Project Title: Sweet Dream - Sleep Health Tracker

Project Summary

Our team observed that many people have sleeping issues, and through research, we have found that “50 to 70 million Americans” have “sleep disorders” (U.S. Department of Health and Human Services) on government records. As sleep becomes a widespread issue in the country, we wanted to create a website to help people track their sleeping schedules and analyze their sleep patterns to determine their sleep quality. We aim to create a website that can store users' sleeping patterns as well as physical conditions, so we can provide personalized recommendations to each user on how to improve their sleep health. Another feature we aim to achieve is that if a user decides to disregard some of the criteria asked on the website, our program should be able to predict a reasonable range of data for the user based on their input.

We evaluate the wellness of a person’s sleep health through various factors, such as whether a person has slept enough hours for their age group, whether a person falls asleep with a reasonable stress level, or whether a person experiences high blood pressure right before they fall asleep, etc. Our webpage should be designed to display stored sleep health records for each user. In addition, our website will contain text boxes or drop-down menus that allow users to manually enter valid inputs and output results based on users’ desired selection of features.

Description

In this project, we will design a sleep health evaluation website. We plan to design two main functionalities for the website. For our first functionality, we will utilize the dataset to train different machine learning methods, trying to find the algorithm that makes the best reflection of the relationship between different data categories. Then when the user is using the website, they will have three ways to interact with the dataset. The first is that we do not require users to enter every data for every category. We will make suggestions about what other data they may need to test to better deal with their sleeping problems based on the relationship between data categories we found using machine learning algorithms. For example, we find there is correlation between sleeping condition A and sleeping condition B. Thus in the UI page, we will implement that “You only need to enter one of the following sleeping condition data”. In that way, we will suggest the user to monitor other sleeping conditions and provide the data that is correlating to the sleeping condition A they provided only if the sleeping condition A data is abnormal. The second is that we will also ask past users whether they would like to help others who experience similar sleeping issues after they solve their problems. Upon agreement, those users will provide their contact information, and we will provide the new users with similar sleeping problems contact information of users who have experienced similar problems in the past. We will also ask whether users want to provide their sleeping data to us for system improvement. If they agree, we will add their data into our dataset, and retrain the model after getting enough new data. In that way, we can make sure our data and the solutions provided are always up to date. The third function is that we will provide some suggestions on what they can do to improve their sleeping

issues to our users based on the data they entered. We will also provide visualization to user's data and our dataset for users to clearly see the comparison.

Technically Challenging features

The first technically challenging feature we may meet is that aside from our original data from Kaggle, most of our user data will have several or even most of the data categories missing. We need to find a way to ignore/replace the missing data when retraining the old model we get from past data and the dataset. We need to make sure that our model will not be heavily influenced by the missing data in a negative way, while still updating based on new data.

The second challenge we may meet is how much we can trust the prediction we made based on the dataset. The dataset might not be large enough to compare the data used by experts when making conclusions about sleeping disabilities, so the pattern we find in the data might be more convincing to users when there are potential theoretical supports. Also, there are probabilities that the correlation we found is caused by the data bias instead of true relationship between categories. We will improve the pattern we found from the old existing dataset with the new user data, but other potential ways may also need to be used when building the model or choosing the training algorithm.

The third challenge we may meet is how to test our functionalities, so we can make further improvements. The solution we have now before we start working on the project is that we found another dataset that can be used as a simulation of user data since it only has some categories that are shared with our main dataset. We can use this as part of test data to solve the first challenge mentioned above. But further solutions may also need to be found in order to better improve the system.

The last challenge we may meet is ethical problems. We will need to get the user data to better improve our model and users' contact information, but we need to ask for their permission because of ethical issues. So we need to think of ways(such as designing the website and building trust) to make our users more willing to provide their data and their personal information. This may need some further research and investigation.

Usefulness

Our web application includes basic functionalities for tracking and evaluating a person's sleep health condition based on their data inputs. In addition, if a user chooses to input only part of the data, our web application will predict the remaining criteria within a reasonable range and display their sleep health condition using clear visuals. Moreover, our web application features a more advanced option that generates personalized recommendations on how to improve their sleep quality.

We have found that there are similar physical products that have been introduced before, they are mostly in the form of smart watches which contain features that could track and store a person's sleep and health patterns. However, we believe that our website differs by allowing

