DeepSleep:

Optimal Sleep Pattern Prediction

National University of Singapore
Deep Learning Capstone Project

Introduction.

- Sleep – Most Important Activity for Human Beings.
- Body and mind recovery takes place during sleep after stressful activities.
- Lack of sleep quality affects overall health and statistically, is a causing factor for most major illnesses.
- May cause medical problems like heart disease, obesity & diabetes.

Problem Statement.

zZZ

Sleep Affects Productivity, Mood & Health.

98%

Sleep Quality and Pattern Was Hard To Determine.

Lack of Sleep Can Be Dangerous.

Was Not Accessible To General Public Without Industry-Grade Equipment.

Our Solution.

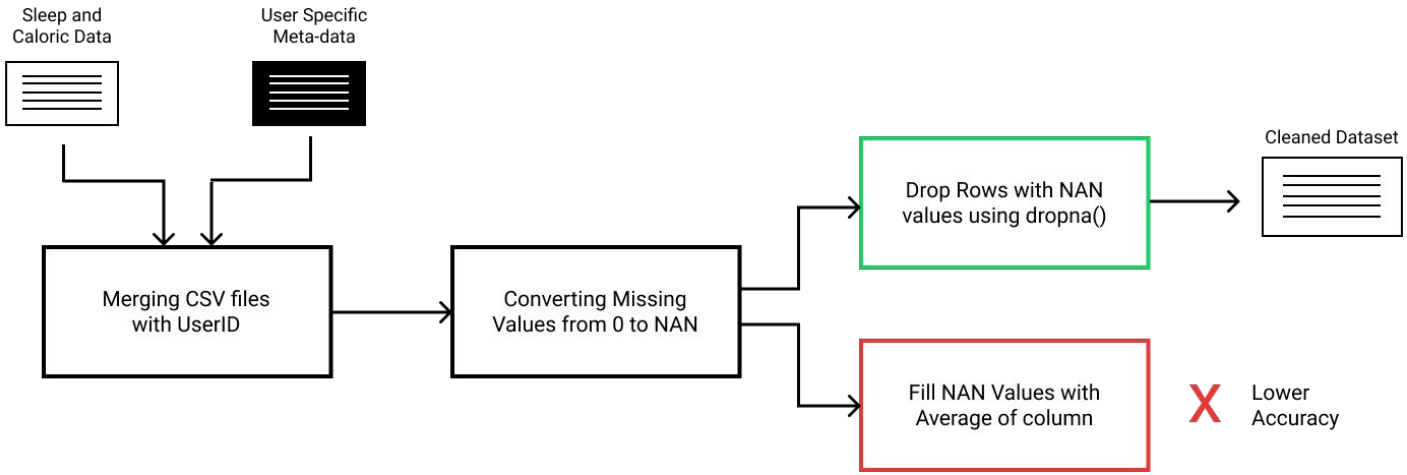
- Optimize Sleep Pattern for high Sleep Efficiency.
- Analyzes Daily Sleep and Caloric data from Smart Bands.
- Predicts **sleep mins** to calculate best sleep start time using Deep Learning model.
- Provides useful dashboard and graphs using ReactJS Web App.

Data Cleaning and Preprocessing.

Data Collection & Cleaning.

- Data collected from (https://github.com/Sungwon-Han/Learning-Sleep-Quality-from-Daily-Logs)
- Two data sets were available. The Sleep and Caloric data and the user specific meta-data consisting of ‘bmi’, ‘age’ and ‘ISI’.
- Merged the datasets together using pd.merge() on UserID.
- Replaced Missing values (set as 0) to NaN values in the dataset.
- ‘awaken_min’, ‘awaken_moments’ and ‘‘sleep_efficiency’ contained moving averages for each user.

Pre-Processing Architecture.



- Two options for handling missing values :
  - Dropping rows with missing values. Percentage of rows dropped= 4.98
  - Fill missing values with average of column.
- Dropping rows resulted in better accuracy.

