# *Sleep and Physical Performance Analysis: A case study of Collegiate Women's Basketball Players*

## *Data Imputations*

Srishti Sharma
School of Engineering and Applied Sciences
Ahmedabad University
Ahmedabad, India
srishti.s1@ahduni.edu.in

***Abstract*—Over the decade, there has been a considerable amount of research done in analyzing the impact of sleep and exercise on the performance and recovery of an athlete. The *long term objective* of this research is to make use of the Machine Learning paradigm to create a predictive model that would help a coach determine whether a player is fit to play in the forthcoming match on the basis of the underlying patterns and correlations between sleep, training load, cognitive state information of the athlete and their performance. In this term, the objective in this project duration is to perform necessary data imputations as a part of preprocessing of data.**

***Keywords—Sports Data Analytics, Basketball, Sleep, Exercise, Injury, Recovery, Data Imputations***

## I. INTRODUCTION

An athlete undergoes both cognitive and physical strain in their life due to the training load imposed upon them, their emotional and physical states, academic routine or regular routine strain, sleep etc. The performance of an athlete is highly dependent upon the quality and quantity of sleep an athlete gets. Sleep as well as exercise has a profound impact on the recovery time for an athlete from an injury.

The research conducted in this domain so far has been concluded on the basis of statistical models or on the basis of correlation drawn between the measured parameters. Issue with statistical models is that they try to fit a generalized equation emphasizing less on the underlying individual patterns. There is no holistic research done for example there is a statistical study of impact of sleep on shooting accuracy of a player or impact of cardiac rhythm on athletic performance. There is a need of a holistic research done taking into account all training load, cognitive stress (emotional and physical state), sleep impacts, physiological recordings in routine etc. There is no research carried out to study the impacts of sleep and exercise on the recovery time from an injury for an athlete.

The aim is examination of the interaction of sleep, stress and training load of athlete and its impact on performance and recovery. The problem statement will be tested and validated over the data collected for 17 women's basketball team players. The data will be gathered by tracking their sleep using a Whoop wearable strap (sleep monitoring & physiological markers), recording their training load by the strength coaches and the athletes and a short recovery and short stress questionnaire that would help evaluate their emotional and physical state.

The data collection period began during the final three weeks of offseason training and continued through a six-week preseason training schedule. During this time the athletes wore the WHOOP straps continuously during the collection period allowing for consistent monitoring of data throughout the day.

We experiment with a few data imputation techniques based on global and local statistics to the missing values in the dataset. Substitution using global mean value, analysis of the most dominant attribute over target attribute, imputation using k-means clustering based on dominant attribute, k-NN algorithm for imputations are experimented and the results are presented in section 3. Section describes a review of existing literature. Section also includes the details about the dataset and imputation techniques. Section 4 concludes the paper.

## II. LITERATURE REVIEW

Sleep has a profound effect on the athletic as well as cognitive performance of the athletes. It is believed that a good sleep helps the athletes recover fast from the cognitive load they undergo [1]. Also, sleep plays an important role in helping the athletes recover from the routine strain thereby improving their performance such as their response time [2].

A study of the soccer players [3] suggested how a degradation in the athletic performance and recovery was directly related to the consumption of substances such as caffeine, alcohol, deprivation of the right quality of the right quantity of sleep, travel fatigue etc. The impact of cardiac rhythm on the athletic performance was studied by analyzing the impacts of deprived sleep due to night soccer matches [2]. It was reported that a deprivation of sleep for 36 long hours would lead to degradation of tolerance to a long duration of exercise of training load [4]. It was also concluded that sleep extension resulted in a better shooting rate and accuracy of basketball players [5].

A week-long sleep deprivation and irregular sleep habits would lead to degradation in attention and alertness of the players [6]. As per recommendations by the National Sleep Foundation, wearable devices were the most appropriate ones to measure the sleep activity or the heart rate measurements of players [7].

