Pipeline for fatigue-prediction package version 1

Welcome to the documentation of fatigue prediction project. Here, we aim to provide you with the latest updates and insights on the data transformations, feature construction and baseline modelling.

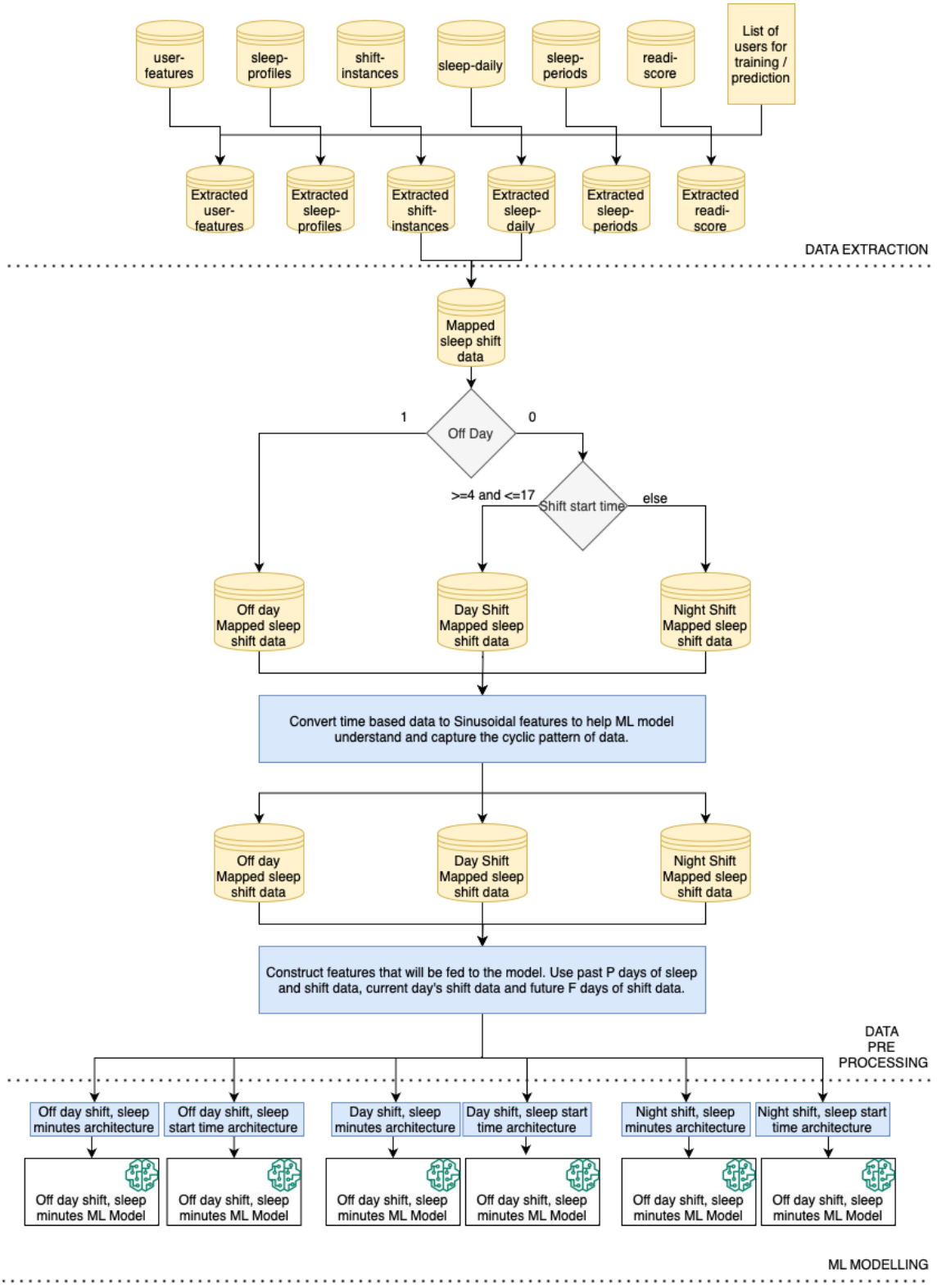
Aim: We aim to predict the sleep start time and sleep minutes & eventually calculate sleep end time as well using post processing.

Pipeline Architecture: The 6 raw datasets (user-features, sleep-profiles, sleep-periods, sleep-daily, shift-instances and ready-score) for the years 2020, 2021, 2022, and 2023 are used. Furthermore, a file is provided to enter the user IDs for which model training and predictions are to be made. For all the users provided in the file, the steps mentioned below are performed iteratively:

Training Pipeline:

The training pipeline is majorly divided into three steps: data extraction for the user, data preprocessing & feature construction, and model creation.

