

Project Task: Predicting Sleep Quality Measures

Objective:

Predict one chosen sleep quality measure (rated 1-10) for each student based on data collected from cell phone sensors and daily sleep questionnaires. Students must select one target variable from the questionnaire ratings (sleep quality, restfulness, duration perception, morning energy, or ease of falling asleep).

General Instructions:

- Flexibility in Task Order: Students can change the order of the tasks if it makes sense and improves workflow.
- Output Requirement: Any outputs produced must be clearly documented and justified.
- Target Selection: Each student must clearly state which sleep measure they chose to predict and justify their choice.
- Clearly state which sleep measure you chose to predict and justify this choice.

Feature Extraction:

Extract meaningful features from the sensor data that can help predict sleep quality measures.

- Only use the raw sensor data and the amount of sleep hours from the previous night (not sleep times) as features, even though other information may be available.
- Time-based Features: Extract features related to different times of day (morning, afternoon, evening, night) to capture circadian patterns and daily activity variations.
- Location-based Features: Extract features related to different places where the subject was located, as location context can significantly impact sleep quality.
- Causal Model Consideration: The model should be causal in nature. While features can be extracted from previous time periods (backwards), ensure that no future information is used to predict current sleep quality.
- Properly handle any NaN values that appear in the data. At this stage, flipping windows for train and test divisions is not allowed.
- Save the output matrices.

Train & Test:

Prepare the data for model training and evaluation.

- Produce a personal prediction model for each student based on individual data changes.
- Divide the data into training and testing sets thoughtfully, ensuring it supports personalized predictions.
- Save the output matrices.

Vetting:

Perform initial data filtering and normalization

- Vet the data so that at the end of the process only 15 features remain.
- Normalize the data appropriately for regression tasks, explaining the chosen methods.
- Consider the continuous nature of the 1-10 rating scales when choosing normalization methods.
- Apply the same normalization to both training and testing data.
- Save the output matrices.

Feature Selection:

Select the final set of features for the model.

- Use the wrapper method with an exhaustive search strategy to select the most relevant features for predicting your chosen sleep measure.
- Justify the number of features selected and explain the discretization method and number of bins if used.
- Save the output matrices.

Note: This is a regression task predicting continuous values (ratings 1-10), so use regression models and evaluation metrics appropriate for continuous outcomes.

Submission:

- Submit all implemented functions along with any supporting code.
- The code should be written in a separate and clear way. Comments should be written in the appropriate places. A grade will be given for code clarity and documentation.
- All functions should run from a file called 'main_##' where ## is the group number. The code should run expecting the data to be in a folder called data, which is in the same path as the 'main_##' file.
- Submit the files in a single zip file named after your group number (e.g., 03.zip for group three).

Submission Deadline:

All function must be submitted by 14.06 (Saturday night at 23:59).

Good luck!

Dvir