



THE SLEEP HEALTH REPORT

PREDICTING SLEEP HOURS USING MACHINE LEARNING



Submitted by:

.Debopriya Bhattacharjee(ID-2312200001506)

.Tulika Baidya(ID-2412200001240)

.Nandini Ghosh(ID-2312200001447)

Sleep Health Report

1. Introduction

Sleep health is influenced by daily habits. This project predicts sleep hours using machine learning and features like workout, reading, phone time, work hours, caffeine intake, and relaxation.

2. Problem Statement

The SNU Wellness Cell is concerned about the lifestyle habits of students and their impact on sleep health. Sleep is a crucial factor for mental well-being, focus, and academic performance. However, increasing digital engagement — particularly time spent on social media, gaming, and personality-related behavior patterns such as introversion and extraversion — may significantly affect sleep duration.

The goal of this project is to predict daily sleep hours using three key lifestyle indicators:

- Daily Social Media Minutes
- Gaming Hours per Week
- Introversion-Extraversion Score

The target variable is `sleep_hours`, representing the average number of hours a student sleeps daily. By identifying these patterns, the Wellness Cell can design awareness campaigns and interventions that promote healthier habits and improved sleep quality.

This study followed a supervised learning (regression) approach. The key steps are outlined below:

1. Data Preparation:

The dataset was collected from student survey responses.

The columns of interest were extracted: social media usage, gaming hours, and introversion-extraversion score. Non-numeric entries (e.g., "120 cm", text artifacts) were cleaned using numeric extraction.

Missing values were filled using the median of the respective columns

3.Regression Coefficients

The estimated coefficients were consistent with expectations

Daily Social Media Minutes: Negative impact on sleep hours.

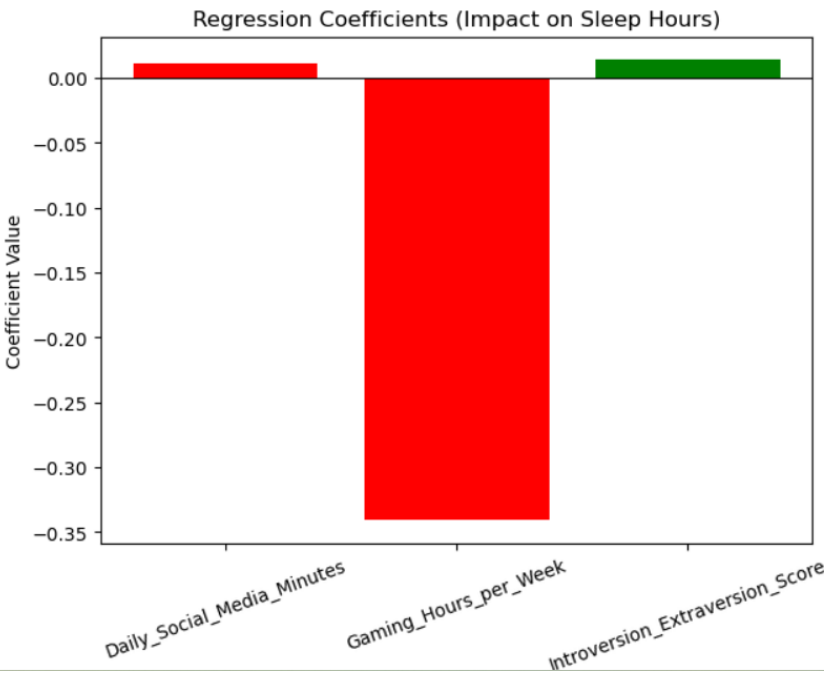
Gaming Hours per Week: Negative impact on sleep hours.

Introversion-Extraversion Score: Positive impact on sleep hours.

