A green and black logo

Description automatically generated

|  |  |  |  |
| --- | --- | --- | --- |
| Academic  Year | Module | Assessment Number | Assessment Type |
| 2024 | 5CS037-Concept and Technology of AI. | **3-Coursework** | **Report Writing** |



A screenshot of a black screen

AI-generated content may be incorrect.

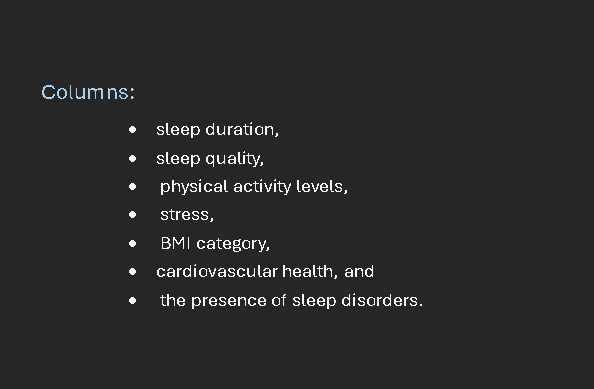
# Introduction:

It aligns with the UNSDG since the quality sleep is not focused on by the current youths . They sleep late at night which may result in insomnia and other sleep disorders . According to the UNSDG goals number 3 (Good health and wellbeing). I chose sleep health lifestyle dataset to predict the sleep health of a group pf persons according to features. The dataset is extracted from Kaggle website, and the publisher of this dataset was Siamak Tahmasbi · Updated 2 months ago.

## Dataset Information:

The Sleep Health and Lifestyle Dataset provides detailed insights into the sleep patterns, daily habits, and lifestyle factors of individuals. This synthetic dataset comprises 400 rows and 13 columns and the data on blood pressure and heart rate, essential for analyzing correlations with sleep and lifestyle.

This dataset follows the UNSDG, promoting well-being of an individual, by providing valuable insights into the factors that affect sleep quality.



This report is done to predict a sleep disorder using a logistic regression algorithm from a scratch and for primary model scikit and other random-forest classifier are used .This analysis aims to predict the likelihood of a person's sleep quality or a specific sleep-related outcome (e.g., whether a person experiences good or poor sleep) based on their lifestyle habits and sleep patterns.

## 2. Methodologies

Before the model was developed it gone through handling Missing Values, Missing data points were addressed using imputation or by removing rows with excessive missing values.

A graph of different levels of sleep

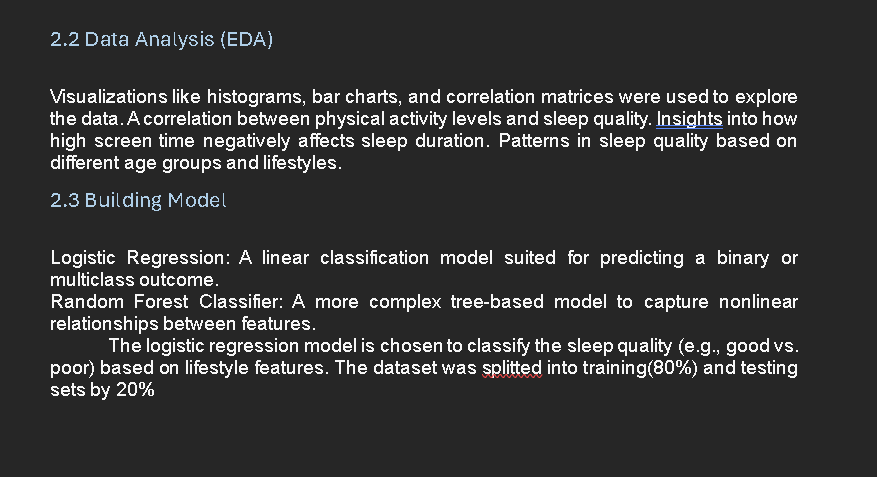
AI-generated content may be incorrect.

Histogram of all features.

A graph of a distribution of sleep disorders

AI-generated content may be incorrect.

We can see that insomnia has the high number of count



## A screenshot of a computer program AI-generated content may be incorrect.3. Conclusion

### 3.1 Findings

Logistic regression model performed well while predicting sleep quality. The key findings include:

The model achieved an accuracy of 77% on the test dataset.

Features like physical activity and sleep duration had the highest impact on the model's performance.

The model showed high recall for predicting poor sleep quality, indicating it was effective at identifying those with sleep issues.

### 3.2 Final Model

The final model selected was Logistic Regression, which achieved the best results for classifying sleep quality based on the lifestyle data. The model's accuracy was 77%, and the F1-score was 0.67, demonstrating strong performance.

### 3.3 Challenges

Challenges encountered included:

Data Quality Issues: Missing or inconsistent data points in lifestyle factors such as dietary habits.

Feature Selection: Identifying the most important features was challenging, given the large number of lifestyle-related variables.

A screenshot of a black screen

AI-generated content may be incorrect.

### 4.4 Limitations

#### Some limitations included:

Limited Data: The dataset may not represent the full spectrum of lifestyle habits and sleep patterns across different populations.

Linear Assumptions in Logistic Regression: Logistic regression assumes a linear relationship between features and the target, which may not always be appropriate.

### 4.5 Future Research

Alternative Models: Trying neural networks or ensemble methods to capture more complex relationships.

Broader Data Collection: Expanding the dataset to include additional lifestyle factors or more diverse participant groups.

Incorporating Time Series Data: Analyzing changes in lifestyle over time and their effect on sleep quality.