## III. EXPERIMENT AND RESULTS

### *A. Dataset*

To study the impacts of sleep and exercise on the athletic performance and recovery from injuries of the

players, an experiment is to be conducted over 17 women's basketball team players from Sacred Heart University. For all these players, their information related to sleep, training load and cognitive stress is recorded using the following:

1. Sleep Monitoring using a WHOOP Strap – This is a wearable devices that helps track the sleep and recovery patterns of the player. It records the data in three categories: attributes dependent on the cardiovascular strain and exertion, other one depending on the resting heart rate and the sleep hours, sleep consistency, sleep score etc. and the last category is related to the physiological markers such as heart rate, respiratory rate, sleep etc.
2. Short Recovery Short Stress Questionnaire – A set of eight questions on emotional and physical state of the athletes.
3. Training Load – Time and the intensity of the metabolic activities of the athletes is computed in the form a score every week and is recorded in the dataset.

The data collection period began during the final three weeks of offseason training and continued through a six-week preseason training schedule. During this time the athletes wore the WHOOP straps continuously during the collection period allowing for consistent monitoring of data throughout the day.

### *B. Discussion*

The data gathered from these three sources is then to be preprocessed which include imputing the missing values that arise due to any fault with the device. Next step is the reduction of dimensionality of the data as the numbers of attributes taken into consideration are too many and hence extraction of the most relevant attributes from the attribute set is an important step. The final phase will be studying the impact of sleep and exercise on the athletic performance of the athletes as well as their recovery from injuries. This predictive analysis will thereby help the coach determine whether a player will be able to play well in a forthcoming match.

As a part of the project, we used various techniques for imputing the missing values.

1. *Global mean substitution:* Each missing value of an attribute is substituted by the global average value of the entire attribute (column)

2. *Local mean substitution:* Techniques like k-means algorithm and k-NN algorithms were used for imputing values.

*a.* *K-means Clustering algorithm:* was applied over the entire dataset and then the mean of an attribute in each cluster was used to impute the missing values of that particular attribute in that particular cluster.

*b.* *Single attribute based K-means Clustering:* For filling missing values, K-means clustering is performed over the dataset fitting using the attribute whose values are to be imputed. Now for each missing value in cluster, the mean of that attribute values in cluster are averaged and substituted.

c. *Permutation based feature importance + Attribute based k-Means Clustering:* Fit a model for predicting target variable using independent variables and store the predictions. For each independent attribute, considering one attribute at a time, shuffle values of only that attribute rows and predict target variable using the fitted model. Calculate the loss suffered from shuffling. The independent variable for which there was maximum loss incurred due to shuffling is the most important attribute for particular target attribute considered. K-means is applied over the dataset fitting it on basis of the most dominant attribute as obtained from feature importance algorithm. Various combinations of dominant attributes and number of clusters were tried.

*d.* *K-NN Algorithm:* kNN algorithm was applied over the entire dataset. For a missing valued tuple, k nearest tuple were found and the missing tuple value was filled using the average value of that attribute values of all near neighbors found.

*e.* *Permutation based feature importance + Attribute based k-NN based Imputations:* kNN algorithm was applied over the entire dataset fitting using the most dominant attribute. For a missing valued tuple, k nearest tuple were found on basis of the most dominant attribute and the missing tuple value was filled using the average value of that attribute values of all near neighbors found.

### *C. Results*

The Dataset currently in consideration is the WHOOP strap dataset. This dataset has 9 weeks of athlete data related to 26 attributes. Some of these attributes have missing values. Out of all such attributes, results of permutation importance based feature importance algorithm applied for Wake Periods attribute are shown here:

