



**Ahmedabad  
University**

## **CSE - 523 Machine Learning**

### **Weekly Report-1**

Project title: Athlete performance in collegiate basketball: Predicting match  
line-up and RSImod

Submitted to faculty: Mehul Raval

Date of Submission: 9-02-2024

Enrolment No.	Student Name	Programme
AU2140031	Priyanshu Thakkar	BTech
AU2140001	Adnan Kadiwala	BTech
AU2140158	Milee Bajpai	BTech
AU2140017	Vandit Shah	BTech

- **Overview of the Problem Statement:**

This ML project wants to create a novel approach to forming the best college basketball lineups that will be able to predict the most effective team composition and thus increase the probability of a win. Usually, it was the head coach's decision, or it was based on subjective judgments and manual performance analyses that were very time-consuming and afflicted with bias. Exploiting ML, this project employs player data and historical information to display coaches with data-driven suggestions, helping team and individual strengths to be exploited in a strategic way. The idea behind this approach is not only to help the selection process run quicker but also to increase the team's success and performance rates by implementing a data-driven approach for the team formation decision. The primary goal behind this project is to ensure that coaches have access to a highly advanced decision-making instrument that is a perfect blend of efficiency and precision and, thus, alter the conventional approach to lineup selection on a collegiate basketball team.

- **How is our Project Different from the base papers?**

This base paper is about the use of machine learning methods for college basketball performance prediction and analysis, with player lineup optimization and injury prevention as the focus. It uses a correlation-based approach to determine the key factors and drop redundant data during detailed analysis. Routines such as XGBoost and Random Forest are applied to determine feature importance and predictions based on decision trees and ensemble methods. Moreover, k-means clustering groups the match performances and the injury cases, making the prediction model more precise for both. Predictive modeling, mainly based on XGBoost, targets precisely forecasting the player performance metrics of RSImod (reactive strength index modified), GS (game score), and PER (Player Efficiency Rating). Through integrating these advanced analytical models, the study deduces effective line-up selection and injury risk management and provides a comprehensive solution for coaches and sports scientists to maximize athlete performance and strategy.

In our project on athlete performance in college basketball, we're using machine learning to ensure that our team decisions and performance are optimal. We're honing in on two critical aspects: RSImod is a tool that helps predict a player's performance and the formation of the best lineup for each game. We use RSImod scores combined with diverse player statistics to determine who is likely to play well. At the same time, we evaluate past performances, player abilities, and opponent strengths in order to design a model, which is used by the coach to create

a lineup that maximizes the possibility of winning. RSImod predictions, when linked to lineup forecasts, give coaches the capability to make informed decisions that lead to the achievement of maximum potential by the team and maintaining a competitive edge. The constant refinement of our models by us makes it possible for the models to perform even better and to adapt to changing conditions over time.

The first paragraph, which introduces the aim and methodologies of research or a project that concentrates on the machine learning approach applied in college basketball performance analysis, It presents the different strategies used, including using correlations for analysis, ensemble modeling with XGBoost and Random Forest, and clustering techniques, to make decisions about the player lineup and keep players off the wrong side of the injury list. The main point is the supply of data on forming lineups and planning injuries for coaches and sports scientists.

In the second paragraph, we are going to add something like that: "A hypothetical project of ours is all about optimizing athlete performance in college basketball with the help of machine learning." The project's aim is underlined in this paragraph and includes determining the utility of RSImod in projecting player performance and establishing the best lineup for each game using the model. The focus is on matching the RSImod scores and player stats to uncover future top performers and then crafting lineup models out of the past stats of top players as well as the key points of the opposing teams. The aim is to give coaches access to data-driven insights for making strategic decisions and staying competitive at the same time.

- **Literature Survey Summary:**

We made the survey focusing on the paper "Prediction of Game Performance and Injury in Division-1 Women's Basketball during the Pandemic Using Multiple Imputation by Chained Equations (MICE), Random Forest (RF), Autocorrelation, and XGBoost." It was a study where the author used the MICE method to predict game performance and injury, identifying key features through RF analysis and autocorrelation, and then using

- **References:**

1. Impact of sleep and training on game performance and injury in division-I women's Basketball Amidst the Pandemic S Senbel, S Sharma, MS Raval, C Taber, J Nolan.. - ieee Access, 2022
2. <https://www.nature.com/articles/s41598-024-51658-8>