

CSE523 Machine Learning

Weekly Report 7

**Athlete Profiling for Division I Basketball Players**

Submitted to faculty: Mehul Raval

Date of Submission: 30-03-24

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 try a trial and error approach in selecting different features to get a better silhouette score and clustering.

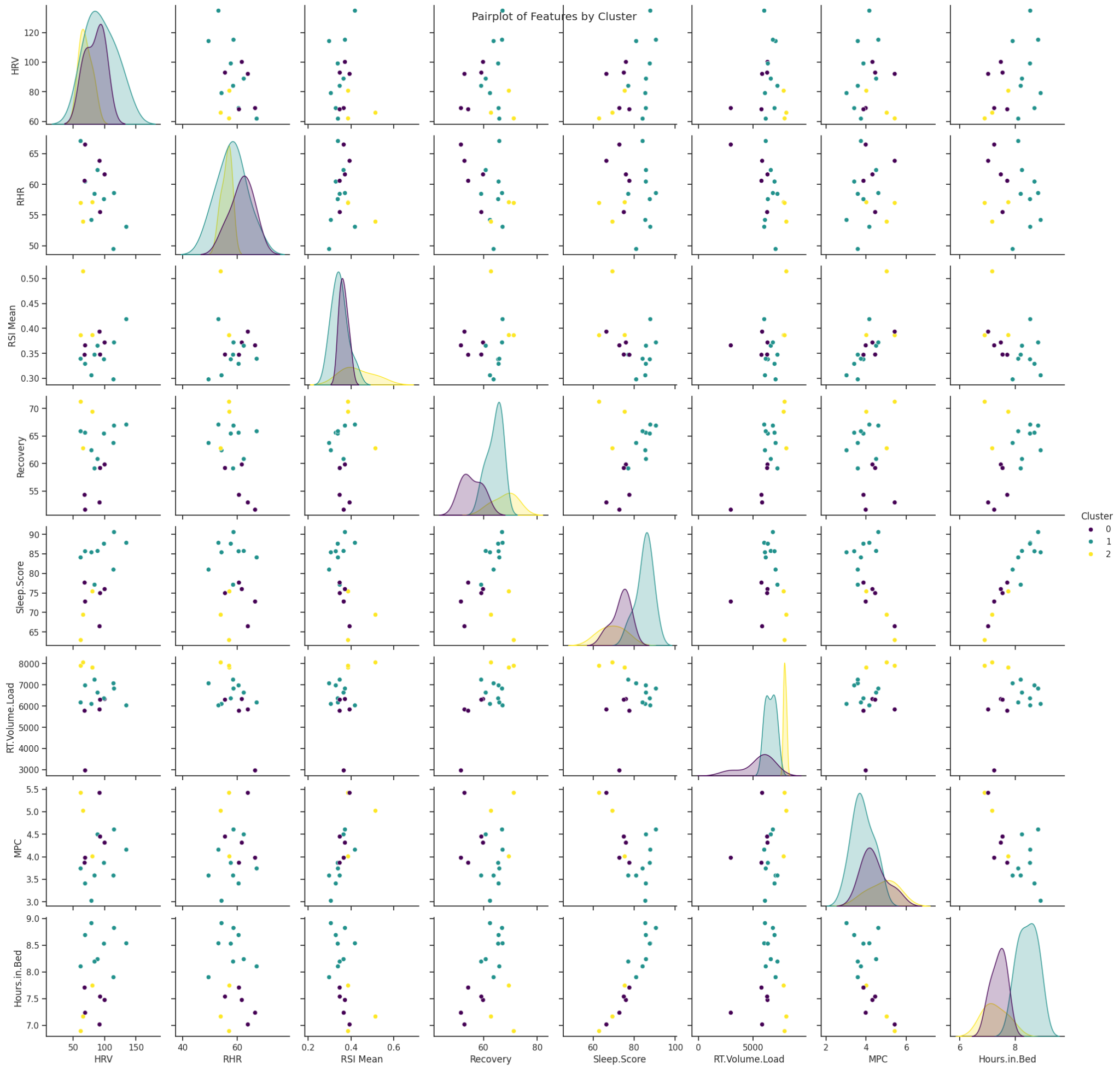
### **Introduction:**

In the beginning, we formed clusters using two crucial features: RSI mean and HRV. We included additional features depending on their significance and relevance to the target variable to improve the findings. We added EB, Weekly.SD, Strain, RT.Volume.Load, Sleep.Need, Sleep.Consistency, Sleep.Disturbances, Respiratory.Rate, RHR, REM.Sleep..hours. and Deep.Sleep..hours to test the outcomes and evaluate the silhouette score.

### **Work Completed:**

* K-Means Clustering:

First, we used the elbow method to find an optimal number of clusters but failed to find an elbow point. We chose k=3 for moving ahead based on the visualization of the data points. Then, we further performed k-means clustering using k-means++ initializations for k-centroids. The following are the clustering results.



* Evaluation of clusters

To evaluate the model, we found the silhouette score for k-means clustering and found silhouette score and inertia as follows:  
Silhouette Score: 0.23258486746372234

Inertia: 70.15563442120009

We will again do preprocessing and feature engineering to improve our silhouette score with the least possible loss of information.

* Back to feature engineering for better results
  + Previously, only two features were selected for clustering, this time we have done preprocessing on 10 significant independent features and used MICE imputation to improve the silhouette score.

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

* Integrating the XAI component to justify and define the clusters and allotting them appropriate labels.
* Building a dashboard for Athlete profiling to better visualize, understand, and provide insights on the clustering.

**Conclusion:**

We tried incorporating more features to do a multimodal clustering, but the silhouette score decreased. We previously got silhouette score as Silhouette Score = 0.43625266469807833 and Inertia = 13.474859108927676. This depicts the importance of RSI mean and HRV for better clustering. We will proceed further with these two variables and will verify the same using XAI.

**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].