

CSE523 Machine Learning

Weekly Report 8

**Athlete Profiling for Division I Basketball Players**

Submitted to faculty: Mehul Raval

Date of Submission: 06-04-24

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:**

Improve athlete profiling and explain the created clusters using factor analysis and feature importance.

### **Introduction:**

As discussed in the previous report, we have created clusters using different features and achieved a maximum of 0.4362 silhouette score. We have imputed both season 2 and season 3 data individually. Then, grouped all the athletes' time series data, averaged them, and did factor analysis. By doing factor analysis with the help of taking different feature combinations with different weights, we got the silhouette score of 0.7038. After that, we used the bootstrapped sampling technique to reach a score of 0.7466. We tried to determine the importance of the feature using PCA to justify our results.

### **Work Completed:**

* Factor analysis

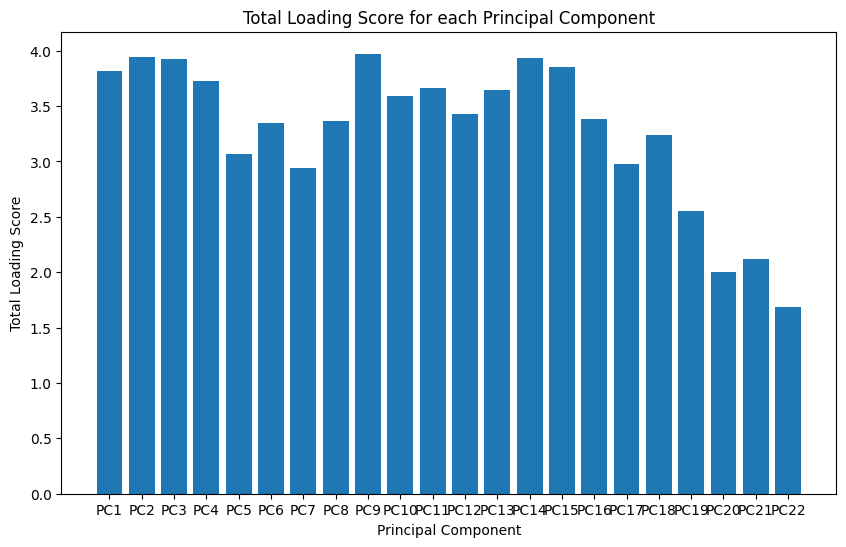
By trying different feature combinations exploiting the multimodality of the dataset and applying KMeans clustering to it, we found the following feature results:

sleep\_weights = {'Sleep.Disturbances': 1}

health\_weights = {'Awake..hours.':0.60, 'Deep.Sleep..hours.':0.11, 'Sleep.Need':0.1, 'RSI Mean': 0.30}

Silhouette Score: 0.7038355432078389

* Feature importance



* Bootstrapped Sampling:

Bootstrapping is a resampling technique where new data samples are generated by sampling from the original dataset with replacement. It increased the dataset size and improved the silhouette score to 0.7466

### **Next Steps and Goals:**

* Preparing Endsem report and working on XAI

**Conclusion:**

Combinations of features provide better results and detailed insights into combinations of features. Furthermore, applying the sampling technique increased the data size and helped improve clustering and athlete profiling. The feature score generated by the PCA algorithm helped obtain the XAI component of the clustering technique.

**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. "ML | K-Means Algorithm," GeeksforGeeks, Last updated: Mar. 21, 2024. [Online]. Available: <https://www.geeksforgeeks.org/ml-k-means-algorithm/>. [Accessed: Mar. 30, 2024].