**CSE 523 Machine Learning**

**Section 1**

**Group Name:** Decision Makers

**Project Number: 6**

**Athlete profiling based on similar characteristics**

**Problem Statement:**  
We have a multi-modal dataset of division I basketball players. This includes their sleep pattern, training details, cardiac rhythm pattern, emotional-mental state information, game score, weekly readiness scores and jump-data (RSImod). You need to apply machine learning algorithms over the comprehensive dataset to group the athletes into clusters (based on similar characteristics). The expected outcome is XAI explained characteristics of each group - most significant feature values and their impact on their weekly readiness score (RSImod).

**Literature References:**

| **Paper title** | **Authors** |
| --- | --- |
| A holistic approach to performance prediction in collegiate athletics: Player, team, and conference perspectives. Scientific Reports, 14(1), 1-10. [link](https://doi.org/10.1038/s41598-024-51658-8) | Taber, C. B., Sharma, S., Raval, M. S., Senbel, S., Keefe, A., Shah, J., Patterson, E., Nolan, J., Sertac Artan, N., & Kaya, T. **(2024).** |
| Impact of sleep and training on game performance and injury in Division-1 women’s basketball amidst the pandemic. [link](https://ieeexplore.ieee.org/document/9690164) | Senbel, S., Sharma, S., Raval, M. S., Taber, C., Nolan, J., Artan, N. S., & Kaya, T. **(2022).** |

**Weekly Report**

Our group started off with reviewing the existing research papers available and shared with us. The core focus of the research papers was to predict the readiness of players at individual, team and conference levels by employing robust measures to quantify the players’ performances. Therefore, our primary task during this time was to understand and explore various methods deployed in the papers like the XG Classifier and random forest for parameter optimizations on features, and algorithms like MICE for data imputation.

Further, we explored the data for our project and planned techniques to derive an approach for our problem definition. We deliberated on various clustering algorithms for our problem definition like K-Mode Clustering, Mean Shift and Gaussian Mixture Models. We decided to use Gaussian Mixture Models considering our dataset.

Our next goal is to develop a model and test it on the dataset, however we are yet to choose a method to clean and organize our data.