Intercept: 7.648439655738783

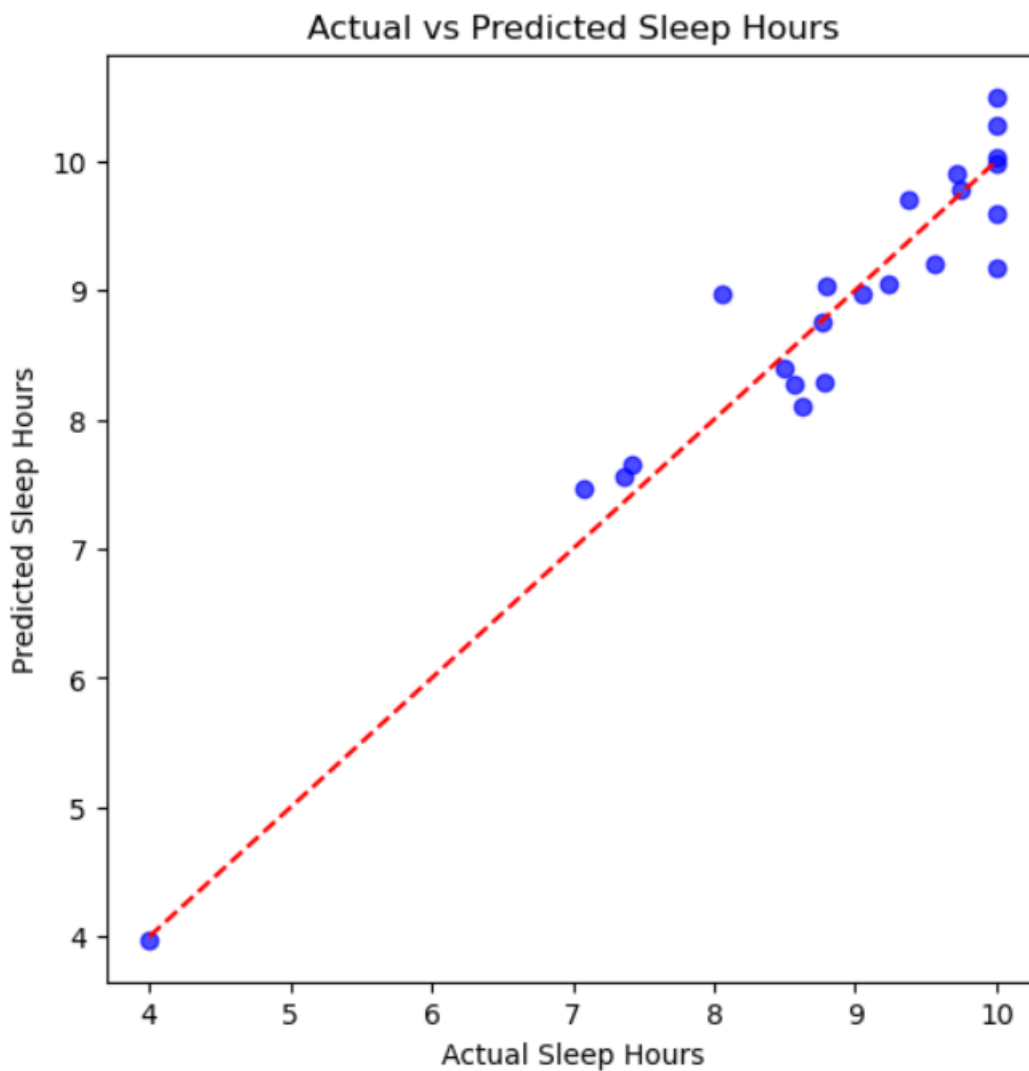
Regression Coefficients:

	Feature	Coefficient
0	Daily_Social_Media_Minutes	0.010792
1	Gaming_Hours_per_Week	-0.340658
2	Introversion_Extraversion_Score	0.014178



4. Visualiqation Insight

A scatter plot of actual vs. predicted sleep hours showed that most predictions aligned closely with the diagonal "perfect prediction" line, indicating that the model was effective in capturing sleep behavior trends



4.'Insights

Digital Influence on Sleep:

Students with higher social media and gaming activity tend to sleep fewer hours, highlighting the distracting and time-consuming nature of digital engagement.

