

CSE523 Machine Learning

Weekly Report 2

**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:**

Data Cleaning and processing (imputation)

### **Introduction:**

The provided dataset includes various basketball-related features for each player, such as Sleep Score, Recovery, Body Weight, Jump Height, etc. However, due to the impact of the COVID-19 pandemic, numerous entries for these features still need to be included, particularly for Season 3. Consequently, a significant portion of the dataset contains NA or null values. To effectively utilize the dataset, we must decide whether to exclude the entire feature with missing values or employ imputation techniques to fill the null cells for the features.

### **Work Completed:**

* The dataset we were given contained many different sheets and files. So, to use the data, we converted and merged all of the data into a single CSV file.
* We used the Multiple Imputation by Chained Equation (MICE) technique to impute the dataset.

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

* To visualize the data and form our methodology.

### **Conclusion:**

We first converted the dataset into a singular CSV file. We spent quite a significant amount of time processing data because we had to think about whether or not this feature was important enough to keep or whether we could drop it. If we impute a cell, should we do it with a mean or use another technique to fill it?

Link to the collab : <https://colab.research.google.com/drive/1WMOik8eT5PZ26DLBYIn0FNncvzSAK5GA?usp=sharing>

### 

### 

### **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>