ML Project Weekly Report

Course Name: Machine Learning (CSE 523)
Week Number: 3 (10th - 16th February 2024)
Group Name: White
Instructor's Name: Mehul S Raval
Project no.: 7

Project 7 | Athlete Statistics Visualization and Prediction

Introduction

In the week of February 1st to 3rd, our team dived into the world of literature to better understand athlete performance analysis. This report covers our activities, challenges, and insights gathered from our literature review.

Weekly Activities

1. Literature Review:

Reviewed two significant research paper suggested by the faculty:

- First paper: "Impact of Sleep and Training on Game Performance and Injury in Division-1 Women's Basketball Amidst the Pandemic" published in IEEE Access, 2022.
- Second paper: "A Holistic Approach to Performance Prediction in Collegiate Athletics: Player, Team, and Conference Perspectives" (2023).

2. Planning and Discussion:

We discussed the project's objectives, and identified challenges, including dimensional reduction and player performance prediction, and planned the initial steps, which involve feature analysis and exploration.

Ahmedabad University

Challenges Identified

1. Dimensional Reduction:

- Addressing the complexity of the dataset with 40 features and the need to reduce it to 5 modalities posed a significant challenge.
- Exploring techniques like feature clustering and model fit-based weighing to address this challenge effectively.

2. Player Performance Prediction:

- Developing accurate models to predict performance scores for each modality based on the original dataset.
- Balancing the predictive accuracy while ensuring the models' interpretability and usability.

Learnings

1. Insights from Literature:

a) For the First Paper

The first paper taught us about the common algorithms used for similar problem solving tasks, including classification, regression, and clustering. We also recognized the importance of correlation-driven features in our analysis. The study revealed significant correlations between sleep patterns, training intensity, and game performance in Division-1 women's basketball, emphasizing the interconnectedness of these factors in athletic performance.

b) For the Second Paper

We discovered the significance of introducing adaptiveness into classification boundaries to address challenging instances. We also learned that synthetic sample generation from the minority class may not fully account for data variability; instead, a combination of over and undersampling techniques works best. Furthermore, the paper highlighted the predictive value of cardiac rhythm analysis for athlete readiness and injury prevention.

2. Data Preprocessing Techniques:

- Explored initial data preprocessing methods to handle missing values, outliers, and feature engineering.

Ahmedabad University

Conclusion

This week, we learned a lot and got a clearer picture of what we need to do next. We studied relevant research papers and found out about common algorithms, the importance of features that show connections between things, and how analyzing heart rhythm can help predict performance. We also started working on getting the data ready for analysis, which involves fixing any missing information, dealing with outliers, and potentially changing how the data is organized. As we move forward, we're ready to tackle the challenges of reducing the number of features and predicting player performance.

References

Senbel, S., Sharma, S., Raval, M. S., Taber, C., Nolan, J., Artan, N. S., ... & Kaya, T. (2022). Impact of sleep and training on game performance and injury in division-1 women's Basketball Amidst the Pandemic. *Ieee Access*, *10*, 15516-15527.

Taber, C. B., Sharma, S., Raval, M. S., Senbel, S., Keefe, A., Shah, J., ... & Kaya, T. (2024). A holistic approach to performance prediction in collegiate athletics: player, team, and conference perspectives. *Scientific Reports*, *14*(1), 1162.