

CSE523 Machine Learning

Weekly Report 1

**Athlete Profiling for Division I Basketball Players**

Submitted to faculty: Mehul Raval

Date of Submission: 02-03-24

Student Details

| Roll No. | Name of the Student |
| --- | --- |
| AU2140024 | Aagam Shah |
| AU2140133 | Dhairya Shah |
| AU2140159 | Aayushi Shah |
| AU2140182 | Aanal Dobariya |

2023-2024 (Winter Semester)

### **Aim:**

Understanding the Problem Statement and Literature Review

### **Introduction:**

The paper uses 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). The task is to apply machine learning algorithms over the comprehensive dataset to group the athletes into clusters (based on similar characteristics). The expected outcome is that XAI explained the characteristics of each group - the most significant feature values and their impact on their weekly readiness score (RSImod).

### **Work Completed:**

* The papers contained K Nearest Neighbors, Decision tree, Random Forest techniques to obtain ranked importance of features and defined key performance indicators (KPIs) at each level to address the three-tiered pyramid. We studied the aforementioned models used in the literature, and how they affected defined key performance. (KPIs)
* We researched different clustering techniques in analyzing basketball player data. It helped us to provide valuable insights for analyzing and choosing our approach for the given problem.

### **Next steps and goals:**

* Analyzing the dataset and preprocessing and cleaning the data.

### **Conclusion:**

After an initial literature review, we are inclined toward utilizing the Random Forest approach for addressing the clustering problem in basketball player data. The versatility and robustness of Random Forests make them particularly suitable for handling complex datasets with multiple variables, which is characteristic of basketball player statistics.

We will also utilize a multiple imputation approach called Multivariate Imputation by Chained Equation (MICE) to impute the data (the approach used by the paper(s)).

### **References:**

1. 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.

<https://ieeexplore.ieee.org/abstract/document/9690164/>

1. Taber, C.B., Sharma, S., Raval, M.S. et al. A holistic approach to performance prediction in collegiate athletics: player, team, and conference perspectives. Sci Rep 14, 1162 (2024). <https://doi.org/10.1038/s41598-024-51658-8>
2. Sharma, S., Divakaran, S., Kaya, T., Taber, C., & Raval, M. S. (2023, October). A Framework for Biomechanical Analysis of Jump Landings for Injury Risk Assessment. In 2023 IEEE 28th Pacific Rim International Symposium on Dependable Computing (PRDC) (pp. 327-331). IEEE. <https://ieeexplore.ieee.org/abstract/document/10356423/?casa_token=yM-1J2vMKG8AAAAA:QnMMyph1mKa6BfvljAqSSABrReChb5hvsEvxFUEv3z5Fiiv1daHpTdbG2oXllJYQ-WQPUijJ6wZr>
3. Sharma, S. U., Divakaran, S., Kaya, T., & Raval, M. (2022, July). A Hybrid Approach for Interpretable Game Performance Prediction in Basketball. In 2022 International Joint Conference on Neural Networks (IJCNN) (pp. 01-08). IEEE. <https://ieeexplore.ieee.org/abstract/document/9892583/?casa_token=Ye3GQJ1JpD0AAAAA:GSlyds24pIa__7Od6UBSNs8nugbEwLvCbI8vG6w-YMYrFX2O-TUlUDo4xej3ulJUAvmO_4ij0J36>
4. K means Clustering: GfG. (2023, December 21). *K means clustering - introduction*. GeeksforGeeks. <https://www.geeksforgeeks.org/k-means-clustering-introduction/>
5. Decision Tree: G. (2023, August 20). *Decision Tree*. GeeksforGeeks. <https://www.geeksforgeeks.org/decision-tree/>
6. G. (2023, December 6). *Random Forest Regression in Python*. GeeksforGeeks. <https://www.geeksforgeeks.org/random-forest-regression-in-python/>
7. G. (2022, June 1). *Bagging vs Boosting in Machine Learning*. GeeksforGeeks. <https://www.geeksforgeeks.org/bagging-vs-boosting-in-machine-learning/>