Leveraging Analytics to Assemble USA's 2024 Olympic Gymnastics Teams

Ryan Talbot

July 5, 2024

Challenge of selecting the optimal Olympic gymnastics team

The challenge of selecting the optimal USA Olympic gymnastics team for the 2024 Paris Olympics involves using advanced data analytics to predict individual performances and optimize team composition. This research utilizes machine learning and metaheuristic algorithms, aiming to maximize Team USA's medal count by analyzing comprehensive data from 2013 to 2023.

Data Sources

- **Thegymter.net** - This source provides comprehensive data on gymnastic competitions, including individual scores, overall performances, and event rankings.
- **UCSAS 2024 USOPC Data Challenge** - This platform offers datasets specifically prepared for analytics challenges, focusing on Olympic sports data.
- **Wikipedia.org** - Used for general information and historical data on athletes and past Olympic events.

Methodology and Analytical Tools

- Data Collection and Preprocessing

  - **Sources**: Data gathered from UCSAS 2024 USOPC Data Challenge, and data scraped from Thegymter.net and Wikipedia.org, covering events from 2013 to 2023.

  - **Preprocessing**: Involves cleaning data for missing values, standardizing column names, and combining data from various competitions. Special attention is given to handling data imbalances.

Methodology and Analytical Tools

Modeling and Analysis

- **Individual Performance Prediction**
  - **Models Used**: Random Forest Classifier, AdaBoost Classifier, Support Vector Classifier (SVC), K-Neighbors Classifier, Decision Tree Classifier, Gaussian Naïve Bayes, and Neural Networks.
  - **Hyperparameter Tuning**: Employing Grid Search to find optimal model parameters.

- **Team Selection Optimization**
  - **Algorithms**: Randomized Heuristics, Stochastic Control, Tabu Search, Variable Neighborhood Search (VNS), and Brute Force.
  - **Data Sets**: Two types of data sets used—one focusing on all-around scores and another on individual apparatus scores.

Expected Challenges

- **Data Quality and Completeness**
- **Handling Data Imbalances**
- **Complex Model Selection and Optimization**

Evaluation

- **Performance Metrics**: Accuracy, Precision, Recall, F1-Score, and F2-Score.
- **Validation Techniques**: Holdout validation and k-folds cross-validation.
- **Confusion Matrix**: Used for detailed error analysis of model predictions.
- **Hyperparameter Tuning**: Employing Grid Search to optimize model parameters.
- **Algorithm Evaluation**: Assessing the effectiveness of optimization algorithms like Tabu Search and Variable Neighborhood Search in team selection based on their ability to maximize scores.

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.

Conclusion