Problem Chosen	2025	Team Control Number
C	MCM/ICM Summary Sheet	2506135

Together, Individuals Make a Difference

Summary

The Olympic Games are not only a stage for athletes to showcase their performance but also a focal point for the competition on the medal table among nations. This paper aims to predict the distribution of medals at the 2028 Olympic Games through a bottom-up approach, estimating each nation's total medal count based on individual athletes' winning expectations. To achieve this, multiple predictive models were developed to analyze historical Olympic medal data, forecast future trends, and uncover deep data features, providing decision-making references for National Olympic Committees (NOCs).

In Task 1, we first conducted data preprocessing and feature engineering, extracting deep data features using the KMeans++ clustering algorithm and other statistical methods. Subsequently, the LightGBM gradient boosting regression algorithm was employed to predict the number of gold, silver, and bronze medals for each country in the 2028 Olympics. The results demonstrate that medal outcomes are closely linked to factors such as home-field advantage, advantageous events, and athletes' individual performance levels. The models performed well, with the R^2 value of the gold medal prediction model reaching 0.890, indicating high reliability. For countries that have never won a medal, our analysis revealed that nations participating in events with growth potential are more likely to secure their first medal. Spearman correlation analysis further explored the dependency of specific countries on advantageous events. The findings indicate that countries with stronger overall athletic capabilities exhibit less reliance on advantageous events and perform more consistently, while countries with weaker athletic capabilities show a stronger dependence on specific events. For example, some countries have a correlation coefficient as high as 0.894 in athletics, suggesting their sports development is concentrated in a limited number of disciplines.

In Task 2, to investigate the "great coach effect," we applied the Bayesian Changepoint Detection (BEAST) model to analyze case studies of notable coaches across various countries. The analysis revealed that performance shifts in certain countries coincide significantly with the tenure of great coaches, underscoring the critical role of coaching. Subsequently, the forward CUSUM algorithm was used to analyze medal variation sequences, identifying the weakest events for each country. The results indicate that while many countries had competitive advantages in specific events during earlier Olympic Games, these strengths have diminished over time. Based on this, targeted recommendations for coach assignments were proposed.

In Task 3, we conducted in-depth data mining and analysis based on the aforementioned models and data, uncovering several novel insights, including gender ratios among athletes, the dominance of certain nations in specific Olympic events, and how changes in medal distributions reflect shifts in global political landscapes. The study found that the gender ratio of Olympic athletes is gradually reaching parity, with the proportion of women's medals showing significant growth. Additionally, some nations maintain dominance in traditional events, with medal shares as high as 88

Keywords: **A, B, C,**

Contents

1 Introduction

1.1 Problem Background

During the 2024 Paris Summer Olympics, fans not only paid attention to the individual events but also showed significant interest in the the ranking of countries on the overall medal count and gold medal table. Ultimately, the United States topped the medal table with a total of 126 medals, while China and the United States shared the top position on the gold medal table, each securing 40 golds. The host country, France, ranked fifth on the gold medal table with 16 gold medals, but finished fourth in the total medal count. Great Britain ranked seventh on the gold medal table with 14 golds, while securing third place in the overall medal count. Despite the prominence of the leading countries, the medal achievements of other nations also attracted attention. For example, Albania, Cape Verde, Dominica, and Saint Lucia each won their first-ever Olympic medal in this edition of the Games, with Dominica and Saint Lucia each earning another gold medal. However, over 60 countries have yet to win a medal in the history of the Olympics.

1.2 Restatement of Problem

While predictions of the final medal counts at the Olympics are common, such forecasts are generally not based on historical medal tables. Instead, they are typically made before the start of the upcoming Olympic Games, once the list of competing athletes is known.To clarify the task, the problem is restated as follows:

- Develop a model for each country's medal count, which should at least include the number of gold medals and the total medal count, and estimate the uncertainty/precision of the model,also evaluate its performance.

  1) Using the model, predict the medal standings for the 2028 Los Angeles Summer Olympics, including prediction intervals for all outcomes. Based on the model's predictions, identify which countries are most likely to improve their performance and which countries are expected to perform worse than in the 2024.

  2) The developed model should include countries that have not yet won any medals, while also predicting the number of countries likely to win their first medal in the upcoming Olympics and provide odds for this estimate.

  3) The model should also consider the number and types of events in each specific Olympic Games, exploring the relationship between the events and the number of medals won by each country. For each country, determine which events are most crucial and why, as well as the impact of the country's selected events on the results.