Cleaned Dataset.

userId	cal consume	**sleep min**	awaken min	stairs
bmi	sleep efficiency	date	active ratio	awaken moments
isi	sleep start time	months	walks	nap total freq
age	sleep end time	active cal	distance	nap total time

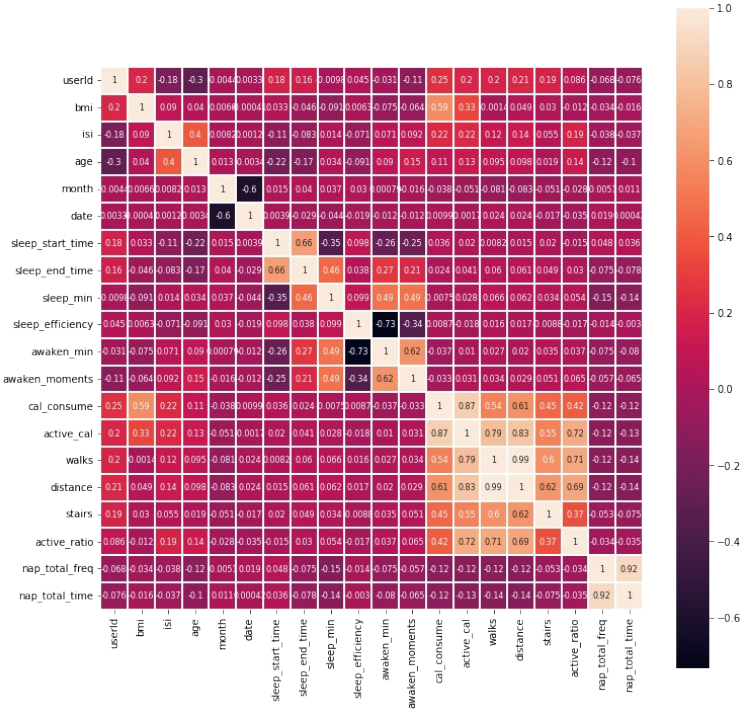
Exploratory Data Analysis.

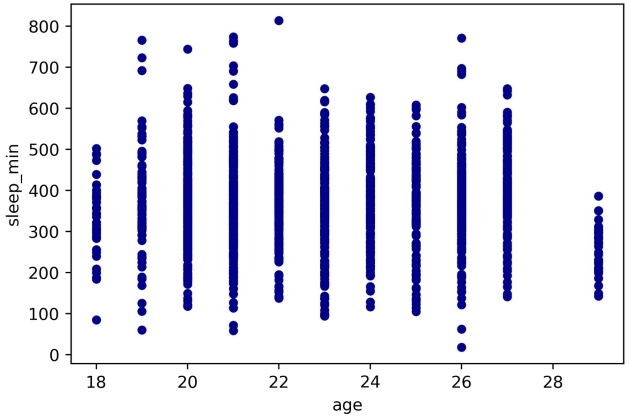
- **Heatmap for correlation:**
  - sleep min has a high positive relationship with
    - sleep_end_time
    - awaken_min
    - awaken_moments.
  - It has a negative relationship with sleep_start_time.

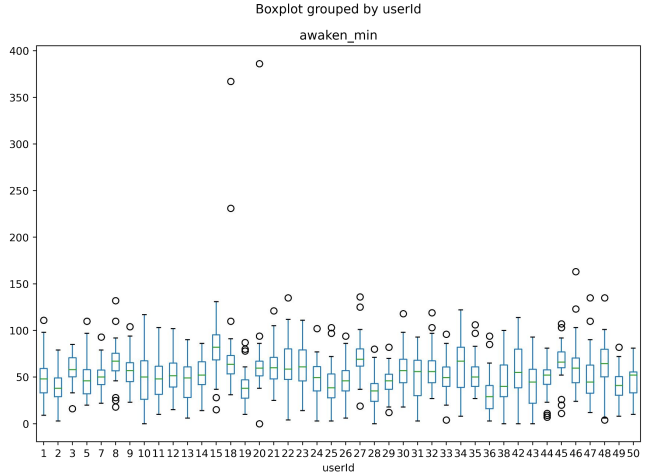
- **Scatter Plot:**
  - People whose age>28 have the least sleep minutes.
  - People whose age<19 and age>28 have sleep minutes<500.

