2024 Paris Olympic Medal Prediction

Haoxian Ruan, Jae Eun Lee
Data Science Department, University of Connecticut

I. Introduction

The prediction model in this paper will identify the group of 5 athletes that will enable the Team USA Olympic Men's and Women's Artistic Gymnastics teams to optimize success in Paris 2024. The teams to optimize success represent the teams consisting of athletes that can maximize the medals. The model will also determine the expected number of medals in the eight medal events for men (individual, team, and floor exercise; pommel horse, still rings, vault, parallel bars, and high bar) and the six medal events for women (balance beam, uneven bars, vault, team, and individual all-around).

II. Dataset

LastName
FirstName
Gender
Country
Date: the date when the competition
Competition: the name of the competition
Round: qual and final
Location
Apparatus: a piece of equipment
Rank
D_Score
E_Score
Penalty
Score: Score is calculated by the formula of D_Score + E_Score – Penalty.

III. Methodology

1. Data Preprocessing
   - Dimension reduction and data reduction
2. Data Analysis and Exploration
   - Conducting Exploratory Data Analysis (EDA) to comprehend data distributions and pinpoint potential patterns

3. Modeling

- The point-ranking system ranks athletes in each apparatus, determining medalists for individual apparatus. For team combinations, the system ranks potential combinations by expected medal count for the selection of combinations likely to secure the highest number of medals.

IV. Function Construction

[RANK(apparatus, gender)]

Ranks athletes based on gender and each apparatus, considering competition importance and weighted scores.

[RANK_AA(gender)]

Ranks all-around gymnasts within each gender, incorporating total scores across different apparatus.

[get_p(data, team)]

Calculates medal points for each team in each apparatus, considering athlete ranks.

[point_apparatus(team, ap_g)]

Calculates the total medal points for a given team in a specific apparatus by considering the top three ranked athletes from the team.

[mean_point_AP(ap, ID, df_ID)]

Computes average points an athlete can achieve in a specific apparatus.

[best_three(ap, team, cdd)]

Identifies the top three athletes for each team in a particular apparatus and calculates their total points.

[team_point(cm, cdd, gender)]

Determines the team all-around points for each combination, considering the best three athletes in each apparatus.

V. Modeling

1. When calculating total points for the team all-around, the focus is on the top 3 players per team, in line with the scoring that prioritizes the best 3 scores in the team final.

2. The scoring system assigns 100 points for a gold medal, 10 for silver, and 1 for bronze, underscoring the importance of achieving gold. This prioritization not only emphasizes the quality of medals but also facilitates a straightforward interpretation of the team's potential medal count from the final total points.

3. The selection of the final five considers the top 5 players from each apparatus in the USA. This ensures that the optimal team all-around combination comprises athletes consistently excelling across all apparatus.

4. Significantly higher weightage is given to Olympics and FIG World Championships compared to World Cup or Asian Games. This decision stems from the likelihood of top players participating in major events like the Olympics and World Championships, in contrast to World Cup games that occur multiple times a year.

VI. Result

All the combinations were ranked based on individual apparatus points. Recognizing that team all-around is determined by total points, our selection process prioritized combinations with the highest total points derived from individual apparatus rankings. After selecting combinations based on individual apparatus points, the function computed team all-around points and identified those combinations with the highest team all-around points.

Two gold, three silver, and three bronze medals are expected. And the team consisting with these members are expected

The Best USA Team (Women)	**The Best USA Team (Men)**
Simone Biles	Brody Malone
Sunisa Lee	Paul Juda
Shilese Jones	Colt Walker
Jade Carey	Asher Hong
Mykayla Skinner	Khoi Young

Limitation:

1. The assigned score weights to each competition are subjectively decided and may introduce bias into the analysis. ~~Consideration should be given to refining this aspect for more robust results.~~
2. The model solely relies on final data, excluding national competition data, limiting its ability to fully leverage all available information and potentially compromising the accuracy of rankings.
3. The prediction assumes an ideal scenario where all top athletes from other countries participate, overlooking the actual composition of final athletes in competing nations.
4. The model lacks the capacity to predict specific medals for the team all-around, as it does not consider the compositions of other competing teams, which can influence the overall results.
5. The ranking system may favor more frequent participants, potentially skewing results based on participation rather than skill level.

VII. Conclusion

In conclusion, our prediction model integrates a comprehensive methodology, optimized the Team USA Olympic Men’s and Women’s Artistic Gymnastics teams' success in Paris 2024. By prioritizing both individual apparatus points and overall team performance, the model identifies top-performing combinations. The functions [Rank], [RANK_AA], [get_p], and [team_point] collectively contribute to a thorough evaluation of athletes and teams. While the model provides valuable insights, the subjective nature of competition score weights remains a limitation, signaling the need for ongoing refinement. As we look ahead to Paris 2024, this data-driven approach proves instrumental in strategic decision-making for the selection of teams poised for optimal success on the international stage.