 to 2024, there is an overall positive upward trend. Furthermore, the model accuracy has fluctuated within a 4% range over the nine events, indicating good stability with minimal variation.

Given that the number of participating countries and events changes with each edition, these variations are also discussed. The number of participating countries and events changes is shown in Figure 5.

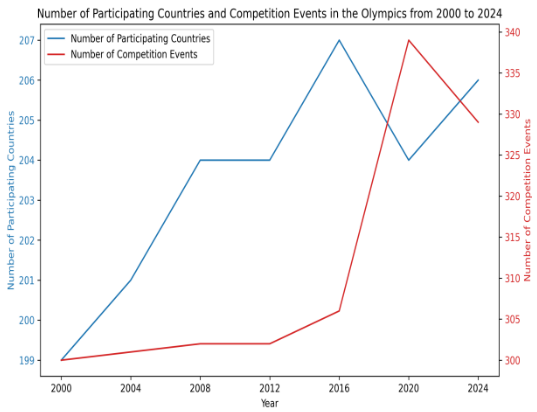

Figure 5: Changes in participating countries and number of projects

As shown in Figure 5, it is not difficult to see that both the number of participating countries and events are increasing. However, the range of fluctuation for participating countries is very small, with an extreme difference of only 7. The number of events was relatively stable from 2000 to 2016 but showed larger fluctuations between 2020 and 2024, with a difference of 10.These variations will have some impact on model accuracy.

However, the ratio of the medal base number to the change number in the early editions is quite small. Additionally, the medal prediction model is determined based on the medal points ranking from previous editions. Therefore, this disturbance factor will only introduce greater uncertainty for the rankings of the last few countries. Consequently, the model still exhibits good robustness.

4. Conclusions and Outlooks

Through analysis and modeling, this paper proposes a more refined prediction method for the Olympic medal table, and combines relevant historical data to obtain the prediction of the medal table in 2028 and the probability of a small country winning a medal for the first time. In this paper, the medal points ranking model and the medal ranking model based on multivariate analysis regression are fused with the overall event and participants, and the binomial distribution is used to calculate the medal winning probability, and the error interval is accurately predicted, and then the accuracy is verifiedIt has provided assistance for the future development and change of countries. In addition, through Pareto analysis, event changes and first-time medal country analysis, it can help different countries discuss potential changes and possible performances in various events, provide guidance for future sports development of the country, and formulate more effective Olympic strategies. At the same time, the Organising Committee of the Olympic Games can pay more attention to and increase support for countries that may win medals for the first time, pay better attention to key events in the competition, and better arrange activities and rationally allocate resources based on the number of participants in each event and public interests.

In this paper: the drawback of the model presented in this paper is that it only identifies numerous potential advantageous events without considering the return on investment for countries. Therefore, consider incorporating a coaching factor to determine which potentially favorable events are most beneficial as key investment projects. In addition, unexpected situations at the Olympics are often unpredictable, so subsequent improvements can include a stability factor for the athlete and focus the probability of winning a medal on the athlete himself, making the prediction more timely and accurate.

References

[1] Scelles N, Andreff W, Bonnal L, et al. Forecasting National Medal Totals at the Summer Olympic Games Reconsidered[J]. Social Science Quarterly, 2020, 101(2): 697-711.

[2] C. R D, Andrew P. Population-adjusted national rankings in the Olympics[J]. Journal of Sports Analytics, 2024, 10(1): 87-104.

[3] Yang X, Yelin F, Keung K L. a Composite Indicator Methodology to Rank National Olympic Committees[J]. Asia-Pacific Journal of Operational Research, 2022, 39(06):

[4] Snowberger D A, Lee H C. An Investigation into the Correlation between a Country’s total Olympic Medal Count, GDP, and Freedom Index Through History[J]. Proceedings of the Integrated Academic Conference of the Korea Information and Communication Society, 2021, 25(2):

[5] Shi Huimin, Zhang Dongying, Zhang Yonghui. Can Olympic medals be predicted? ——Based on the perspective of interpretable machine learning[J]. Journal of Shanghai Sports University, 2024, 48(04): 26-36.

[6] Huang Caihong, Yang Yiwu, Lu Wangang, et al. Distribution of gold medals, influencing factors and strategies for improving competitive sports strength at the Chinese sports delegation at the Summer Olympics[J]. Journal of Capital Sports College, 2022, 34(04): 438-446.

[7] CsurillaG, FertőI. How to win the first Olympic medal? And the second?[J]. Social Science Quarterly, 2024, 105(5): 1544-1564.

[8] Jiao Fangqian, Song Zhigang. Research on my country's strategic response under the evolution trend of new events in the modern Summer Olympics[J]. Journal of Beijing Sports University, 2023, 46(04): 118-127.

[9] Zhang Haiping. Analysis of the evolution of project group characteristics of my country's advantageous events in the Summer Olympics[J]. Sports Science and Technology Literature Bulletin, 2024, 32(05): 1-3.

[10] Yang Binglong. Analysis of the characteristics of medals won by China in the Olympic Games [J]. Journal of Shanxi University (Philosophy and Social Sciences Edition), 2010, 33(05): 117-120.

[11] Guo Aimin, Zhao Mingfa. Predicting the order of the gold medal list of the 2016 Summer Olympics based on gray theory[J]. China Science and Technology Information, 2013, (09):173-174.

[12] Vladimír B. Olympic medalists of the modern summer Olympic games 1896–2016[J]. Journal of Maps, 2021, 17(1): 145-153.