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

Together, Individuals Make a Difference

Summary

The Olympic Games serve as both a stage for athletes' performances and a focal point for global competition in the medal table. This study aims to predict the medal distribution for the 2028 Olympics using a bottom-up approach, estimating the total medal count for each country based on individual athletes' expected performance. To achieve this, multiple predictive models were developed to forecast the future medal table and uncover deeper data features, providing valuable decision-making insights for national Olympic committees.

In Task 1, data preprocessing and feature engineering were first performed, utilizing KMeans++ and other statistical methods to extract latent features. Subsequently, the LightGBM model was employed to predict the number of gold, silver, and bronze medals for each country at the 2028 Olympics. The results indicate that a country's medal performance is closely related to factors such as home advantage, adeptness in specific events, and the individual performance levels of athletes. The model demonstrated high accuracy, with an R^2 value of 0.890 for the gold medal prediction, suggesting the reliability of the results.

For countries yet to win medals, our research revealed that participation in high-growth potential events increases the likelihood of securing their first medal. Spearman Correlation Analysis further examined the degree of dependency on specific events by certain countries. The findings show that countries with strong overall sports performance exhibit a lower dependency on specific events, maintaining stable results, while weaker sports nations tend to rely more on select events. In some cases, the correlation coefficient in athletics events was as high as 0.894, indicating a concentration of athletic development in a few specific areas.

In Task 2, we explored the "great coach effect" by applying Bayesian Change Point Detection (BEAST) to analyze instances from multiple countries. The study found a significant correlation between the points of inflection in some countries' sports performance and the tenure of a "great coach," validating the importance of this effect.

Additionally, forward CUSUM analysis was employed to examine changes in medal sequences, identifying the weakest areas for each country. The results show that many countries, which once had competitive advantages in certain events, have seen a decline in these advantages over time. Based on this, we propose targeted coach recommendations for countries in need of improvement in specific areas.

In Task 3, we conducted in-depth data mining and analysis based on the models and data mentioned earlier, revealing several novel insights, such as the gender ratio of athletes, advantages derived from a country's traditional events, and changes in medal counts reflecting shifts in international political dynamics.

The study found that the gender ratio in the Olympics is gradually balancing, with a significant increase in the proportion of female medals.

Furthermore, some countries dominate traditional events, with their medal share reaching up to 88. Consequently, Olympic committees seeking to increase their medal counts should consider investing more in their national traditional events.

Finally, through time-series analysis of changes in medal distribution, multiple anomalies were identified, reflecting not only the evolution of the Olympics from its early stages to maturity but also their connection to the political context of the 20th century.

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

Contents

1 Introduction **4**
1.1 Problem Background . . . 4
1.2 Restatement of Problem . . . 4
1.3 Our work . . . 5

2 Question Preparation **5**
2.1 Assumptions . . . 5
2.2 Notations . . . 5

3 Data Preprocessing **6**
3.1 Basic Data Preprocessing . . . 6
3.2 Data Mining . . . 8
3.2.1 Athlete Service Status . . . 8
3.2.2 Distribution of Countries' Strength Sports . . . 9
3.2.3 Changes in Medal Count . . . 10

4 Task1:Medal Prediction Model Based on LightGBM **10**
4.1 Medal Standings . . . 10
4.2 Countries that Win Their First Medal . . . 13
4.3 Events and Medal Counts by Countries . . . 14
4.4 Model Performances . . . 15

5 Task2:A "Great Coach" Effect **16**

6 Verification of the "Great Coach" Effect **16**

7 Task3 **17**
7.1 Global Geopolitics . . . 19
7.2 Strength events . . . 19
7.3 Global Geopolitics . . . 19

8 Sensitivity Analysis **19**

9 Model Evaluation **19**
9.1 Strengths . . . 19
9.2 Weaknesses . . . 19

10 Conclusion **21**

References **21**

Appendix A: Countries' Medal Count Prediction **22**

Appendix B: Program Codes **24**

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 medal table. Ultimately, the United States topped the medal table with a total of 126 medals, while China and the United States are tied for first place in the number of gold medals, each securing 40 golds. The host country, France ranks 5th in the total number of gold medals with 16 gold medals, but finished 4th in the total medal count. Great Britain ranked 7th on the gold medal table with 14 golds, while securing 3rd 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 3 countries, determine the events in which they should consider investing in "great" 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

Assumption 1 *The number of medals won by each country typically fluctuates only slightly from year to year, without significant increases or decreases.*

The total medal count of a country more accurately reflects the long-term investments and developmental outcomes in sports, rather than short-term results that can be drastically altered. Furthermore, external factors that influence medal counts, such as the organization of international competitions, rule changes, and judging standards, tend to remain relatively stable in the short term.

Assumption 2 *Each year, a number of new athletes join the Olympic Games, some of whom may be competing for the first time. The specific proportion of these athletes will be estimated through our simulation.*

The performance variability of athletes is assumed to be limited, with minimal occurrences of exceptional overperformance or underperformance.

Assumption 3 *The performance variability of athletes is assumed to be limited, with minimal occurrences of exceptional overperformance or underperformance.*

The performance variability of athletes is assumed to be limited, with minimal occurrences of exceptional overperformance or underperformance. Due to the systematic nature of athlete training and the relative consistency of competition environments, their performance tends to remain stable. By mitigating the impact of outliers, the model can avoid being influenced by extreme values.

2.2 Notations

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

Table 1: Notations

Symbol	Definition
C_i	Certain country
E_i	Specific event
G	Gold medal count
V	Silver medal count
B	Bronze medal count
M	Total count of the 3 types of medals
A	Number of new athletes
T	Times of Olympic participations
γ	Sport Advantage Coefficient
m_{c_i}	Total medal count of a certain country
m_{c_i,e_j}	Total medal count of a certain country in a specific event
L_i	Country clustering level
g_i	the second one
X	Historical total medal counts (gold, silver, and bronze) of all countries
ξ	The Probability of First Medal
σ	
	Normalized version of probability of first medal

3 Data Preprocessing

3.1 Basic Data Preprocessing

Due to various influencing factors, there are significant differences in the competitive levels of countries in the Olympic Games, which have been reflected in past competitions, specifically in the total number of gold, silver, and bronze medals won by each country.