| Weight | Feature |
|---|---|
| 65.1498 ± 8.5121 | Total Cycle Sleep Time (hours) |
| 31.4923 ± 3.6556 | Light Sleep (hours) |
| 21.3889 ± 3.2909 | Hours of Sleep |
| 14.7714 ± 2.2244 | Hours in Bed |
| 13.9012 ± 0.7925 | Total Cycle Nap Time (hours) |
| 13.8682 ± 3.1412 | Restorative Sleep (hours) |
| 9.4451 ± 0.7743 | Restorative Sleep (%) |
| 7.7635 ± 1.0626 | REM Percentage |
| 1.9439 ± 0.2195 | REM Sleep (hours) |
| 1.9303 ± 0.3693 | Deep Sleep Percentage |
| 1.2048 ± 0.1659 | Sleep Disturbances |
| 0.5300 ± 0.0858 | Deep Sleep (hours) |
| 0.4819 ± 0.1464 | Sleep Efficiency (%) |
| 0.0164 ± 0.0094 | Sleep Need |
| 0.0105 ± 0.0096 | Recovery |
| 0.0089 ± 0.0081 | Sleep Score |
| 0.0052 ± 0.0043 | HRV |
| 0.0039 ± 0.0086 | RHR |
| 0.0015 ± 0.0016 | Respiratory Rate |
| 0.0007 ± 0.0081 | Sleep Consistency |
| 0.0004 ± 0.0061 | Cycles |
| 0.0003 ± 0.0021 | Latency (min) |
| -0.0038 ± 0.0033 | Awake (hours) |
| -0.0066 ± 0.0086 | Sleep Debt (hours) |
| -0.0085 ± 0.0551 | Missing Data (hours) |

Figure 1: Important Features for Wake Periods attribute

For the Wake Periods attribute, each of the techniques discussed above were and their comparison is shown below in table 1 in form of column averages post substitution. Table 1 also contains average of original dataset that can be used as a benchmark.

| Parameters | Wake Periods Average |
|---|---|
| Original Dataset | 11.98 |
| Global mean | 11.98 |
| K-Means over entire dataset (6 clusters) | 11.98 |
| K-Means (6 clusters, 1 attributes) | 11.95 |
| K-Means (4 clusters, 1 attributes) | 11.95 |
| K-Means (10 clusters, 1 attributes) | 11.96 |
| K-Means (6 clusters, 2 attributes) | 11.94 |
| K-Means (6 clusters, 5 attributes) | 11.93 |
| K-Means (10 clusters, 5 attributes) | 11.93 |
| K-Means (10 clusters, 2 attributes) | 11.94 |
| K-NN over Entire Dataset (k=3) | 11.66 |
| K-NN Attribute based (k=3) | 12.02 |

Table1: Average Comparison between various techniques

## IV. Conclusion

Permutation based feature importance is an effective technique for finding the most dominant attribute for a given target attribute whose values are to be imputed. Data Imputation using Multivariate Regression using combination of important attributes selected using permutation based feature importance or recursive feature elimination has to be implemented ahead. Also, better metrics of comparison between the experimented techniques is to be worked upon.

## References


1. Vyazovskiy, V.V., 2015. Sleep, recovery, and metaregulation: explaining the benefits of sleep. Nature and science of sleep, 7, p.171.
2. Fullagar, H.H., Skorski, S., Duffield, R., Hammes, D., Coutts, A.J. and Meyer, T., 2015. Sleep and
3. athletic performance: the effects of sleep loss on exercise performance, and physiological and cognitive responses to exercise. Sports medicine, 45(2), pp.161-186.
4. Nédélec, M., Halson, S., Abaidia, A.E., Ahmaidi, S. and Dupont, G., 2015. Stress, sleep and recovery in elite soccer: a critical review of the literature. Sports Medicine, 45(10), pp.1387- 1400.
5. Martin, B.J., 1981. Effect of sleep deprivation on tolerance of prolonged exercise. European journal of applied physiology and occupational physiology, 47(4), pp.345-354.
6. Mah, C.D., Mah, K.E., Kezirian, E.J. and Dement, W.C., 2011. The effects of sleep extension on the athletic performance of collegiate basketball players. Sleep, 34(7), pp.943-950.
7. Al-Kandari, S., Alsalem, A., Al-Mutairi, S., Al-Lumai, D., Dawoud, A. and Moussa, M., 2017. Association between sleep hygiene awareness and practice with sleep quality among Kuwait University students. Sleep health, 3(5), pp.342-347.