

CSE523 Machine Learning

Weekly Report 4

**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:**

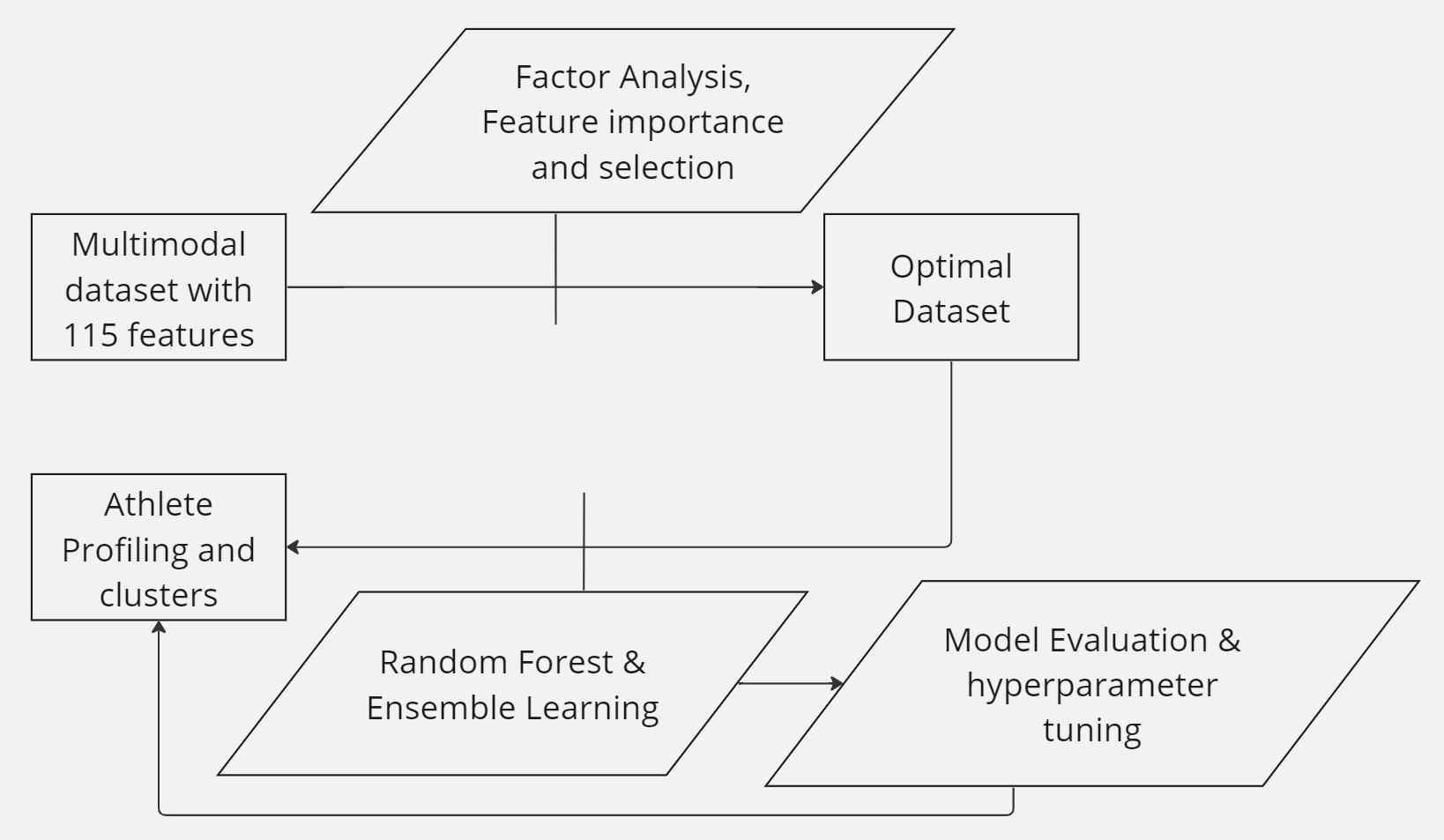
To formule an approach for athlete profiling and clustering

### **Introduction:**

We formulated a customized strategy to reveal implicit trends and derive significant data by utilizing correlation heat maps and descriptive statistics, which gave us important information about feature interdependencies. The present strategic study paves the way for subsequent investigations, such as factor analysis and feature engineering, which can yield new understandings of the fundamental structure and interdependencies of the features.

### **Work Completed:**

* Approach



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

* Factor analysis
* Feature engineering (Feature clustering)

### **Conclusion:**

The approach of clustering and optimizing datasets through feature engineering and factor analysis, coupled with the random forest algorithm. This presents a robust methodology for data analysis and prediction. By using feature engineering and factor analysis, we would effectively extract meaningful patterns and reduce dimensionality, enhancing the quality of our dataset. Adopting the random forest algorithm methodology, accurate model evaluation, and efficient hyperparameter tuning improves predictive performance.

### **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. 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>
3. K means Clustering: GfG. (2023, December 21). *K means clustering - introduction*. GeeksforGeeks. <https://www.geeksforgeeks.org/k-means-clustering-introduction/>
4. Decision Tree: G. (2023, August 20). *Decision Tree*. GeeksforGeeks. <https://www.geeksforgeeks.org/decision-tree/>
5. G. (2023, December 6). *Random Forest Regression in Python*. GeeksforGeeks. <https://www.geeksforgeeks.org/random-forest-regression-in-python/>
6. G. (2022, June 1). *Bagging vs Boosting in Machine Learning*. GeeksforGeeks. <https://www.geeksforgeeks.org/bagging-vs-boosting-in-machine-learning/>