To illustrate the differences in the levels of countries in previous Olympic Games (i.e. national level), we employed the KMeans++ clustering algorithm. Based on the number of gold, silver, and bronze medals, as well as the total medal count of each country in past Olympics, we classified the countries into five levels. These five levels, L1, L2, L3, L4, and L5, form a partition of the sample set X (the total number of gold, silver, and bronze medals won by countries in previous Olympic Games).

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

After the classification, the distribution of countries across the 5 levels is shown in the figures:

In addition, we created 10 new features based on the original variables, which were utilized during the modeling process with the LightGBM model.

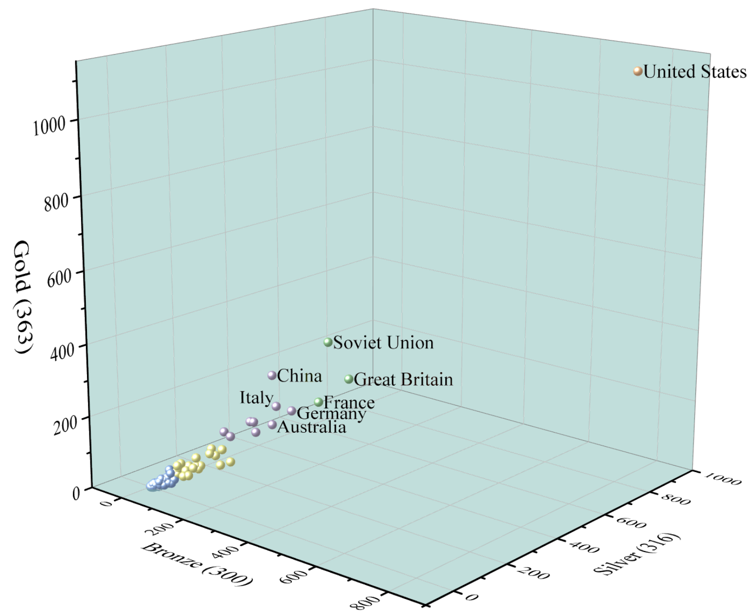


Figure 1: Scatter plot of national level classification (based on Kmeans++clustering algorithm)

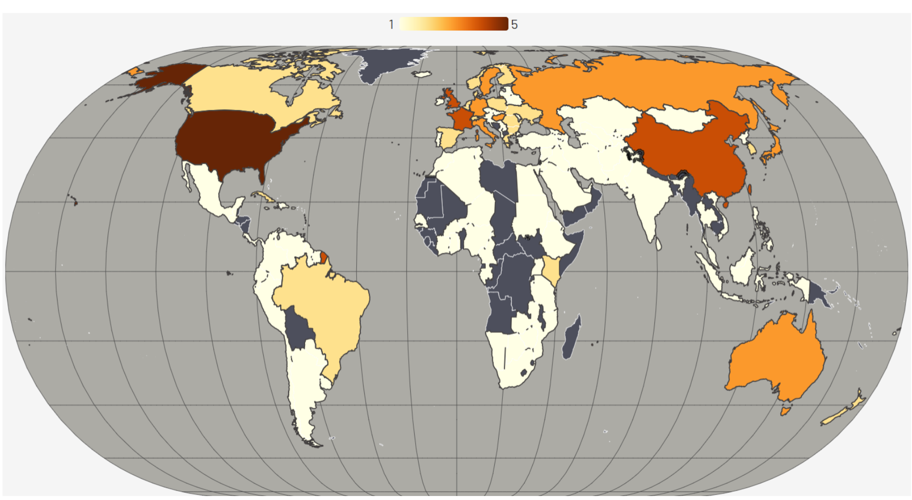


Figure 2: Geographical Distribution Map of Country Level Classification

Table 2: Variable Name

Variable Name	*Code*	Definition
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

3.2 Data Mining

3.2.1 Athlete Service Status

To predict future scenarios, we analyzed the trends in the number of participants for each country, event and year, and applied **Linear Regression** for fitting.

Then, we compiled the distribution of the times of participations and the distribution of athlete tenure based on historical data, as shown in the figure.

From the distribution chart, it can be seen that nearly 80

To predict the Participant List, we employed Monte Carlo simulation, which consisted of two parts: the first part involved selecting 20

In the previous sections, we determined which athletes will continue to compete. To simulate new random athletes, we simplified each individual into three parameters: the probabilities of winning a gold, silver, or bronze medal.

For a given country in a specific event, the average number of medals won by all athletes serves as the standard number of medals for a representative athlete. The medal expectations were

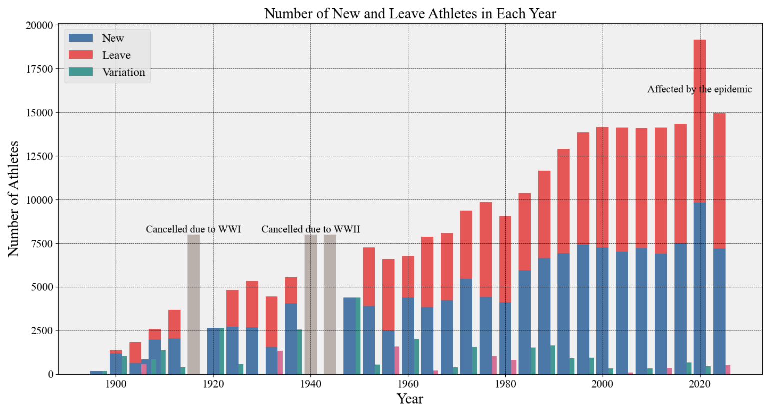
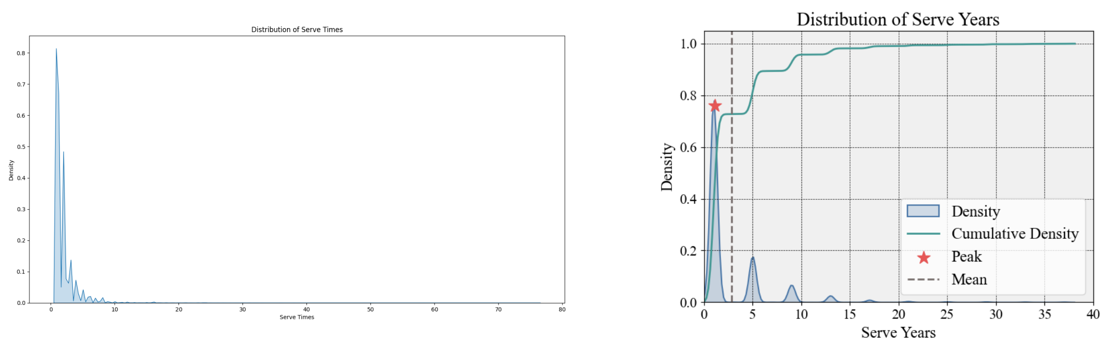