- **Box Plot:**
  - Identifying the presence of outliers.
  - Approximate range of awaken minutes is less than 150.

Graphs.


userId
bmi
isi
age
month
date
sleep_start_time
sleep_end_time
sleep_min
sleep_efficiency
awaken_min
awaken_moments
cal_consume
active_cal
walks
distance
stairs
active_ratio
nap_total_freq
nap_total_time
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6


sleep_min
age
800
700
600
500
400
300
200
100
0
18
20
22
24
26
28


Boxplot grouped by userId
awaken_min
userId
400
350
300
250
200
150
100
50
0

Machine Learning.

Workflow.

- Columns with low correlation with target column are dropped.
- **Train-Validation-Test Split:**
  - Train set - 70% of the dataset.
  - Validation set - 20% of the dataset.
  - Test set - 10% of the dataset.
- MinMax normalization of train, validation and test sets.
- **Models:**
  - Linear Regression
  - Decision Tree Regression
  - Random Forest Regression
  - Linear Support Vector Regression
  - XGBoost Regression

ML Models.

Model Name	R2 Score	MAE	MAPE
Linear Regression	0.097	67.374	25.769
Decision Tree Regression	-0.104	76.578	29.729
Random Forest Regression	0.124	67.238	25.166
Linear Support Vector Regression	0.101	67.418	25.644
XGBoost Regression	-0.228	81.438	25.786

- In spite of using different models, there is no significant rise in accuracy.

Deep Learning.

Overview.

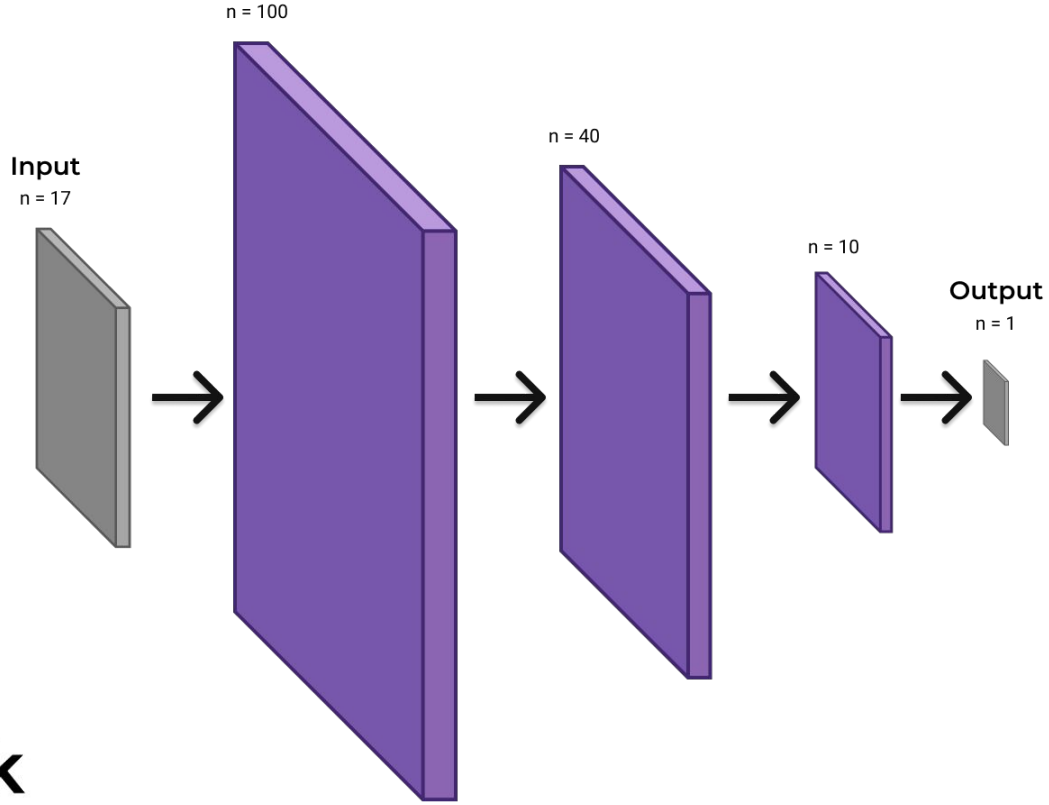
- Traditional Neural Network and LSTMs were used.
- Neural Networks turned out to be better.
- awaken_min, awaken_moments and sleep_efficiency contained their averages till previous day.
- Early stopping to prevent overfitting.