- In contrast to athletes, who must change their citizenship in order to represent different countries in competitions, coaches do not need to change their citizenship to coach, which makes it easier for them to move to other countries. As a result, the "Great Coach" effect may arise. For example, Lang Ping has coached the volleyball teams of both China and the U.S., leading both to championships, while Bela Karolyi coached the Romanian and U.S. women's gymnastics teams, achieving great success. The task is to identify evidence in the data that may suggest changes driven by the "Great Coach" effect and estimate its impact on the medal count. Select three countries, determine the events in which they should consider investing in "outstanding" coaches, and estimate the potential impact of such investments.

- Explain the unique insights regarding Olympic medal counts that are included in the developed model and how these insights can provide valuable information to the National Olympic Committees of various countries.

1.3 Our work

We do such things ...

1. We do ...

2. We do ...

3. We do ...

2 Question Preparation

2.1 Assumptions

2.2 Notations

The primary notations used in this paper are listed in Table 1.

3 Data Preprocessing

3.1 Basic Data Preprocessing

$$G_i \cap G_j = \emptyset, \bigcup_{i=1}^{5} G_i = X \tag{1}$$

Table 1: Notations

Symbol	Definition
C_i	the first one
S_i	the second one
G	the second one
V	the second one
B	the second one
M	the second one
A	the second one
D	the second one
E	the second one
γ	Sport Advantagen Coefficient
m_{c_i}	the second one
$m_{c_i s_j}$	the second one
g_i	the second one
ξ	The Probability of First Medal

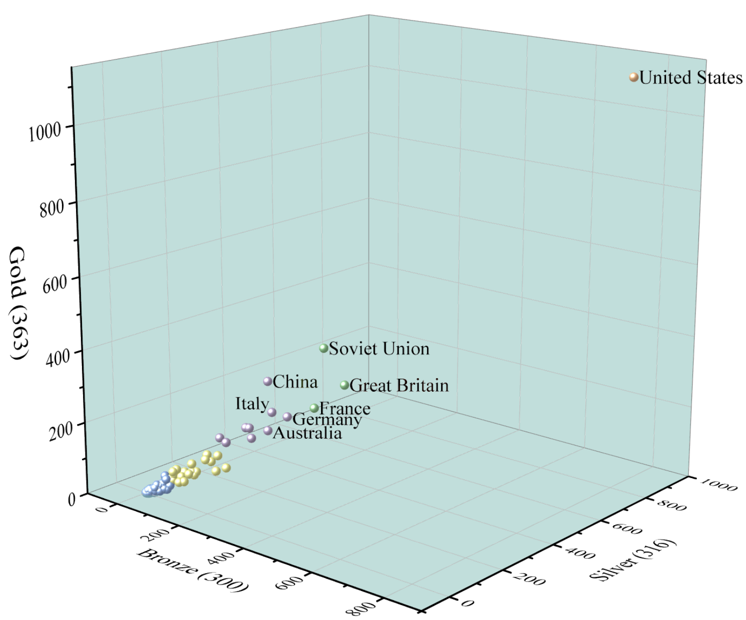


Figure 1: Scatter plot of national level classification (based on Kmeans++clustering algorithm)