Figure 3: Number of New and Leave Athletes in Each Year



(a) Distribution of Serve Times (b) Distribution of Serve Years

Figure 4: Two images

categorized into 3 scenarios: optimistic, moderate, and pessimistic. The optimistic expectation was calculated by averaging the top 50

Using the Previous Participants' Records as the basis for Monte Carlo simulation, we ultimately determined the Participant List Prediction.

3.2.2 Distribution of Countries' Strength Sports

Due to factors such as cultural and geographical environments, countries exhibit distinct advantages in different events. For example, Kenya and Ethiopia excel in long-distance running, particularly in Athletics, largely due to the training conditions in high-altitude regions. Brazil, with its deep football culture, has produced many world-class players, a result of its warm climate and environment conducive to year-round outdoor sports.

By analyzing the performance of various countries in different events across past Olympic Games, we extracted a feature variable γ(Event Adeptness Coefficient), to represent the advantages of these countries in certain events..

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

By extracting features, we mined the level of expertise of each country in different events within the dataset, such as China's performance across various events, as shown in the figure below:

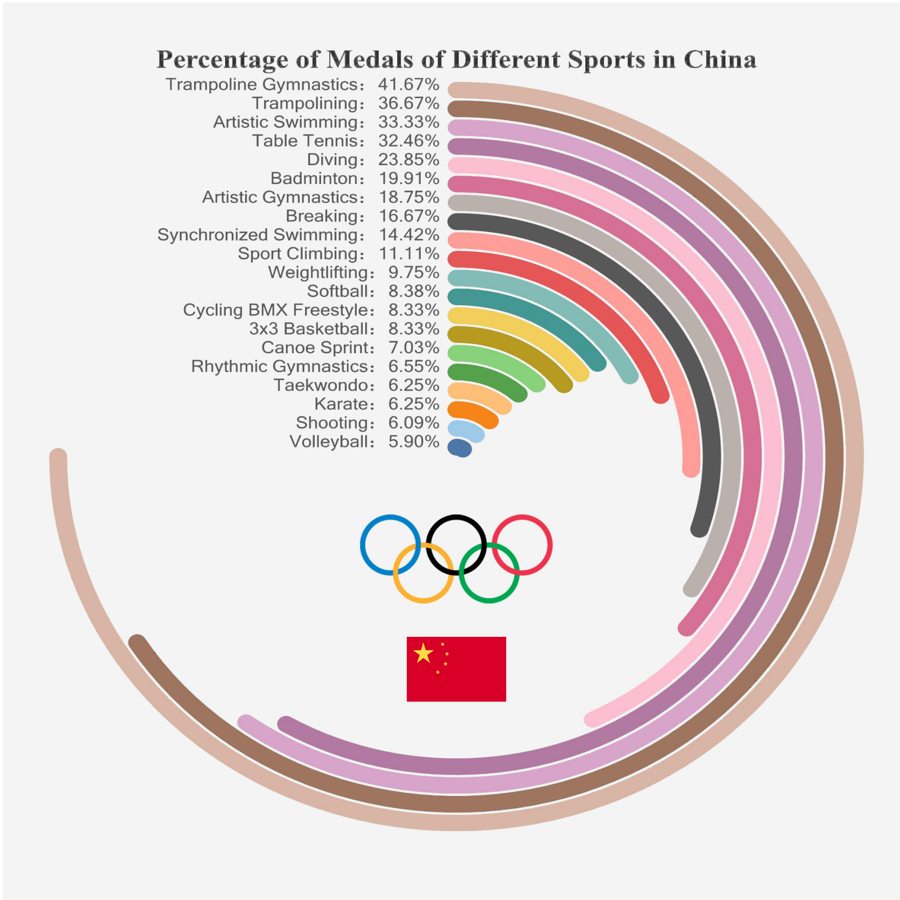


Figure 5: Percentage of Medals of Different Sports in China

3.2.3 Changes in Medal Count

Since the IOC makes adjustments to events each year, the total number of medals awarded in each event does not remain exactly the same each year. To address this, we applied **L**inear Regression to fit the number of gold, silver, and bronze medals awarded in each event across past Olympic Games and predicted the medal distribution for the upcoming Olympic Games, providing a reference for subsequent operations.

4 Task1:Medal Prediction Model Based on LightGBM

4.1 Medal Standings

Numerous factors influence the medal table. To predict the medal standings for the next Olympics, we incorporated factors identified during the data preprocessing stage, including nation level, the list of participating athletes, each country's adeptness in specific events, and adjustments to medal counts.

Additionally, as host countries typically benefit from home advantage, the factor of "whether a country is the host"or "whether hosting for the first time" was also considered. This factor was categorized into 3 categories and quantified for inclusion in the model. Ultimately, these five factors were collectively summarized as **National Background Factors**.

Category	Definition
0	Not host
1	Host for the 1st time
2	Host for more than once

The prediction of the Olympic medal table is essentially a multi-task learning problem, with the goal of predicting the number of gold, silver, and bronze medals each country is likely to win in different events. To accomplish this task, we need to input a large and diverse set of data while outputting predictions for multiple medal categories. These features are highly suitable for the LightGBM model.