Intuition for LSTM.

$$N_h = \frac{N_s}{(\alpha * (N_i + N_o))}$$

N_i = number of input neurons.
N_o = number of output neurons.
N_s = number of samples in training data set.
α = an arbitrary scaling factor usually 2-10.

- Nh = 1600 / (2 * (17+1)) = 44.44
- We therefore used 42 neurons per layer in our LSTM, Bi-LSTM and GRU
- 2 Hidden layers and Timesteps : 21 days



Neural Network Design.

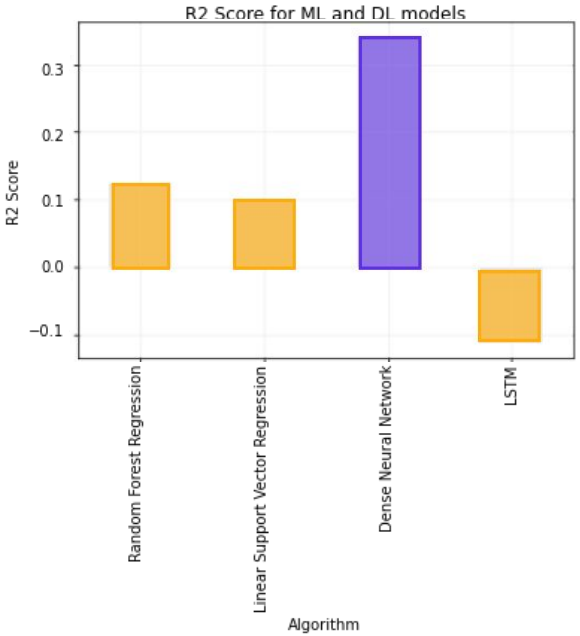
Results.

Model Name	R2 Score	MAE	MAPE
Dense Neural Network	0.334	62.842	27.238
Bi-LSTM	-0.149	82.915	34.968
LSTM	-0.112	82.400	34.982
GRU	-0.241	88.166	37.593

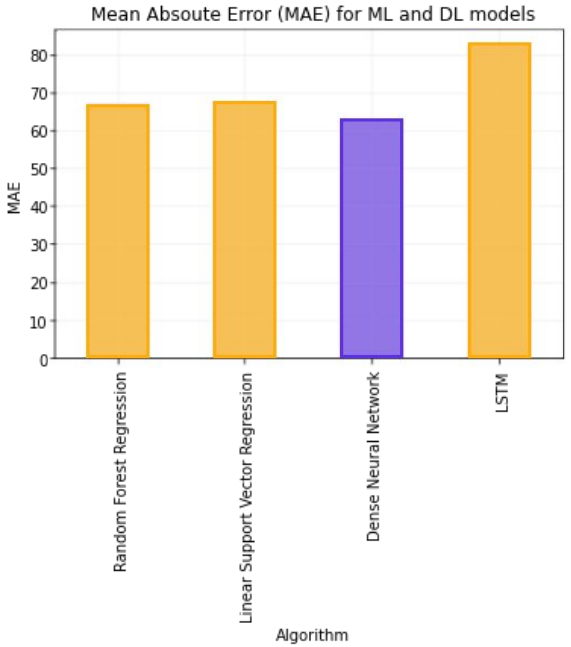
- Behavioral data, therefore the low R2 score for NN.
- More dataset size could have improved model.
- Ensemble of neural networks could have improved predictions.

