Ahmedabad University

ML Project Weekly Report

Course Name: Machine Learning (CSE 523)
Week Number: 1 (1st - 3rd February 2024)
Group Name: White
Instructor's Name: Mehul S Raval
Project no.: 7

Project 7 | Athlete Statistics Visualization and Prediction

Introduction

This is our first report for the week of February 1st to 3rd. During this time, we started working on our project titled "Athlete Statistics Visualization and Prediction" This report details the initial progress of our project, and we are still trying to figure out how to leverage machine learning algorithms to analyze a diverse dataset of Division I basketball players. The primary goal is to reduce the 40 features into 5 modalities, including sleep, training, cardiac rhythm, jump, and cognitive aspects, and develop models to predict scores for each modality.

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.

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.

Conclusion

The first week of the project provided valuable insights into the project's scope, challenges, and initial steps required for upcoming weeks implementation of the project. The initial data exploration highlighted the need for feature engineering and dimensionality reduction techniques.

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.