Building on GBDT, LightGBM accelerates the computation using histograms, optimizing training efficiency and enabling the fast processing of high-dimensional, sparse, large-scale data. Compared to traditional logistic regression, LightGBM can capture nonlinear relationships through tree splitting, offering a performance advantage in accurately handling complex data and nonlinear relationships, significantly improving accuracy. Additionally, it can automatically handle missing values, providing great convenience for data preprocessing.

As a gradient boosting-based regression model, in LightGBM, the objective function consists of two parts: the loss function and the regularization term.

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

The optimized loss function based on GBDT is represented as squared error in the medal prediction regression task.

$y_i$denotes the true value. $\hat{y}_i^{(m-1)}$is the current model's predicted value. $\hat{y}_i^{(m-1)}$+$\eta f_m(x_i)$is the model's predicted value at the m-th step during the gradient boosting model update process. $\Omega(f_m)$ Regularization term, which prevents overfitting of the model and controls the complexity of the trees.

We input the previously obtained data into LightGBM for iterative model training. After training, we input the factors affecting the predictions for the next edition into the model, ultimately obtaining the predicted medal table for the 2028 Summer Olympics in Los Angeles, USA, as follows: According to the model's predictions, the United States is expected to see a significant increase in medal count compared to the previous year, while France will experience a notable decrease, which is related to the home-field advantage of the host country. Additionally, due to changes in the number of medals for certain events, the medal counts for Australia, Japan, and the Netherlands are expected to rise, while Italy and South Korea will see a decline. China's medal count is projected to remain relatively stable.

To more accurately simulate the actual number of medals won, we simplify each individual into 3 parameters: the probabilities of winning gold, silver, and bronze medals. For a specific country and event, the average number of medals won by all team members represents the probability that a standard "virtual athlete" should have.

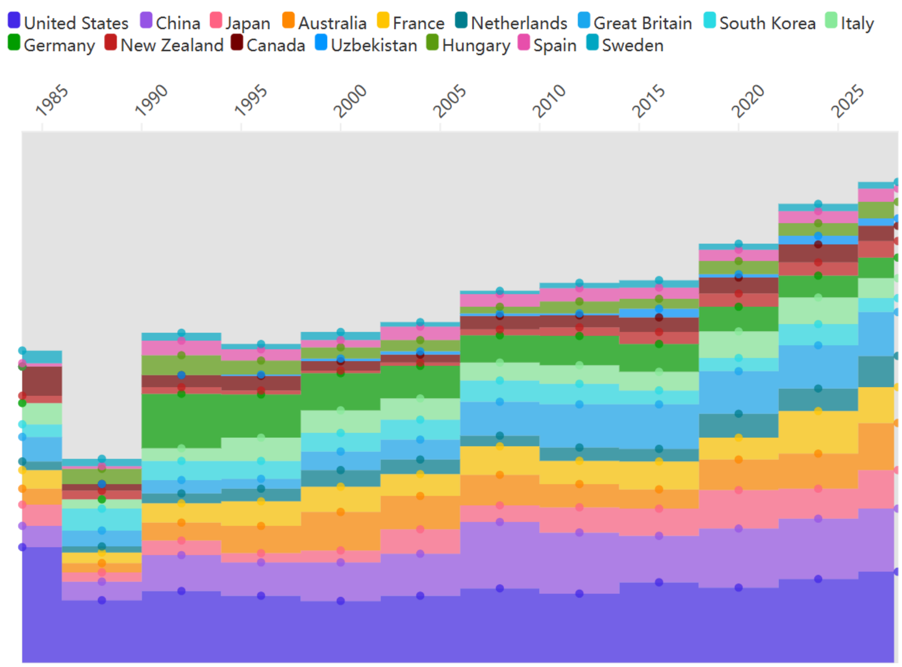


Figure 6: Medal count prediction

For the expected outcomes, the optimistic expectation is calculated by averaging the top 50

The "Previous Participants Record" serves as the basis for the Monte Carlo simulation, which consists of two parts: the first part involves selecting 20

Based on the model, we predict the medal outcomes for each country in the 2028 Olympic Games. A partial result is provided here, with the full version available in the appendix. The predicted outcomes for the United States are presented, showing that, overall, the optimistic expectation > moderate expectation > pessimistic expectation.

Table 4: 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
...	..	..	..	..	..	..	..	..	..	..

4.2 Countries that Win Their First Medal

Since the model used for predicting the medal table in the previous section yields varying levels of accuracy for countries of different rankings, the predictions for higher-ranked countries, such as the United States and China, are more accurate. In contrast, the prediction errors for lower-ranked countries are larger. Clearly, countries that have not yet won medals fall into the category with higher prediction errors, so the previous model is not applicable in this case.

Therefore, for countries that have never won a medal before, we calculate their first medal index ξ.

The number of competitions is counted from the original dataset, and the expected increase in the number of medals for the next edition is predicted through a Linear Regression model based on historical data.

As the number of competitions and the expected increase in medals for the next edition increase for countries that have not won medals, the likelihood of these countries winning medals becomes higher. Therefore, the first medal index ξ is positively correlated with the probability of a country that has not won a medal securing its first medal in the next Olympic Games, meaning that the larger the ξ value, the higher the likelihood of winning the first medal.

In order to visualize the accuracy of a country's first medal prediction, we normalized the first medal index using the Z-score and then multiplied it by 0.8 to obtain the first medal coefficient κ.