1. Data Extraction: For all six datasets, the data for each year will be combined with its respective dataset type, and subsequently, only the user's data will be extracted from each of these datasets.
2. Data pre processing: The data preprocessing step is further divided into 4 steps.
    a. Mapping sleep and shift data: The sleep and shift data are mapped in such a way that the next sleep start time occurs after the shift end time. This mapping ensures that each sleep duration corresponds to a particular shift and is not overlapping with the previous shift. The detailed information about mapping is available here, Sleep-Shift mapping.
    b. Divide data based on shift instance: With the off day feature provided, we can categorize the data into off days and working days, and then further subdivide them into day shift and night shift. If the off day feature is set to 1, it will represent an off day. On the other hand, if the feature has a value of 1 and the shift start time is between 4 and 17, it will be considered a day shift; otherwise, it will be classified as a night shift.
    c. Convert time data to sinusoidal features: Instead of representing time using conventional numeric labels (e.g., 1 to 24 for hours), sinusoidal features might be used, such as encoding hours as $\sin(2\pi * \text{hour}/24)$ and $\cos(2\pi * \text{hour}/24)$. This representation condenses time information into two continuous values that encapsulate the cyclic nature of the data. By converting time data to sinusoidal features, machine learning models can better understand and utilize the inherent cyclical patterns present in time-related data, leading to improved performance. Features like sleep start time, sleep end time, shift start time, shift end time, etc., will be converted.
    d. Feature construction: For each day, the training data will consist of the shift and sleep data for the past P days. Additionally, it will include the shift data for the current day and the shift data for the future F days. Before passing the features to the model, the features will be normalized.
3. Modelling: The modeling process is further divided into two parts. The first part involves determining the shift instance in which the user is currently working. The second part utilizes the model architecture for sleep minutes and sleep start time based on the determined shift instance.
    a. Sleep minutes model: The target variable will be the duration of sleep in minutes. The model architecture is customized to handle either the prediction of day shift sleep minutes or night shift sleep start time.
    b. Sleep start time model:The target variable will be sinusoidal value of sleep start time feature. The model architecture is customized to handle either the prediction of day shift start time or night shift sleep start time.

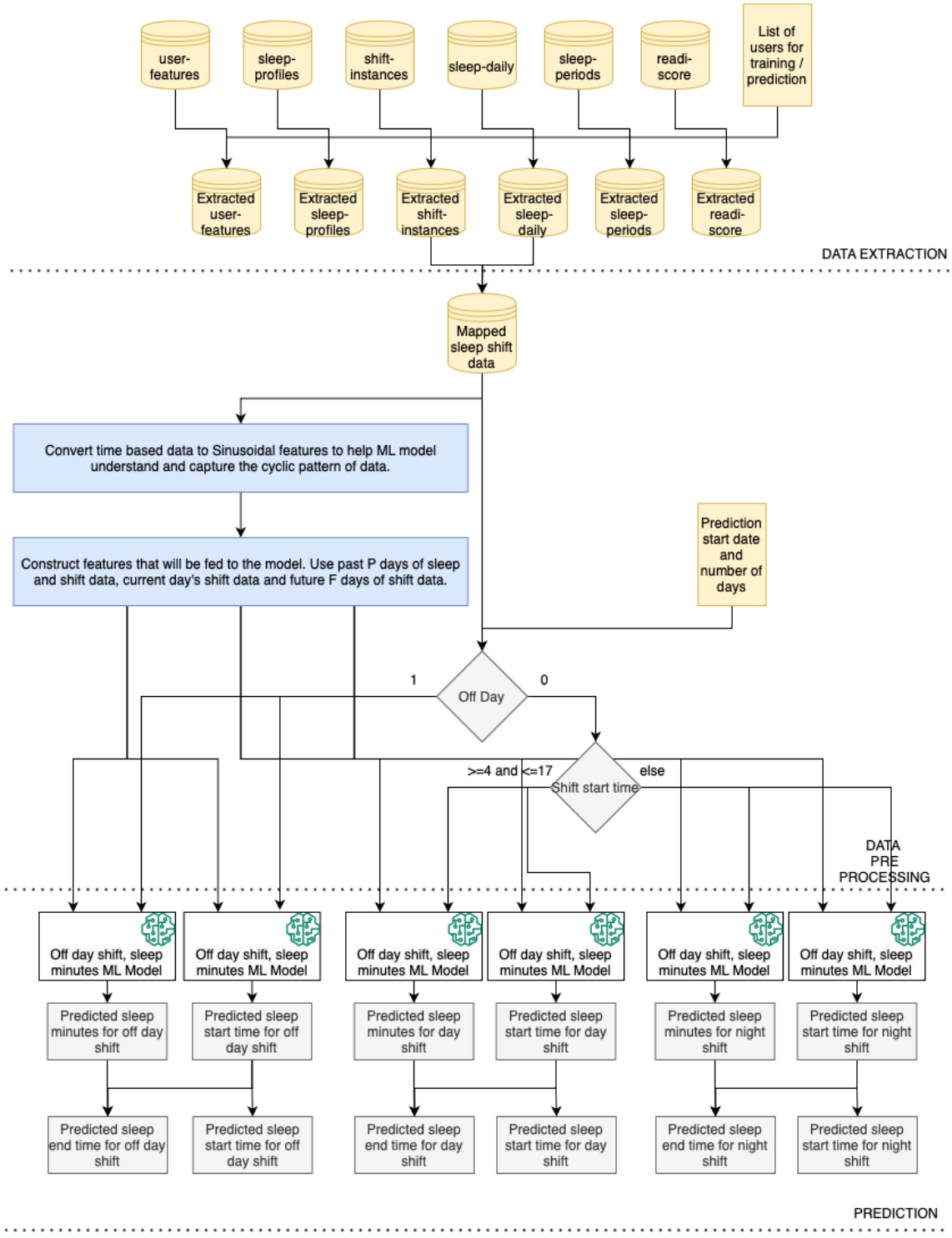


Training pipeline

Prediction Pipeline:

1. Extraction and preprocessing of data will follow the same procedures as explained in the training pipeline.
2. The prediction start date and the number of days to predict will be extracted from the file. The model will be used to predict both sleep minutes and sleep start time (represented as a sinusoidal value) for the specified day.

3. Post processing: The sleep minutes and sleep start time will be denormalized to retrieve their original values. Additionally, for the sleep start time, the sinusoidal values will be converted back to their original values, providing the sleep start time in minutes. Using both the sleep minutes and sleep start time, the sleep end time will also be determined.



Prediction Pipeline

Thanks for reading the documentation!