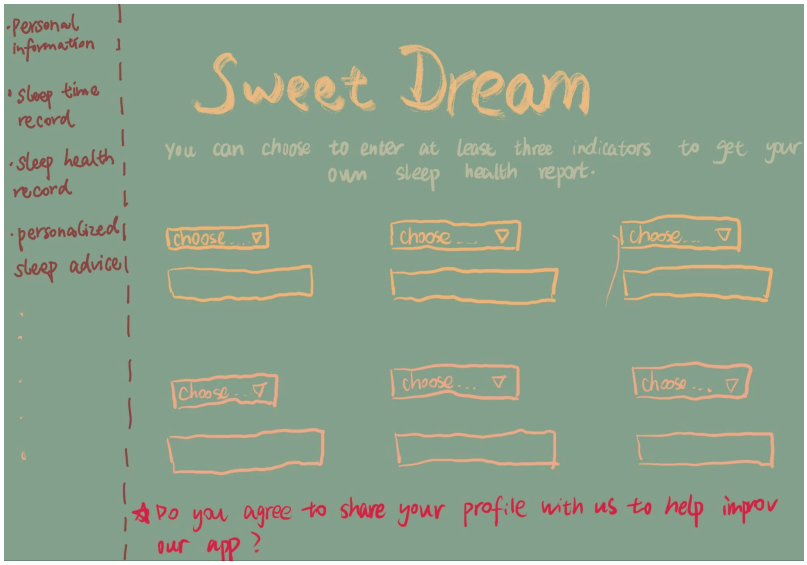
users to flexibly input data, making it possible for users to repeatedly test various sleep health conditions based on their desired data inputs. For example, if a person was advised by their doctor to use medication for high blood pressure problems, and if they want to know prior to taking the medication on how lower blood pressure can make an effect on their sleep condition, the person can input their predicted blood pressure value and the website can predict the result.

Realness

According to the functional requirements of this application, our team plans to use at least two datasets to analyze sleep conditions more accurately. The format of the first dataset is CSV. The degree is 13 and the cardinality is 373. It covers a wide range of variables related to sleep and physical data. It includes information such as gender, age, occupation, sleep duration, quality of sleep, physical activity level, stress levels, BMI category, blood pressure, heart rate, daily steps, and the presence or absence of sleep disorders. The format of the second dataset is CSV. The degree is 9 and the cardinality is 1000. It is a dataset measuring the comprehensive sleep and health of some sample of people. The first dataset provides a wide variety of physical data that are closely related to the evaluation of a person's sleep quality. Providing the degree of detailed data and usefulness of its application, the first dataset appears to be genuine. The second dataset captures similar data with fewer criteria, making it suitable for testing purposes as well as a potential supplement for the first dataset. This dataset also demonstrated a degree of reliability conveyed through its size and content. All these datasets are from Kaggle, which has a positive image for having reliable datasets. By using the real data from Kaggle, we can make sure that we are using accurate data to produce promising analysis results for the users.

Functionality

1. Low-fidelity UI mockup



Our app allows users to enter at least three categories to get their own sleep health report. For example, our database includes metrics such as blood pressure, sleep duration, BMI, heart rate, body temperature, stress levels, etc. However, users may not know all of their own metrics—they may only know three or four of them. In that case, they only need to input the metrics they are aware of (a minimum of three), and the remaining data will be generated based on our database to provide them with a personalized health report.

2. Project work distribution

Front-end Tasks:

User Interface design and implementation

Xiaoyu Li

Data visualization

Feihao F

Input validation

Hezi Jiang

Back-end Tasks:

User authentication and authorization

Selena Wang

Responsibility:

Implement secure user registration, login, log out, and password reset functionality.
Using authentication protocols for session management.
Encrypt user credentials using hashing algorithm.

Database design and management

Selena Wang

Responsibility:

Design and using the database schema
Define tables for users, sleep data entries, health metrics, and recommendations.
Establish relationships between tables (e.g., one-to-many between users and sleep entries).
Optimize queries and set up indexing for efficient data retrieval.
Implement data backup and recovery strategies.
Ensure data integrity and implement constraints where necessary.

API implementation for client server interaction
Feihao
Responsibility:
Develop RESTful API endpoints to facilitate communication between the frontend and backend.
Implement CRUD operations for user profiles, sleep data, and recommendations.
Ensure APIs are secure by implementing token-based authentication and input validation.
Handle error responses and status codes appropriately.

ML model integration
Feihao
Responsibility:
Develop machine learning models to predict missing user data and assess sleep quality.
Preprocess data including cleaning, normalization, and handling missing values.
Train models using algorithms suitable for time-series and health data.
Evaluate model performance using metrics, and adjust as necessary
Implement a retraining mechanism to update models with new user data periodically

Logging, error handling
Hezi Jiang
Responsibility:
Implement a logging framework to record backend events and errors.
Define log levels (DEBUG, INFO, WARN, ERROR) for effective debugging.
Develop error handling mechanisms to catch and manage exceptions.
Set up centralized log management.

Datasets
Dataset 1: https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset
Dataset 2: https://www.kaggle.com/datasets/uom190346a/sleep-and-health-metrics
Here are the two dataset we might use. The first one is our primary choice because it has more detailed information and categories. We might also consider using part of the second dataset as our input user data for testing, since they share some of the common categories(e.g. Sleep duration, physical activities, stress level…), which can be used as the simulation of users that do not enter their complete data for all categories.

Reference

- U.S. Department of Health and Human Services. (n.d.). *What are sleep deprivation and deficiency?*. National Heart Lung and Blood Institute.

https://www.nhlbi.nih.gov/health/sleep-deprivation#:~:text=Also%2C%20an%20estimated%2050%20to,a%20greater%20likelihood%20of%20death.