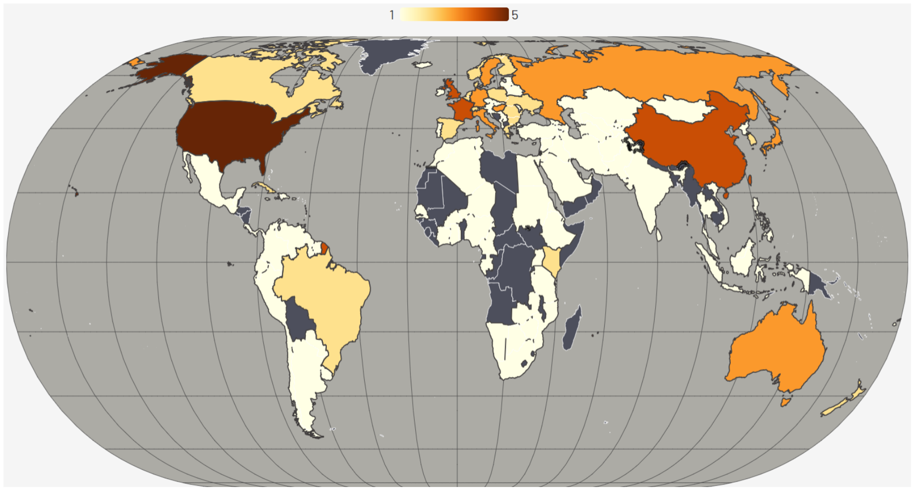


Figure 2: aa

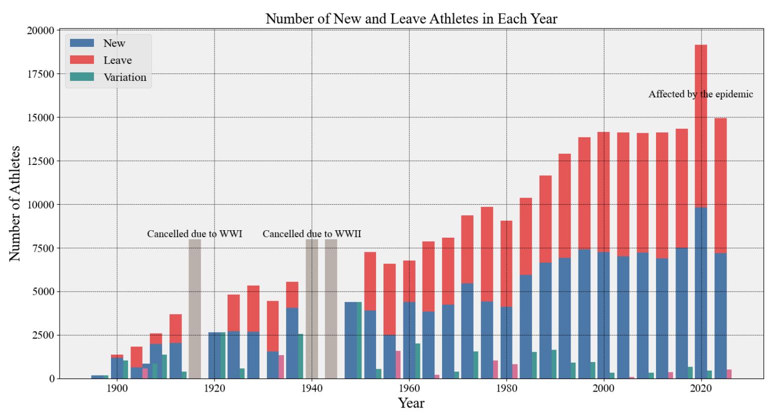
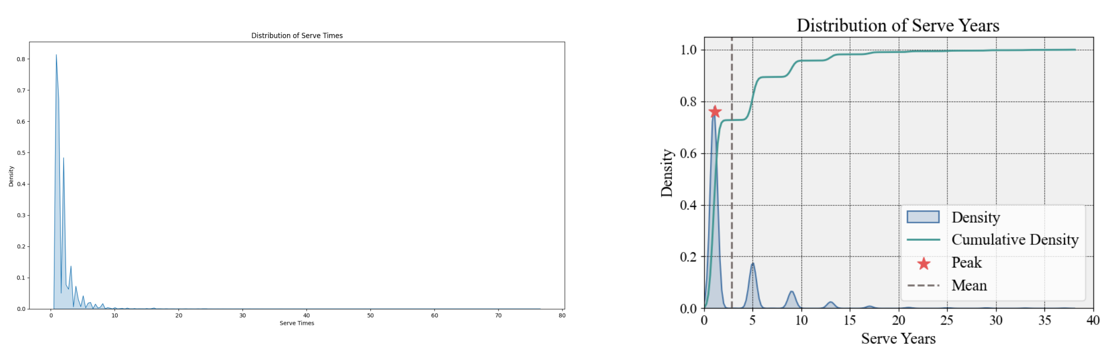


Figure 3: Number of New and Leave Athletes in Each Year



(a) Distribution of Serve Times

(b) Distribution of Serve Years

Figure 4: Two images

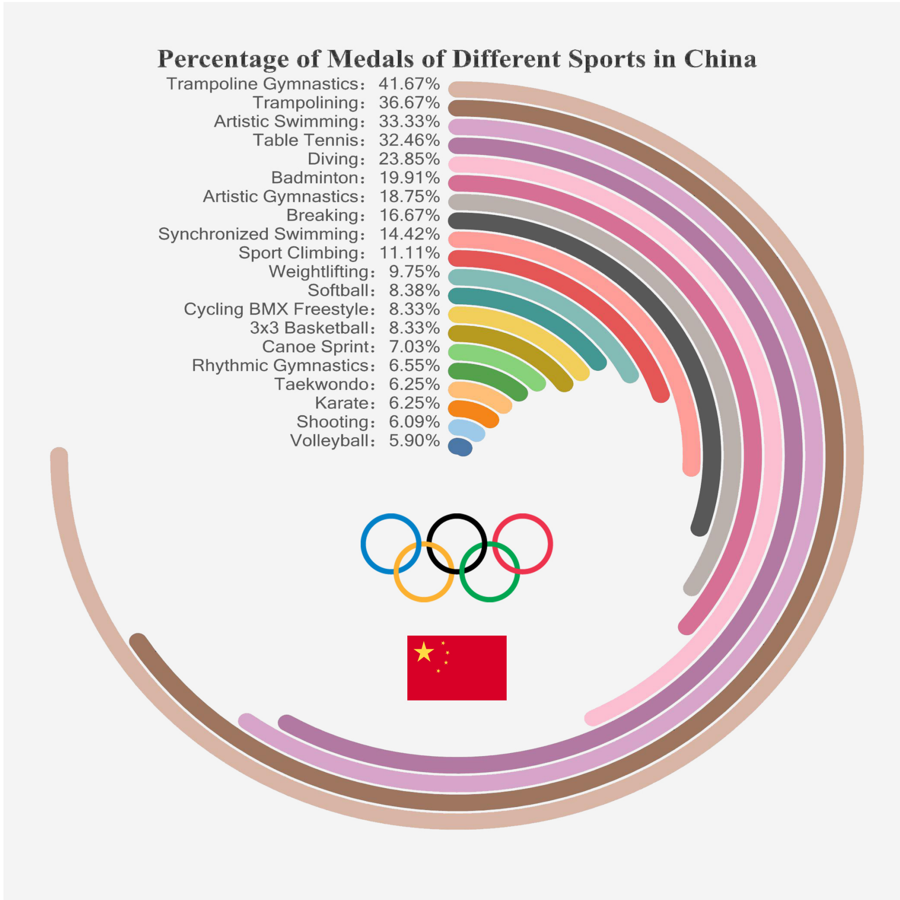


Figure 5: Percentage of Medals of Different Sports in China

3.2 Data Mining

3.2.1 Athlete Service Status

3.2.2 Distribution of Countries' Strength Sports

4 Task1:Medal Prediction Model Based on LightGBM

4.1 Medal Standings

The detail can be described by equation (**??**):**??**:

$$\gamma = \frac{m_{c_i s_j}}{\sum m_{c_i s_j}} \tag{2}$$

Table 2: Variable Name

Variable Name	*Code*	Definition

Continued on next page

Table 2: Variable Name (Continued)

Whether Host Country	*is host*	Whether the country is the host(1 for host,0 for non-host)
Medal Expectation Increment *Personnel Expectation lncrement	*medal_increment * personnel_increment*	Product of medal expectation increment and personnel expectation increment
Sport Advantage Coefficient	*sport_adv*	Advantage coefficient of a specific sport
Country Level	*country_lvl*	The level of the country in the competition (ordered by rank)
Project Medal Expectation /Project Personnel Expection	*sport_medal _per_ person*	Ratio of sport medals to projected personnel for a specific sport
Gold Medal Probability	*gold_prob*	Probability of an athlete winning a gold medal
Silver Medal Probability	*silver_prob*	Probability of an athlete winning a silver medal
Bronze Medal Probability	*silver_prob*	Probability of an athletewinning a bronze medal
No Medal Probability	*no_medal_probe*	Probability of an athlete winning no medal

Table 3: Countries' Medal Count Prediction(part)

Country	G_ pes.	S_ pes.	B_ pes.	G_ opt.	S_ opt.	B_ opt.	Gold	Silver	Bronze	Total
United States	47	36	23	53	41	25	51	40	25	117
China	35	35	15	46	38	20	40	36	17	95
Australia	28	20	22	33	24	24	29	22	23	71
France	23	19	19	27	20	22	24	19	21	64
Germany	18	24	14	23	26	16	21	25	15	61
United Kingdom	20	19	14	26	24	15	25	21	15	61
Japan	16	25	14	19	27	16	17	26	16	58
Italy	15	20	15	17	20	22	16	20	16	53
Spain	21	10	13	30	12	14	25	12	13	50
Netherlands	24	6	13	29	9	15	26	8	15	47
New Zealand	20	13	9	22	15	10	22	14	9	45
...	..	..	..	..	..	..	..	..	..	..