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

Here,$\overline{\xi}$ refers to the mean of the original data, andσrefers to the standard deviation of the original data. The calculated results are as follows:

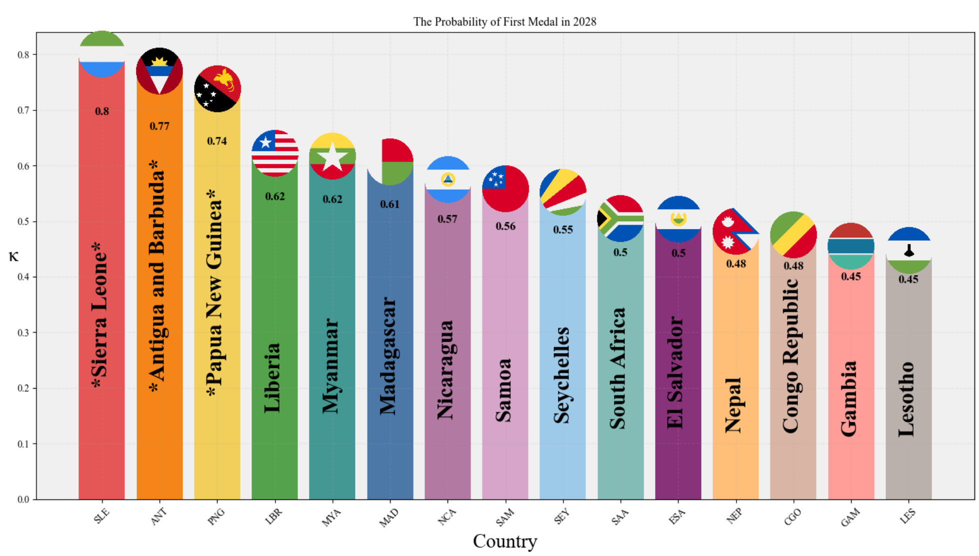


Figure 7: The Probability of First Medal in 2028

As shown in the figure, Sierra Leone (SLE), Antigua and Barbuda (ANT), and Papua New

Guinea (PNG) are the three countries most likely to win their first medal in the next Olympic Games, with accuracies approximately 0.8, 0.77, and 0.74, respectively.

4.3 Events and Medal Counts by Countries

In the data preprocessing section, we have calculated the adeptness of each country in each event, thus allowing us to roughly estimate the events in which each country excels.

We input the data of the dominant events of multiple countries into the prediction model and use **Spearman's Rank Correlation Coefficient** to analyze the correlation between the total medal count of a country and the medal count of a specific event, as well as the number of events in that event. This is because we cannot confirm that the data sample follows a normal distribution, nor can we ascertain that the relationship between the dominant events and the total medal count is linearly correlated. Spearman's Rank Correlation Coefficient is suitable for evaluating **nonlinear** relationships.

We studied the relationship between China's total medal count, gold medal count, and the number of awards and events in table tennis. For each pair of variables, we ranked the data points and assigned ranks to each data point. The rank difference for each data point is calculated as:

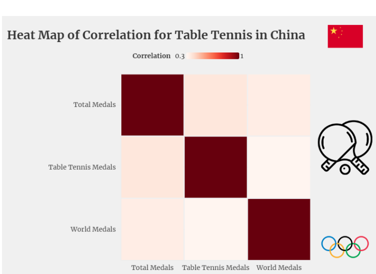
$$d_i = R(X_i) - R(Y_i) \tag{5}$$

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

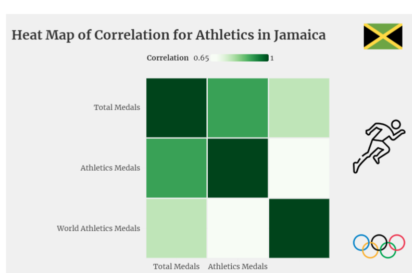
Here, d_i is the rank difference of the two variables for the iii-th data point. Next, we compute the Spearman's Rank Correlation Coefficient (R_s).

The absolute value of R_s closer to 1 indicates a stronger correlation, while closer to 0 indicates a weaker correlation.A positive R_s indicates a positive correlation between the two variables, meaning an increase in one variable typically accompanies an increase in the other.

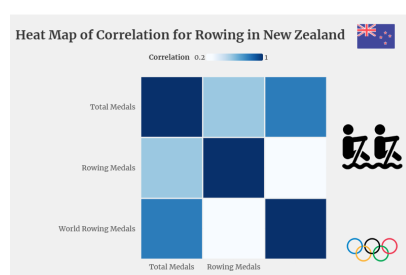
Through extensive data analysis, we identified 3 typical examples: table tennis in China, Athletics in Jamaica, and rowing in New Zealand, as shown in the figure.



(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

We found that although China has a significant advantage in table tennis, the relationship between the development of the event and China's total medal count is relatively weak. This suggests that China's overall athletic adeptness is stable and less influenced by a single dominant event.

Additionally, Jamaica's total medal count shows a strong positive correlation with its performance in athletics, indicating the country's heavy reliance on athletics competitions. Since athletics is a well-established and stable event with minimal fluctuations in the number of events held each year, the relationship between the number of events and Jamaica's total medal count is relatively weak.

New Zealand also has a high dependency on rowing, with a substantial portion of its medals coming from this event. However, unlike Jamaica, rowing is not as established as athletics, and the number of competitions fluctuates from year to year. As a result, New Zealand's overall medal count is highly dependent on the rise and fall of rowing's prominence.

4.4 Model Performances

We calculated four evaluation metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the Coefficient of Determination ($\mathbf{R}^2$).

Among these, smaller values for the first three metrics and a larger value for R^2 indicate better model performance.

Table 5: 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

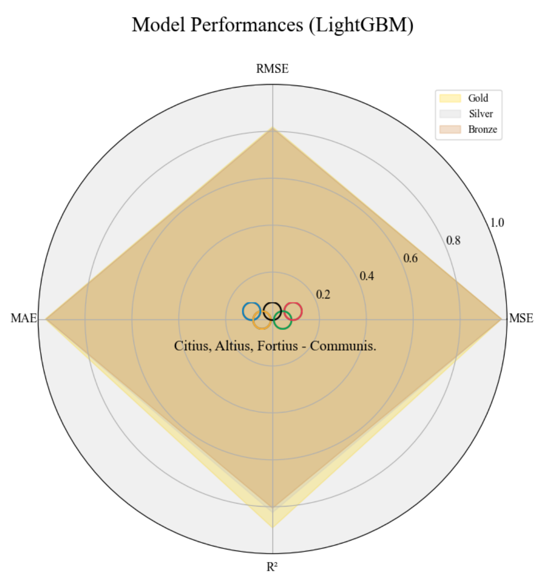


Figure 9: Model Performance Radar Chart

It can be observed that the evaluation metrics for the prediction of the 3 types of medals show good performance with relatively small differences.

