A logo for college computing

Description automatically generated

**Assessment Cover Page**

|  |  |
| --- | --- |
| *Student Full Name* | Denisse Garcia Vivolo |
| *Student Number* | sbs24050 |
| *Module Title* | Machine Learning |
| *Assessment Title* | CA1 Project |
| *Assessment Due Date* | 28/04/2024 |
| *Date of Submission* | 28/04/2024 |

**Declaration**

By submitting this assessment, I confirm that I have read the CCT policy on academic misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source.

I declare it to be my own work and that all material from third parties has been appropriately referenced.

I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution.

**Influence of Lifestyle Factors on Sleep Disorders and Quality**

Contents

[Introduction 1](#_Toc165145277)

[Motivation 1](#_Toc165145278)

[Problem Statement 1](#_Toc165145279)

[Objectives 1](#_Toc165145280)

[Data Characterization and Pre-processing 1](#_Toc165145281)

[Hyperparameter Tuning 2](#_Toc165145282)

[Results Interpretation and Model Rationale 2](#_Toc165145283)

[References 5](#_Toc165145284)

[NOTE: The table of contents above has been included for your convenience. To refresh the table, simply click on it, then select 'Update Table' using the mouse. You can choose to update either the page numbers exclusively or the entire table as needed.]

# Introduction

This report explores the application of machine learning techniques to the Sleep Health and Lifestyle Dataset. The goal is to uncover insights into sleep patterns and disorders, which can inform health interventions and lifestyle adjustments

Analysis of Sleep Health and Lifestyle Dataset using Machine Learning Techniques

sleep patterns and disorders, which can inform health interventions and lifestyle adjustments.

# Motivation

Sleep health is a critical component of overall well-being, influencing physical health, mental health, and quality of life. Understanding factors that impact sleep can help in developing targeted treatments and recommendations.

# Problem Statement

The dataset contains information on various aspects of individual lifestyles and their sleep patterns, including age, gender, occupation, sleep duration, and health-related metrics.

# Objectives

The primary objective is to use prediction and classification algorithms to:

- Predict the quality of sleep based on lifestyle factors.

- Classify individuals with potential sleep disorders.

These goals are pursued to enhance the understanding and management of sleep health.

# Data Characterization and Pre-processing

\*\*Data Characterization\*\*

The dataset includes 374 entries across 13 attributes such as age, sleep duration, physical activity, and more. Each attribute's role is briefly explained with statistical summaries.

\*\*Pre-processing Steps\*\*

Data cleaning involved removing duplicates and handling missing values. The dataset was randomly sampled to create different subsets for training (20%, 25%, 30% splits) to evaluate the models under varying conditions.

\*\*Cross Validation\*\*

Cross-validation techniques were used to assess the robustness of the model predictions. The variance in model accuracy across different splits is discussed to illustrate the stability of the predictive models.

# Hyperparameter Tuning

\*\*Purpose of Hyperparameter Tuning\*\*

Hyperparameter tuning is essential to optimize the performance of machine learning models. It involves adjusting parameters that govern the model training process.

\*\*Tuning Techniques\*\*

GridSearchCV was utilized to systematically work through multiple combinations of parameter tunes, cross-validating as it goes to determine which tune gives the best performance.

# Results Interpretation and Model Rationale

\*\*Model Interpretation\*\*

The outcomes are analyzed, focusing on the accuracy and robustness of predictions. The performance metrics are illustrated through confusion matrices and ROC curves.

\*\*Discussion on Model Fit\*\*

The analysis discusses potential overfitting or underfitting, highlighting how the models generalize to new data.

\*\*Rationale for Model Choice\*\*

The choice of models, including decision trees and SVM, is justified based on their performance and suitability for the dataset's characteristics.

Given the variables available in your "Sleep Health and Lifestyle Dataset," there are several interesting questions and analyses you can explore. Here are some potential questions based on the data:

1. \*\*Gender Differences in Sleep Health\*\*:

- What is the impact of gender on sleep quality? Are males or females more likely to report poor sleep quality?

- How does BMI influence sleep quality across genders? For example, you can analyze whether an increase in BMI category (from normal to overweight or obese) affects sleep quality differently in males vs. females.

2. \*\*Age and Sleep Duration\*\*:

- How does age affect sleep duration and quality? Is there a correlation between increasing age and changes in sleep duration or quality?

3. \*\*Occupational Impact on Sleep\*\*:

- Does occupation type correlate with sleep quality or duration? For example, comparing high-stress jobs (like doctors) vs. lower-stress jobs.

4. \*\*Physical Activity and Sleep Health\*\*:

- How does the level of physical activity affect sleep quality? Is higher physical activity associated with better sleep quality?

5. \*\*Stress and Sleep Disorders\*\*:

- What is the relationship between reported stress levels and the presence of sleep disorders? Can high stress levels predict the occurrence of specific sleep disorders like Sleep Apnea?

6. \*\*BMI, Blood Pressure, and Heart Rate as Predictors\*\*:

- Can BMI, blood pressure, and heart rate predict sleep quality or the presence of sleep disorders? This can be an exploration of how physiological measures correlate with sleep health.

### Example Analysis: Impact of BMI on Sleep Quality by Gender

To address the question about the impact of BMI on sleep quality and how it varies by gender, you could perform the following analysis:

- \*\*Statistical Analysis\*\*: Use statistical tests to compare sleep quality across different BMI categories within each gender. This might involve using ANOVA or Kruskal-Wallis tests if the sleep quality is ordinal or not normally distributed.

- \*\*Visualization\*\*: Create boxplots or bar charts to visualize the differences in sleep quality across BMI categories for each gender.

- \*\*Predictive Modeling\*\*: Use logistic regression or decision trees to predict sleep quality based on BMI, gender, and their interaction.

Let's start with a statistical summary and a visualization to explore the relationship between BMI, gender, and sleep quality. I will generate the necessary code for this analysis. If this aligns with your needs, we can proceed!

\*\*Visualizations\*\*

Graphs and charts are included to visualize the data distributions and model performances.

\*\*Python Code Comments\*\*

Commented Python code is provided to show the implementation of data processing, model training, and evaluation.

### Conclusion

The report concludes with a summary of findings, implications for sleep health management, and suggestions for further research.

---

# References