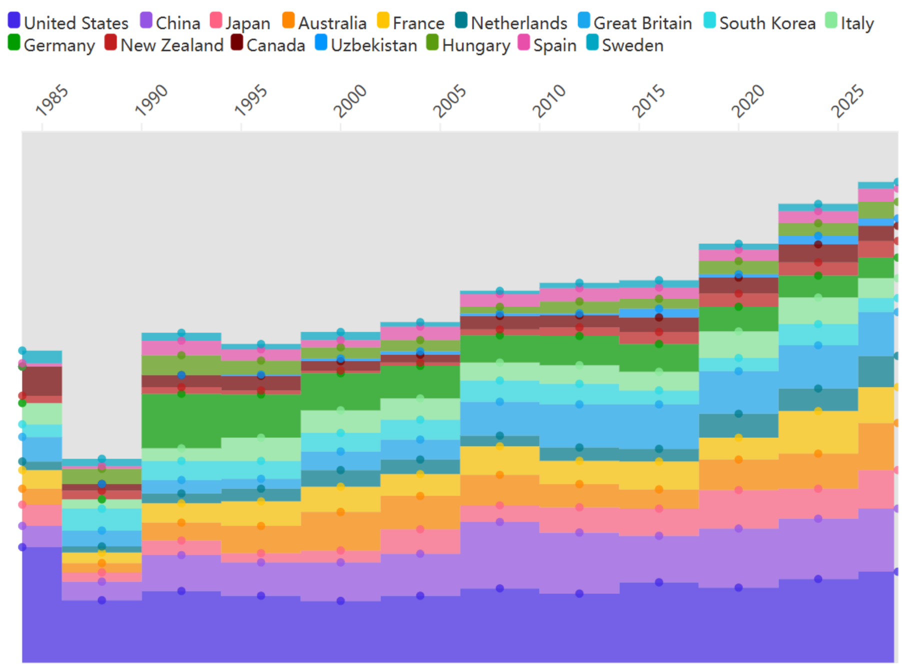


Figure 6: Medal count prediction

4.2 Countries that Win Their First Medal

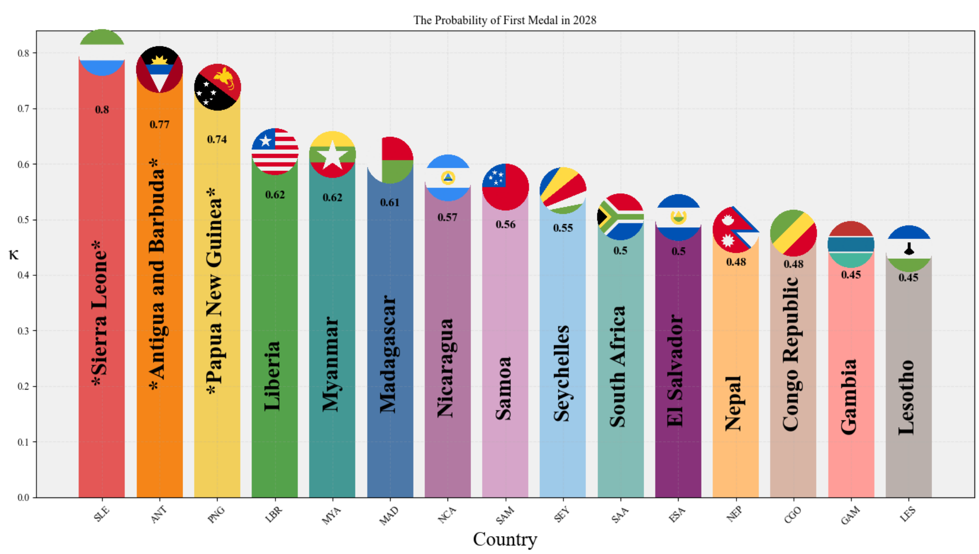
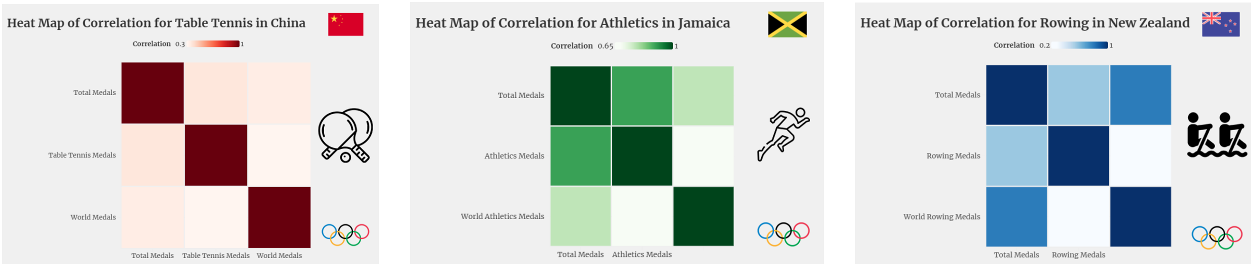


Figure 7: The Probability of First Medal in 2028



(a) Heatmap of Correlation for Table Tennis in China

(b) Heatmap of Correlation for Athletics in Jamaica

(c) Heatmap of Correlation for Rowing in New Zealand

Figure 8: Three images

4.3 Events and Medal Counts by Countries

5 Task2:

The results are shown in Figure ??, where t denotes the time in seconds, and c refers to the concentration of water in the boiler.