A comparison reveals that the model's prediction accuracy for gold medals is relatively higher, with the highest $\mathbf{R}^2$ and the lowest Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE).

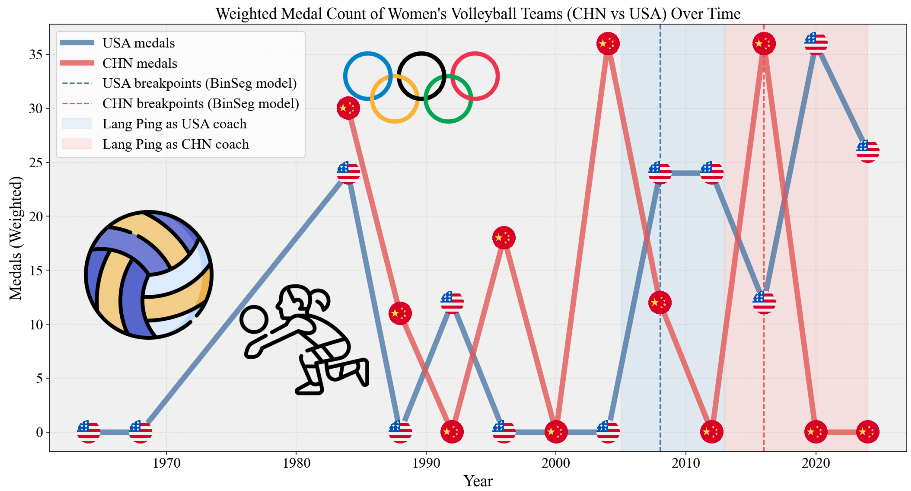
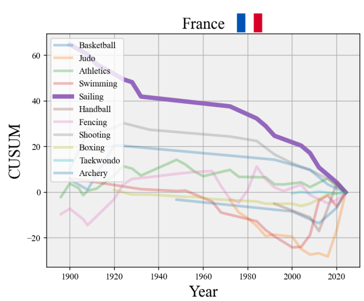
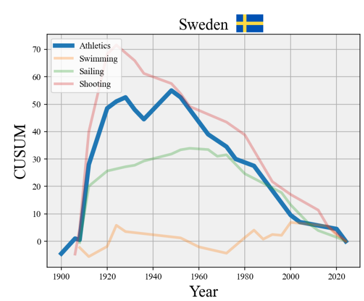


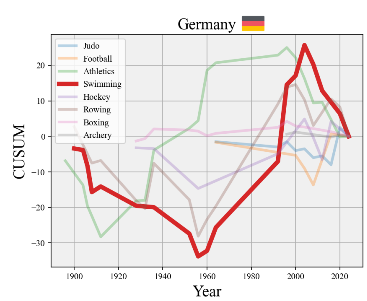
Figure 10: Medal Count of Women's Volleyball Teams(CHN vs USA) Over Time



(a) France



(b) Sweden



(c) Germany

Figure 11: Three images

5 Task2:A "Great Coach" Effect

6 Verification of the "Great Coach" Effect

The "Great Coach" effect can cause significant differences in the performance of relevant countries in corresponding events. To collect evidence of changes caused by this effect, we use a Bayesian Changepoint Detection (BEAST) model.

The BEAST model uses Bayesian inference to calculate the posterior probabilities of changes occurring at different time points, determining which time points are potential change points.

In the time series dataset $X = (x_1, x_2, x_3, \ldots, x_T)$, we identify the set of change points $C = \{c_1, c_2, \ldots, c_K\}$, where each c_k represents a change point. Using Bayes' theorem, we can compute the posterior probability for each possible change point c_k:

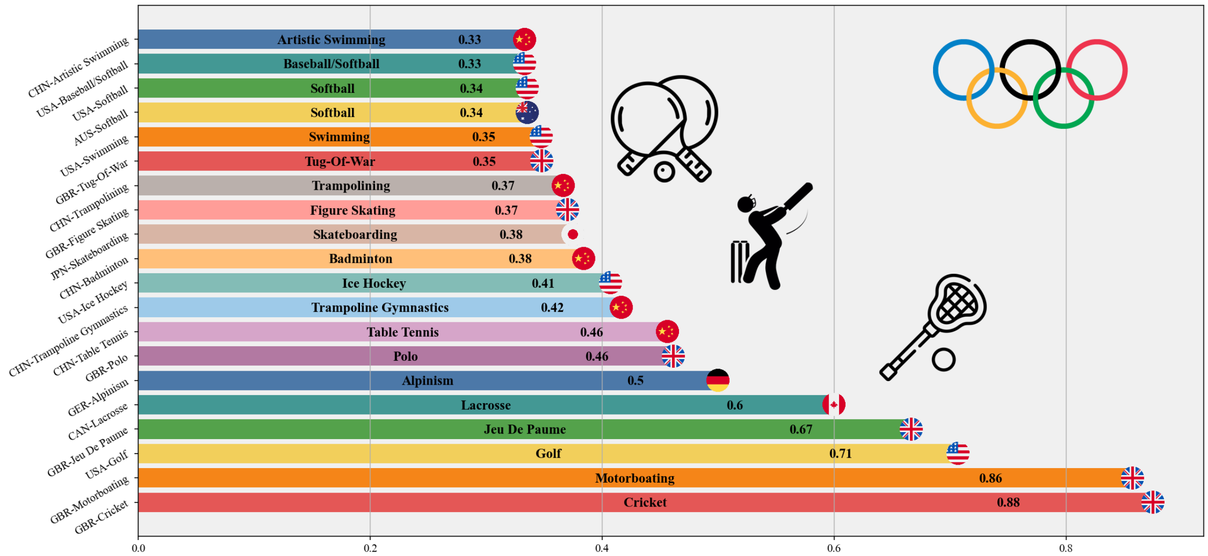


Figure 12: Adeptness of Countries in Specitific Project

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

In this case, $P(X|c_k)$ represents the likelihood of the data at the change point c_k, $P(c_k)$ is the prior probability, and $P(X)$ is the marginal likelihood of the data.

