

Leveraging Analytics to Assemble USA's 2024 Olympic Gymnastics Teams

Abstract

This research utilizes a decade's worth of data (2013-2023) from diverse sources to employ data scraping, cleaning, exploratory data analysis (EDA), and modeling to determine the optimal composition of USA Olympic Artistic Gymnastics teams for maximizing medal counts at the 2024 Olympic Games. By integrating advanced machine learning models and optimization algorithms, this project aims to strategically position Team USA for the highest potential medal haul.

Introduction

The challenge of this research is to identify the optimal Men's and Women's USA Olympic Artistic Gymnastics teams for the Paris 2024 Olympics. This involves forecasting medal counts based on historical performance data to aid in strategic decision-making. The project incorporates advanced analytics to predict individual apparatus outcomes and optimize team selection through a novel application of Metaheuristics Algorithms.

Literature Review

Prior research in the field of sports analytics has extensively utilized predictive modeling and optimization algorithms. Studies such as those presented at the UConn Sports Analytics Symposium 2024 have discussed various optimization methods. However, this research uniquely integrates Randomized Heuristic, Tabu Search, and Variable Neighborhood Search (VNS) for team composition optimization.

Project Details

- **Tools:** Python for data scraping and modeling, TensorFlow/Keras for neural networks, Scikit-learn for other ML models.
- **Datasets:** Data from Thegymter.net, Wikipedia.org, and UCSAS 2024, focusing on international competitions.
- **Techniques:** Machine Learning models for binary classification, Metaheuristics for optimization.
- **Tasks:** Data scraping and preprocessing, individual apparatus prediction, team composition optimization.

Evaluation Plan

The project's success will be measured by the accuracy, precision, recall, F1-Score, and F2-Score of the predictive models and the effectiveness of the optimization algorithms in maximizing the theoretical medal count. The project will consider a model successful if it significantly outperforms baseline models (ZeroR).

Timeline

- **Week 1-3:** Data collection and preprocessing.
- **Week 4-6:** Development and training of machine learning models.
- **Week 7-9:** Optimization algorithm implementation and testing.
- **Week 10:** Integration and final testing.
- **Week 11-12:** Analysis of results and preparation of findings for publication.

References

This research utilizes publicly available data from:

- Thegymter.net
- Wikipedia.org
- UConn Sports Analytics Symposium (2024).