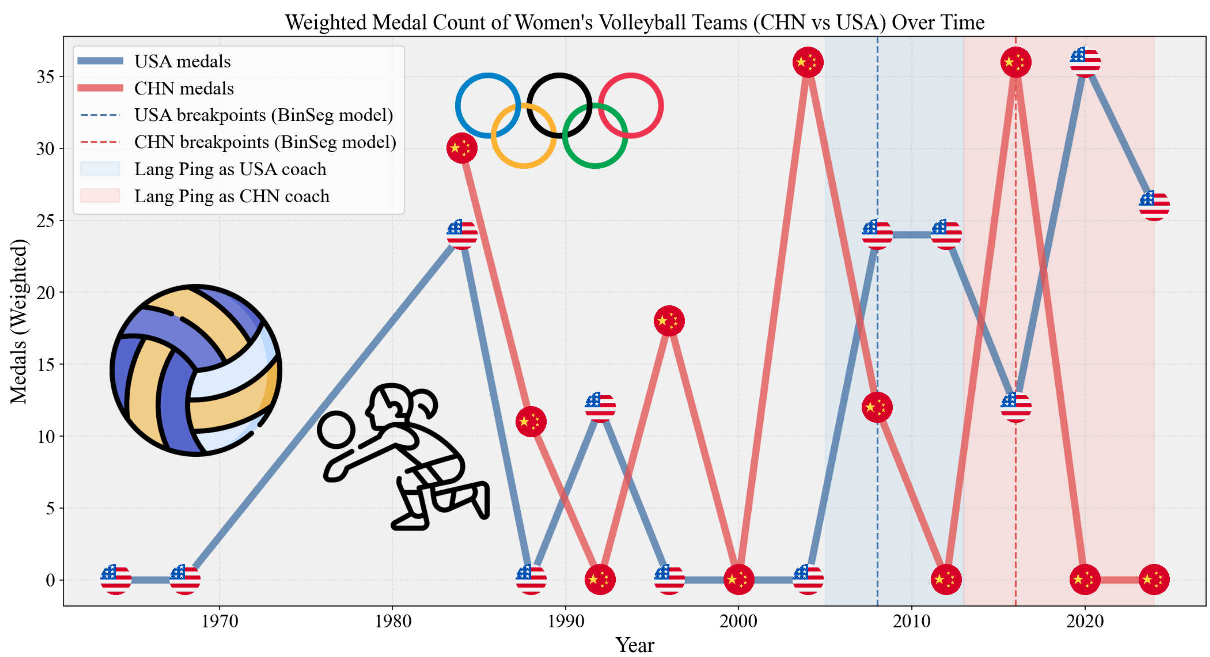
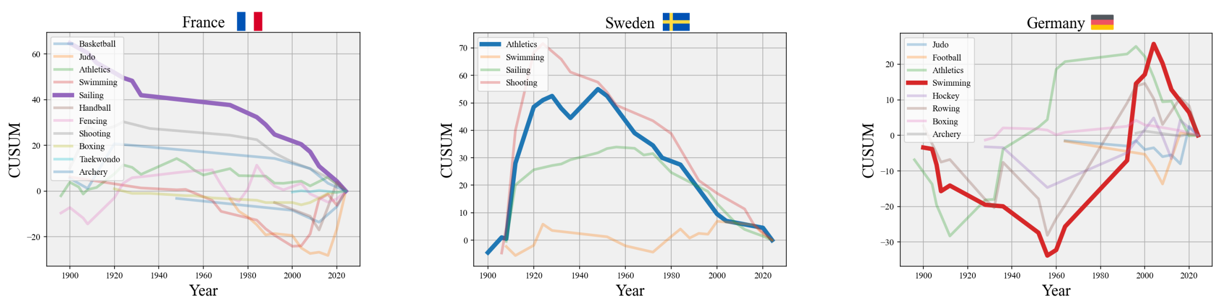


Figure 9: Medal Count of Women’s Volleyball Teams(CHN vs USA) Over Time



(a) France (b) Sweden (c) Germany

Figure 10: Three images

Table 4: Model Performances(LightGBM)

Model	**MSE**	**RMSE**	**MAE**	$\mathbf{R^2}$
Gold Prediction	0.0235	0.180	0.030	0.890
Silver Prediction	0.0239	0.185	0.033	0.825
Bronze Prediction	0.0231	0.187	0.034	0.806

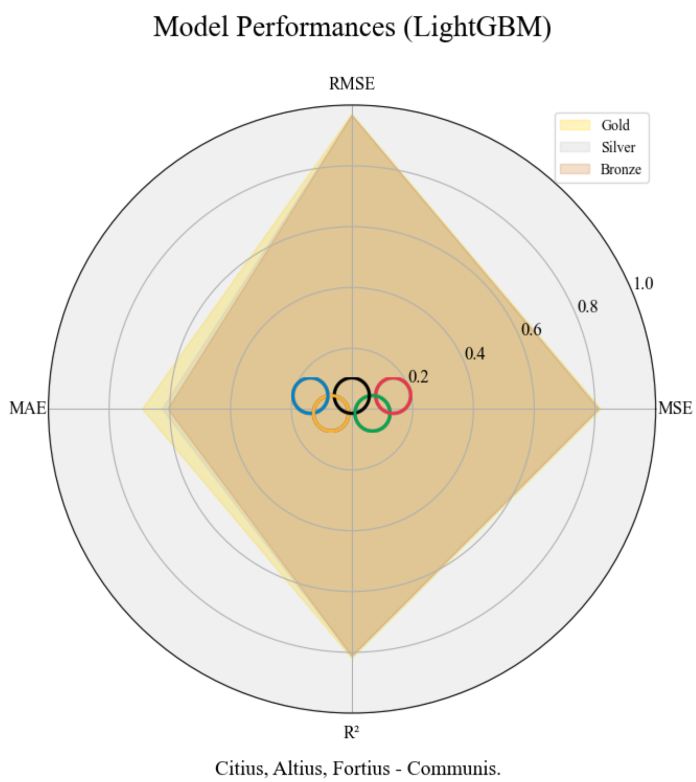


Figure 11: Model Performance Radar Chart

6 Task3

$$\mathscr{L} = \sum_{i=1}^{n} L(y_i, \hat{y}_i^{(m-1)} + \eta f_m(x_i)) + \Omega(f_m) \tag{3}$$