7 Task3

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

C_t^+

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

$\phi_i(f)$

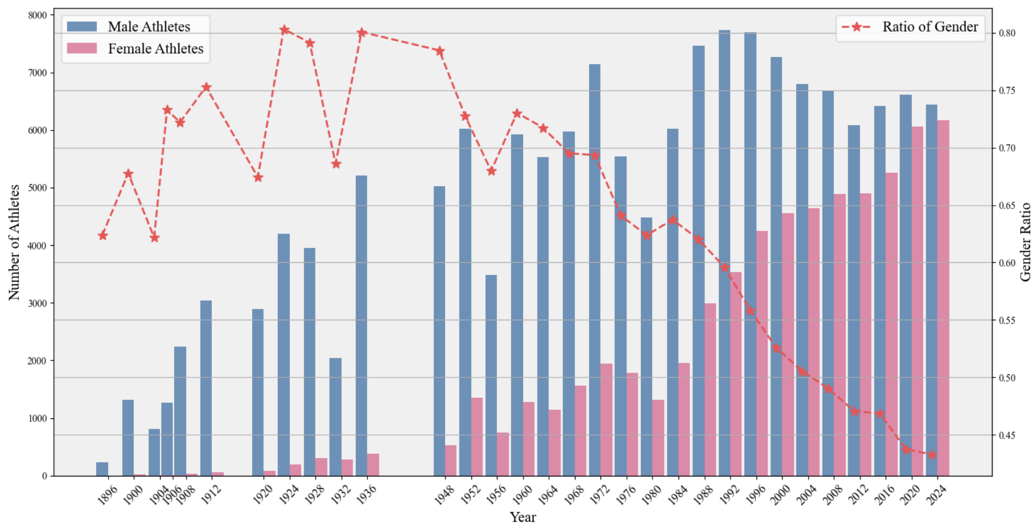


Figure 13: Gender Distribution of Athletes(with Award)

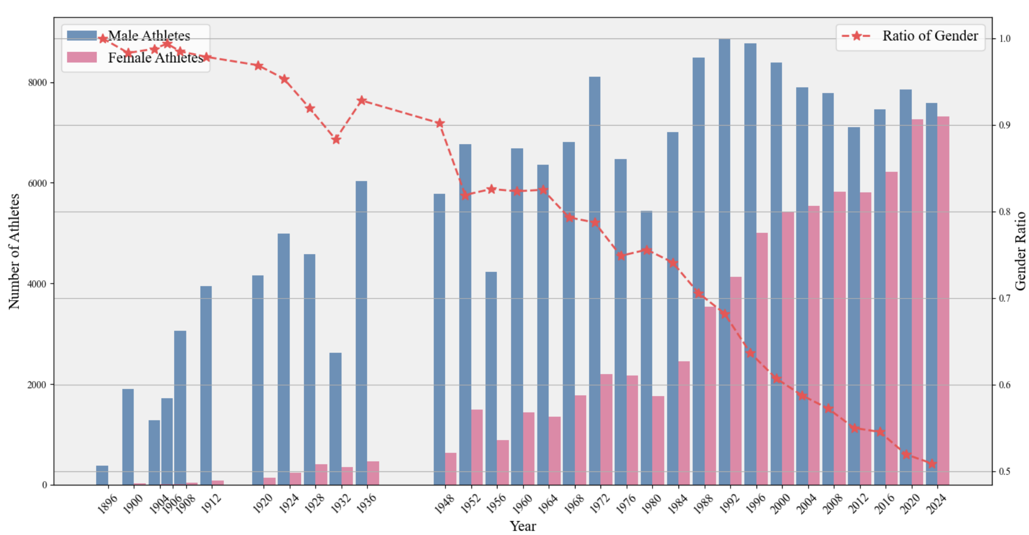


Figure 14: Gender Distribution of Athletes

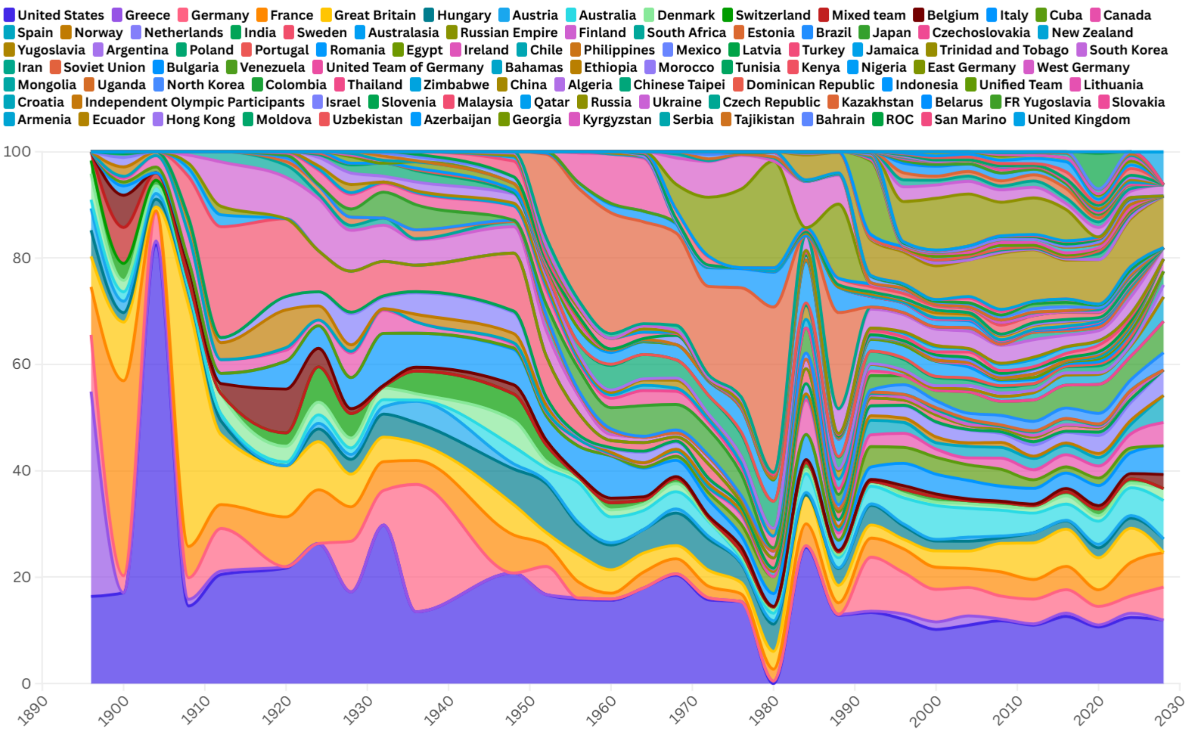


Figure 15: Trend of Medal Proportions by Country Annually

7.1 Global Geopolitics

7.2 Strength events

7.3 Global Geopolitics

8 Sensitivity Analysis

9 Model Evaluation

9.1 Strengths

1.
2.
3.
4.

