

CSE 523 ML Project

Faculty mentor:
Prof. Mehul Raval

Project Number: 9

Project Title: Data-driven imputation scheme for human-subject-based dataset

Group name: Syntellect

Group member details:

ENROLLMENT NUMBER	NAME
Kashish Jethmalani	AU2140029
Srushti Thakar	AU2140117
Priyal Patel	AU2140204
Riya Patel	AU2140214

Progress report : (Week 1)

Date: 1-02-2024 to 8-02-2024

Understanding the problem statement:

This project aims to develop a model for imputing missing values in a multi-modal dataset of Division I women's basketball players, focusing on sleep patterns, training details, cardiac rhythm patterns, emotional-mental state, game scores, readiness scores, and jump data. The model will be compared against state-of-the-art imputers like MICE. The study's findings will contribute to understanding the impact of sleep and training on game performance and injury, particularly in the context of the pandemic.

Understanding the dataset:

- The dataset includes Season 2 with Polar.csv, Season 3.csv, Vertical Jump Season 2.xlsx, and Vertical Jump Season 3.csv
- Season 2 with Polar.csv contains sleep, training, questionnaires, and Polar data for season 2.
- Season 3.csv contains similar data for season 3.
- Vertical Jump Season 2.xlsx contains RSI-related data for season 2, with multiple sheets.
- Vertical Jump Season 3.csv contains RSI-related data for season 3.

Data Conversion: We need to convert Vertical Jump Season 2.xlsx into a format similar to Vertical Jump Season 3.csv

Data Merging: We will then merge Vertical Jump Season 2.csv into Season 2 with Polar.csv using "date" as the merge key.

- This step combines all modalities of data for each athlete date-wise. Be cautious with athlete names to avoid errors in merging.

Preprocessing and Analysis: After merging, we will perform the preprocessing steps mentioned in the papers provided.

- We will refer to the papers for information on features to be kept, collinearity issues, and further analysis techniques.
- We will also consider dropping features that exhibit collinearity to avoid redundancy in the analysis.

Finally, we will use the merged CSV files for analysis and modeling.

Papers:

<https://ieeexplore.ieee.org/document/9690164>

<https://www.nature.com/articles/s41598-024-51658-8>