y_i
$\hat{y}_i^{(m-1)}$
$\eta f_m(x_i)$
$\Omega(f_m)$

$$\kappa = 0.8 \times \frac{\xi - \overline{\xi}}{\sigma} \tag{4}$$

$\overline{\xi}$ σ

$$d_i = R(X_i) - R(Y_i) \tag{5}$$

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

d_i R_s

$$P(c_k \mid X) = \frac{P(X \mid c_k)P(c_k)}{P(X)} \tag{7}$$

$$C_t^+ = \max(0, C_{t-1}^+ + (X_t - \mu - k)) \tag{8}$$

C_t^+

$$\phi_i(f) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!}[f(S \cup \{i\}) - f(S)] \tag{9}$$

$\phi_i(f)$

7 Sensitivity Analysis

8 Model Evaluation

8.1 Strengths

1.
2.
3.
4.

8.2 Weaknesses

1.

9 Conclusion

- Task1

  AA
  bbb

- Task2
  AAA
- Task3
  AAA
- Task4
  AAA

References

[1] Einstein, A., Podolsky, B., & Rosen, N. (1935). Can quantum-mechanical description of physical reality be considered complete?. *Physical review*, 47(10), 777.

[2] *A simple, easy LaTeX template for MCM/ICM: EasyMCM*. (2018). Retrieved December 1, 2019, from `https://www.cnblogs.com/xjtu-blacksmith/p/easymcm.html`

Appendix A: Countries' Medal Count Prediction

Table 5: Countries' Medal Count Prediction

Country	G_ pes.	S_ pes.	B_ pes.	G_ opt.	S_ opt.	B_ opt.	Gold	Silver	Bronze	Total
United States	47	36	23	53	41	25	51	40	25	117
China	35	35	15	46	38	20	40	36	17	95
Australia	28	20	22	33	24	24	29	22	23	71
France	23	19	19	27	20	22	24	19	21	64
Germany	18	24	14	23	26	16	21	25	15	61
United Kingdom	20	19	14	26	24	15	25	21	15	61
Japan	16	25	14	19	27	16	17	26	16	58
Italy	15	20	15	17	20	22	16	20	16	53
Spain	21	10	13	30	12	14	25	12	13	50
Netherlands	24	6	13	29	9	15	26	8	15	47
New Zealand	20	13	9	22	15	10	22	14	9	45
Canada	15	10	16	19	10	18	17	10	16	43
Brazil	8	10	13	8	11	15	8	10	14	32
Belgium	6	8	7	15	10	8	11	8	7	26
Hungary	7	6	8	8	8	9	8	8	9	25
Poland	10	7	6	12	10	8	11	9	6	25
Ireland	6	10	6	7	11	6	6	10	6	23
Argentina	4	8	7	5	10	9	4	9	8	22
Denmark	10	7	5	12	8	8	10	8	6	22
Ukraine	3	7	8	5	8	9	4	8	8	21
South Korea	10	5	4	12	6	4	11	6	4	21
Romania	10	7	3	12	8	3	12	7	3	20
Norway	11	1	4	11	3	6	11	2	6	19
South Africa	6	4	4	8	8	5	8	7	5	18
Slovenia	4	3	8	6	3	8	4	3	8	17
Kenya	5	4	4	5	7	5	5	7	4	16
Serbia	4	5	6	5	9	7	4	8	6	16
Mexico	4	4	3	6	4	5	5	4	4	15
India	1	1	4	8	4	5	7	3	4	15
Greece	3	4	2	7	4	3	7	4	3	14
Switzerland	2	4	5	5	4	6	4	4	6	14
Jamaica	1	7	2	5	12	3	4	8	2	13
Czech Republic	5	3	3	6	4	5	6	4	4	13
Uzbekistan	4	2	2	5	3	2	4	2	2	10
Nigeria	4	2	4	4	3	4	4	2	4	10
Sweden	4	4	2	5	5	5	4	4	2	10
Turkey	3	2	5	3	2	6	3	2	5	10
Finland	2	6	2	2	6	2	2	6	2	10
Colombia	5	2	0	5	3	1	5	3	0	9