9.2 Weaknesses

1.

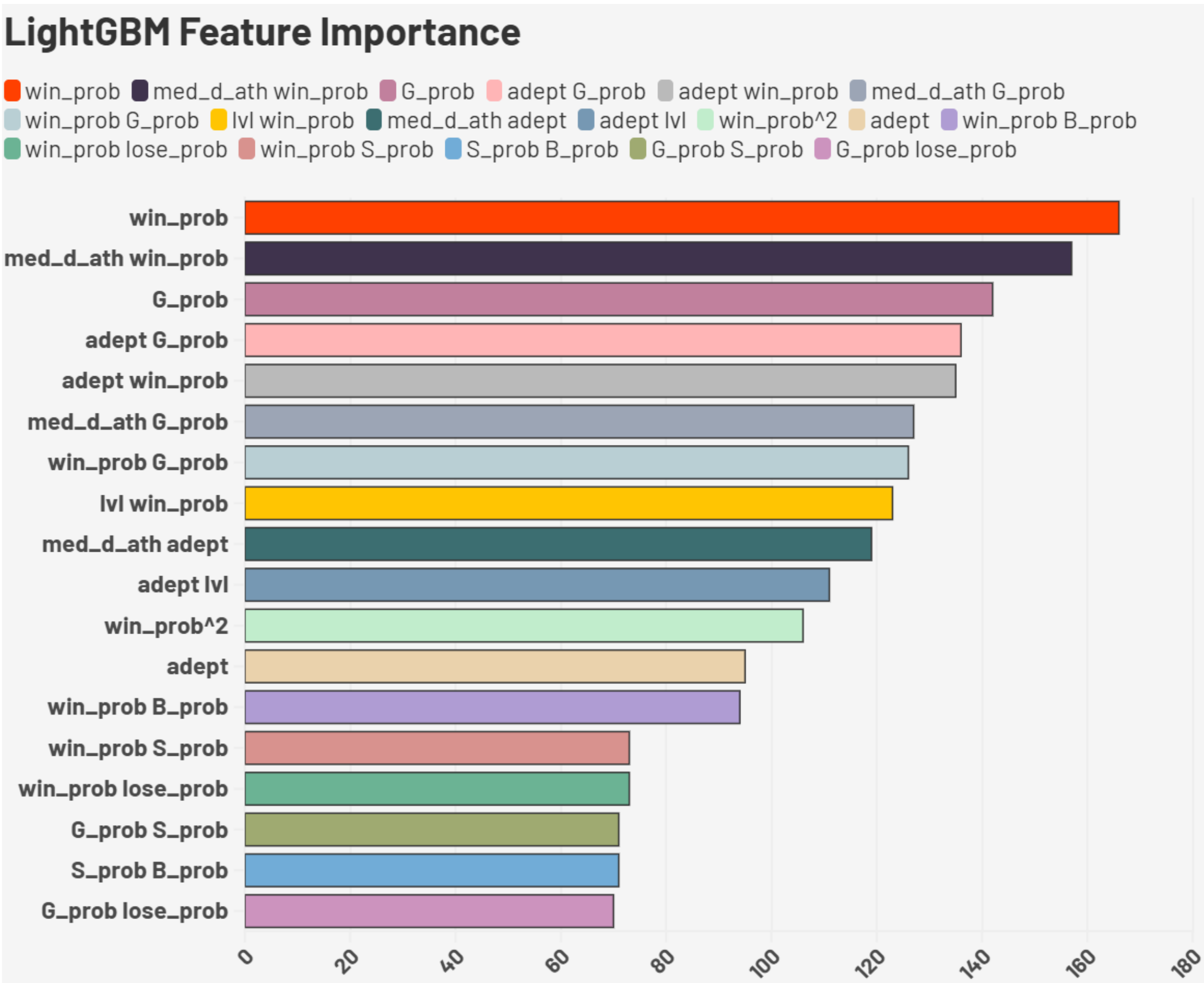
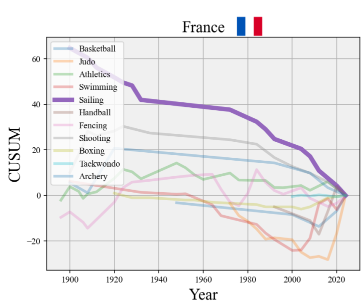
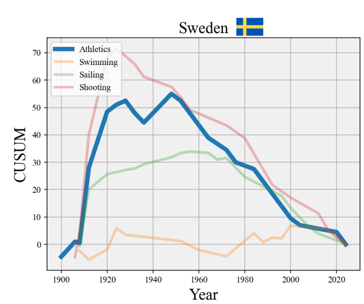


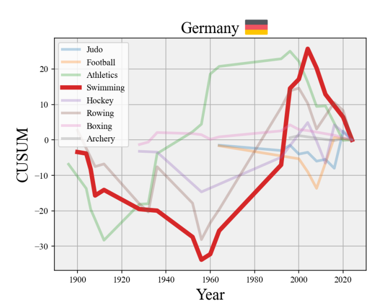
Figure 16: LightGBM Feature Importance



(a) France



(b) Sweden



(c) Germany

Figure 17: Three images

10 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 6: 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;
}
```