

CSE523 Machine Learning

Weekly Report 6

**Athlete Profiling for Division I Basketball Players**

Submitted to faculty: Mehul Raval

Date of Submission: 23-03-24

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 include more features to form clusters and check if the silhouette score increases

### **Introduction:**

RSI mean and HRV were the primary two features we dealt with. After using three clustering techniques, we chose K-means as it had a higher silhouette score. We attempted to incorporate more features based on feature relevance to see whether we might obtain better clusters.

### **Work Completed:**

1. Clustering Techniques and Silhouette Score:

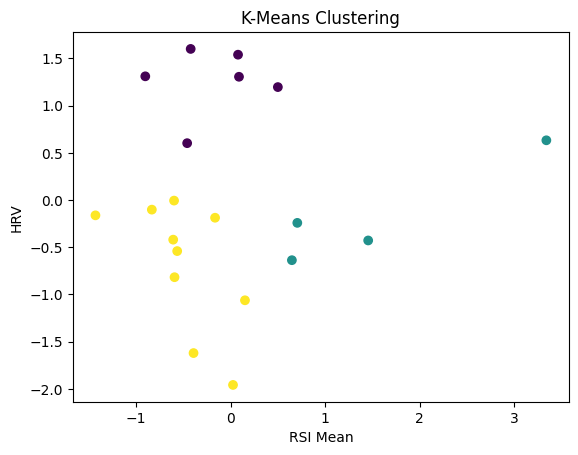
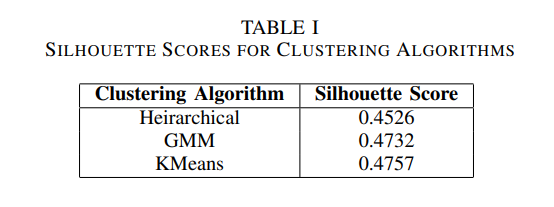
 \

Table I summarizes the clustering findings for athlete profiling using KMeans, Gaussian mixture model (GMM), and hierarchical clustering. Each clustering technique generated a silhouette score. This score indicates the clustering quality, with higher values denoting better-defined clusters.

1. Incorporating New features

We added features based on the correlation matrix, identifying traits highly correlated with the target variable. We added RHR, Recovery, and Sleep. scores along with the RSI mean and HRV. As the data for the other three features was available for all days, we calculated the weekly average and placed it on Monday. The processed dataset contained data values for Mondays.

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

We incorporated the new features in addition to RSI mean, and HRV by taking the weekly average of those features and putting it in Monday, similar to what we did for HRV. If, for certain features, no values existed for an entire week, then we used the MICE imputation technique to impute the weekly average for that week, but instead, what we can do is first impute the values and then calculate the weekly average to see if it provides us with better results.

We will also use more features, not just 5, and see if the K-means clustering model will give us better results. And also incorporate more explainable AI techniques.

### **Conclusion:**

In addition to using HRV and RSI Mean and RHR, Recovery, and Sleep Score features to cluster the athletes. However, the silhouette score did not improve. Instead, it decreased to 0.3657 for K-means clustering.

This was also true for hierarchical clustering and Gaussian mixture models. So, we either try a trial and error approach in selecting different features to get better results or try better imputation methods on the dataset.

### **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/>