Model Comparison.

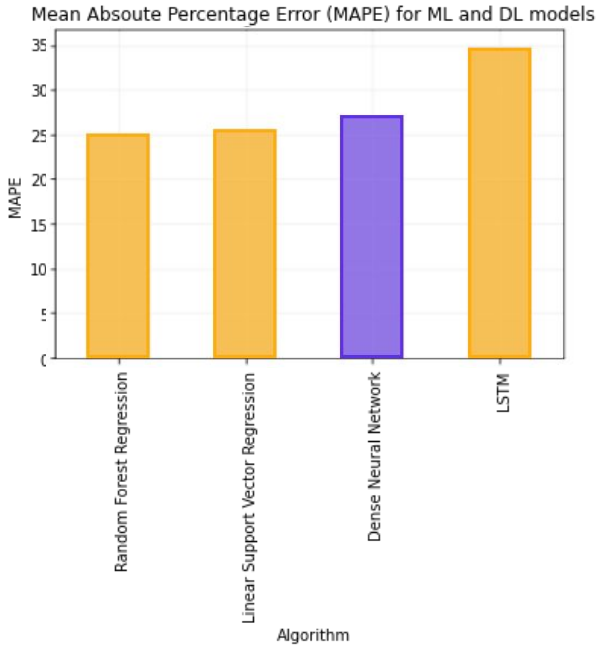
R2 Score:



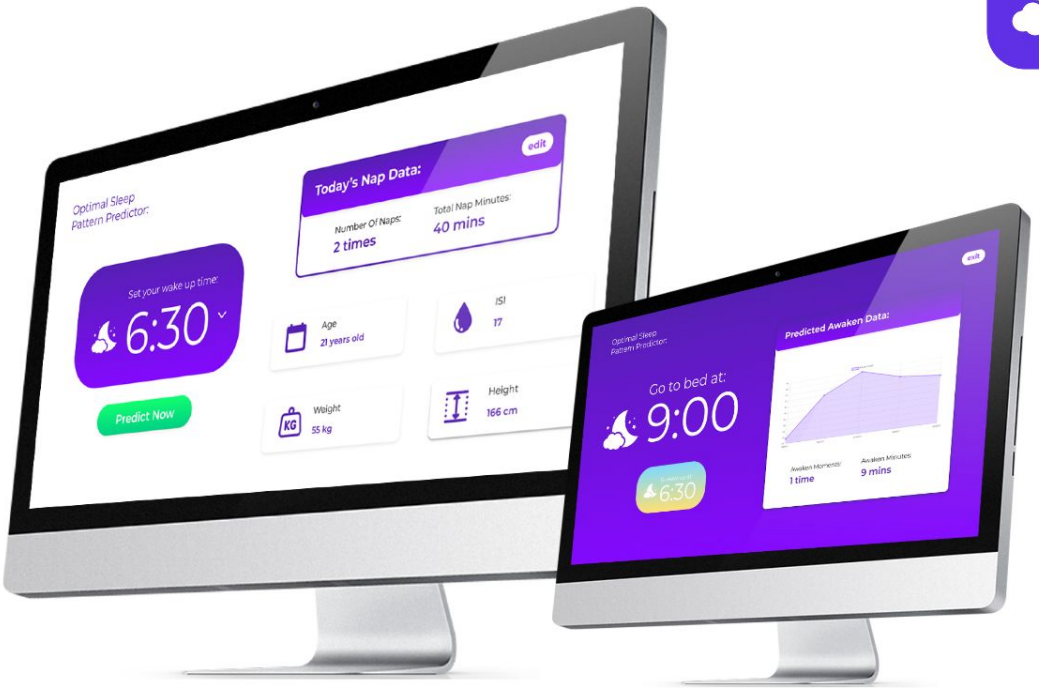
MAE:



MAPE:



App Development.



Front End.