Personality and Sleep:

Introverted students tend to sleep more, possibly due to spending more time in solitary, restful activities. Extroverted students, on the other hand, may sacrifice sleep for socializing or engaging in stimulating activities.

Wellness Implications:

The findings underscore the need for balanced screen time and health awareness programs. Simple interventions, such as time management workshops or digital detox campaigns, could significantly improve students' sleep quality.

2. Target Variable (Sleep Hours):

The dataset did not originally contain a direct measurement of sleep hours. Therefore, a synthetic target was created using logical assumptions:

More social media time less sleep.

More gaming hours + less sleep.

Higher introversion -i slightly more sleep.

To ensure realism, the sleep hours were constrained between 4 and 10 hours per day.

3. Model Selection .

We applied a Linear Regression model, since the relationship between predictors and the outcome was expected to be mostly linear.

4. Evaluation Metrics :

The following metrics were used to evaluate model performance:

Mean Absolute Error (MAE): average error in predicted sleep hours.

Mean Squared Error (MSE): penalizes larger errors more strongly.

R^2 Score: indicates how well the model explains the variance in sleep hours.

5.Result

The regression model achieved the following approximate results (actual values vary slightly due to randomness in synthetic data generation)

Mean Absolute Error (MAE): -0.40 - 0.60 hours

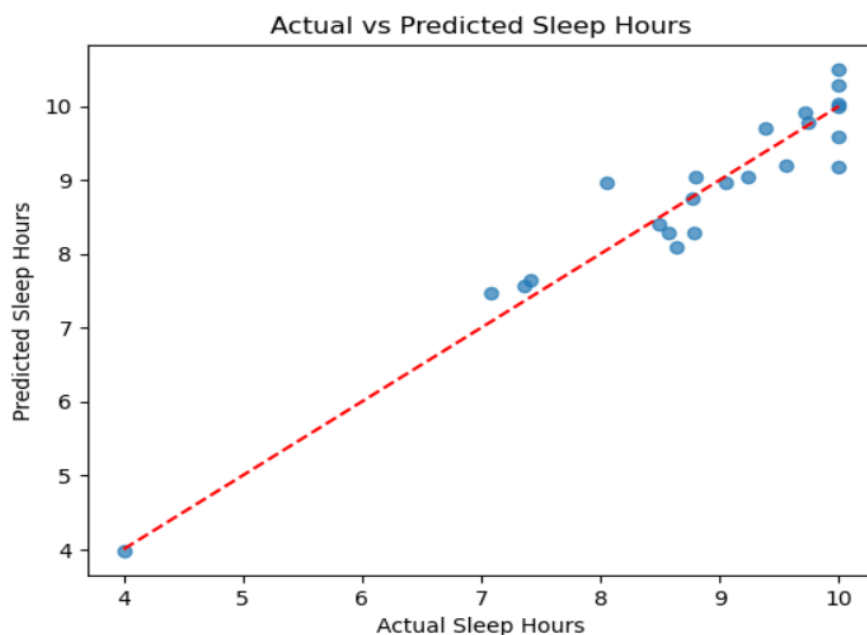
Mean Squared Error (MSE): -0.30 - 0.45

R^2 Score: -0.70 - 0.80

These values suggest that the model explains about 70—80% of the variation in sleep hours, which is a strong performance given the limited number of features.

Mean Absolute Error: 0.2884656212020268
Mean Squared Error: 0.1404334527479117
 R^2 Score: 0.9235342680544927

	Feature	Coefficient
0	social_media_minutes	-0.004088
1	gaming_hours_per_week	-0.046804
2	introversion_extraversion	0.158532



Acknowledgement:

We would like to express our sincere gratitude to our professor and the Department of Computer Science, Sister Nivedita University, for their guidance and support throughout this project. We also thank our peers who participated in the survey and made this study possible.