Egypt	0	3	3	1	4	4	0	4	4	8
Kazakhstan	1	2	2	4	4	2	3	3	2	8
Iran	2	2	1	5	3	2	4	2	1	8
Dominican Republic	3	3	2	6	4	3	5	3	3	8
Portugal	2	3	3	2	4	3	2	3	3	8
Cuba	2	2	1	3	2	2	3	2	2	7
Unknown	1	2	2	1	2	4	1	2	2	7
Latvia	3	1	3	3	1	3	3	1	3	7
Ethiopia	2	1	0	2	6	2	2	2	2	6
Algeria	2	2	1	2	3	2	2	2	1	6
Morocco	0	2	4	0	2	5	0	2	5	6
Austria	3	0	1	3	1	2	3	0	2	6
Puerto Rico	2	3	1	4	3	2	2	3	2	6
Thailand	1	2	1	2	2	2	2	2	1	5
Croatia	2	2	0	2	2	1	2	2	1	5
Ecuador	2	0	2	3	1	3	2	0	3	5
Uganda	0	1	1	1	4	1	1	3	1	5
Angola	0	2	0	1	3	1	0	2	0	5
Israel	2	2	1	2	2	2	2	2	1	5
Mongolia	1	3	1	1	4	1	1	4	1	5
Fiji	0	1	0	2	5	1	1	4	0	5
Uruguay	0	2	0	1	3	1	0	2	0	5
Chile	0	2	0	1	3	1	0	2	0	5
Unknown	1	1	0	3	1	0	1	1	0	4
Zambia	1	3	0	1	3	0	1	3	0	4
Azerbaijan	1	0	2	1	1	2	1	1	2	4
Guinea	0	4	0	0	4	0	0	4	0	4
Georgia	1	1	0	2	2	0	2	2	0	3
Mali	0	3	0	0	3	0	0	3	0	3
Chinese Taipei	2	1	0	2	1	0	2	1	0	3
Venezuela	0	2	1	1	2	2	0	2	2	3
Paraguay	0	3	0	0	4	0	0	4	0	3
Indonesia	0	1	1	1	1	2	1	1	2	3
Peru	0	1	0	1	2	0	0	2	0	3
Iraq	0	3	0	0	3	0	0	3	0	3
Bulgaria	1	1	0	1	1	1	1	1	0	3
Lithuania	2	1	0	2	2	1	2	1	1	3
Tunisia	0	0	0	1	1	0	0	0	0	2
Guatemala	1	0	1	1	0	1	1	0	1	2
Bahamas	0	2	0	1	2	0	0	2	0	2
Trinidad and Tobago	0	1	0	1	1	0	0	1	0	2

Montenegro	0	1	0	0	2	0	0	2	0	2
Hong Kong	0	1	1	0	1	1	0	1	1	2
Moldova	0	1	0	0	2	0	0	2	0	1
South Sudan	0	1	0	0	2	0	0	2	0	1
Malaysia	0	1	0	0	1	0	0	1	0	1
Philippines	0	0	0	1	0	0	0	0	0	1
Bahrain	0	0	0	0	2	0	0	1	0	1
Estonia	0	0	1	0	0	1	0	0	1	1
Liberia	0	1	0	0	2	0	0	1	0	1
United Arab Emirates	0	0	0	0	1	0	0	0	0	1
Jordan	0	0	0	0	1	0	0	0	0	1
Eritrea	0	1	0	0	2	0	0	2	0	1
Singapore	0	0	0	0	1	1	0	0	1	1
Luxembourg	0	0	0	0	1	0	0	0	0	1
Cyprus	0	0	0	0	1	0	0	0	0	1
Samoa	0	1	0	0	1	0	0	1	0	1
Qatar	0	0	0	0	0	0	0	0	0	0
Tanzania	0	0	0	0	0	0	0	0	0	0
Djibouti	0	0	0	0	0	0	0	0	0	0
Ghana	0	0	0	0	0	0	0	0	0	0
Ivory Coast	0	0	0	0	0	0	0	0	0	0
Tajikistan	0	0	0	0	0	0	0	0	0	0
Kyrgyzstan	0	0	0	0	1	0	0	0	0	0
Armenia	0	0	0	0	0	0	0	0	0	0
Slovakia	0	0	0	0	0	0	0	0	0	0
Botswana	0	0	0	0	2	0	0	0	0	0
Grenada	0	0	0	0	0	0	0	0	0	0
Zimbabwe	0	0	0	0	0	0	0	0	0	0
Burundi	0	0	0	0	0	0	0	0	0	0
Kosovo	0	0	0	0	0	0	0	0	0	0

Appendix B: Program Codes

Here are the program codes we used in our research.

test.py

```
# Python code example
for i in range(10):
    print('Hello, world!')
```

test.m

```
% MATLAB code example
```

```
for i = 1:10
    disp("hello, world!");
end
```

test.cpp

```
// C++ code example
#include <iostream>
using namespace std;

int main() {
    for (int i = 0; i < 10; i++)
        cout << "hello, world" << endl;
    return 0;
}
```