- Developed React JS FrontEnd User Interface.
- Used Axios to send HTTP requests.
- Provides Sleep & Caloric Analytics Dashboard.
- Dynamic & Easy To Use.

Back End.

- Used Flask Framework for backend.
- Created API endpoints.
- Refactored and Connected Deep Learning Model.
- Processed the data and predicted the output.

Workflow.

User enters data in the frontend

Other data like age, weight, height etc is taken from the backend directly.

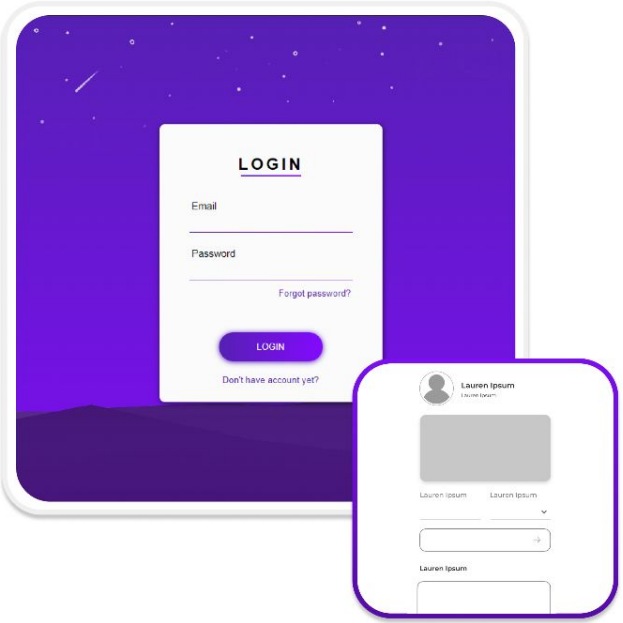
Using Axios a post request is made to Flask backend

Data is retrieved, processed and given to the DL model.

We get the number of sleep minutes as the output, using this we predict the time at which user should go to sleep

Display output on the frontend

Project Demo.



Future Enhancements.

- User Login Page.
- Dedicated Profile Page.
- CI & CD based On User-Data.
- Smart Device Integration.

References.

- Original Paper: http://shaun.kr/file/ads1910p-parkA.pdf
- Dataset: https://github.com/Sungwon-Han/Learning-Sleep-Quality-from-Daily-Logs/tree/master/data
- Deep learning models: https://towardsdatascience.com/predictive-analysis-rnn-lstm-and-gru-to-predict-water-consumption-e6bb3c2b4b02
- Grid SearchCV: https://machinelearningmastery.com/hyperparameter-optimization-with-random-search-and-grid-search/
- Randomized SearchCV: https://www.section.io/engineering-education/random-search-hyperparameters/

The Conclusion.

Jack Praveen Raj Ilango

Data Processing, Machine Learning, Deep Learning, Frontend (HTML CSS) and Presentation Design

Aditya G

Data Processing, Feature Engineering and Deep Learning

Megha N

Exploratory Data Analysis, Machine Learning and Deep Learning

Shriya Dutta

Exploratory Data Analysis, Hyperparameter Tuning and Machine Learning

Rishit Puri

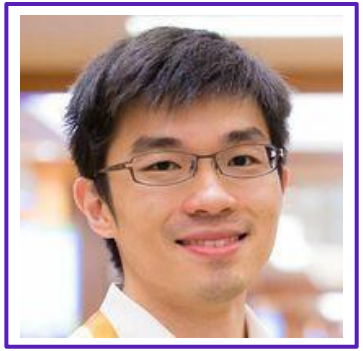
Data Cleaning, Data Processing and Machine Learning

Tanmay Garg

Frontend, Backend, Deployment and Hyperparameter Tuning

Thank You.

National University of Singapore
Deep Learning Capstone Project



Dr. LEK Hsiang Hui

Senior Lecturer

Dr. TAN Wee Kek

Senior Lecturer

Dr. Amirhassan Monajemi

Senior Lecturer

Thank You.