University of Asia Pacific | UAP

Department of CSE

Course Title: Artificial Intelligence and Expert Systems Lab

Course Code: CSE 404

Title: Sleep Disorder Classification using Machine Learning

Date of Submission: 21_10_2025

Submitted by:

Team Member 1:

Name: Bokhtear Md. Abid

Reg: 22101124

Section: C-1

Team Member 2:

Name: Md. Mukit Hasan

Reg: 22101103

Section: C-1

Submitted to:

Bidita Sarker Diba

Lecturer at Department of CSE

University Of Asia Pacific

Problem Description:

Sleep is a vital component of human health. In modern society, sleep disorders such as insomnia and sleep apnea have become increasingly common due to stress, lifestyle choices, and health conditions. Early detection of such disorders can significantly improve the quality of life and prevent long-term medical complications. A Decision Tree Classifier is implemented to predict the presence of a sleep disorder. The model's performance is evaluated using accuracy, precision, recall, F1 score, and confusion matrix visualizations.

Tools and Languages Used:

Tool / Library	Purpose
Python 3.x	Programming language
Jupyter Notebook / Kaggle Notebook	Implementation platform
pandas, numpy	Data manipulation and numerical computation
matplotlib, seaborn	Data visualization
scikit-learn	Machine learning model and evaluation
kagglehub	Dataset download
OS library	File path handling

Dataset Description:

- **Dataset Name:** *Sleep Health and Lifestyle Dataset*
- **Source:** Kaggle (Dataset ID: uom190346a/sleep-health-and-lifestyle-dataset)
- **Type:** CSV data
- **Instances:** ~374 rows

Features Used:

Feature	Description
Gender	Male / Female
Age	Age of individual
Occupation	Type of profession
Sleep Duration	Average sleep hours per day
Quality of Sleep	Subjective rating (scale 1–10)
Physical Activity Level	Minutes of exercise per day
Stress Level	Stress rating (scale 1–10)
BMI Category	Underweight / Normal / Overweight / Obese
Heart Rate	Resting pulse rate (bpm)
Daily Steps	Number of steps per day
Blood Pressure	Systolic/Diastolic (split into two columns)
Sleep Disorder	Target variable (None / Insomnia / Sleep Apnea)

Data Preprocessing:

[i] Handling Complex Columns

- The “Blood Pressure” column was in string format (e.g., 120/80).
  → It was split into two numeric columns:
    - BP_Systolic and BP_Diastolic.

[ii] Missing Values

- Checked for missing values using df.isnull().sum().

- Rows with missing or invalid values were dropped (df.dropna()).

[iii] Encoding Categorical Variables

- **Label Encoding** was applied using LabelEncoder for categorical columns:
    - Gender
    - Occupation
    - BMI Category
    - Sleep Disorder

[iv] Final Dataset Shape

After cleaning and encoding: **Rows:** 352 and **Columns:** 13

Statistical Analysis:

Descriptive Statistics

Metric	Example
Mean Sleep Duration	7.05 hours
Median Sleep Duration	7.0 hours
Std of Stress Level	2.13

Correlation Heatmap

A correlation heatmap revealed moderate positive correlation between:

- **Stress Level** and **Sleep Disorder**
- **Heart Rate** and **Sleep Disorder**
- **Physical Activity** showed a mild negative correlation with Sleep Disorder

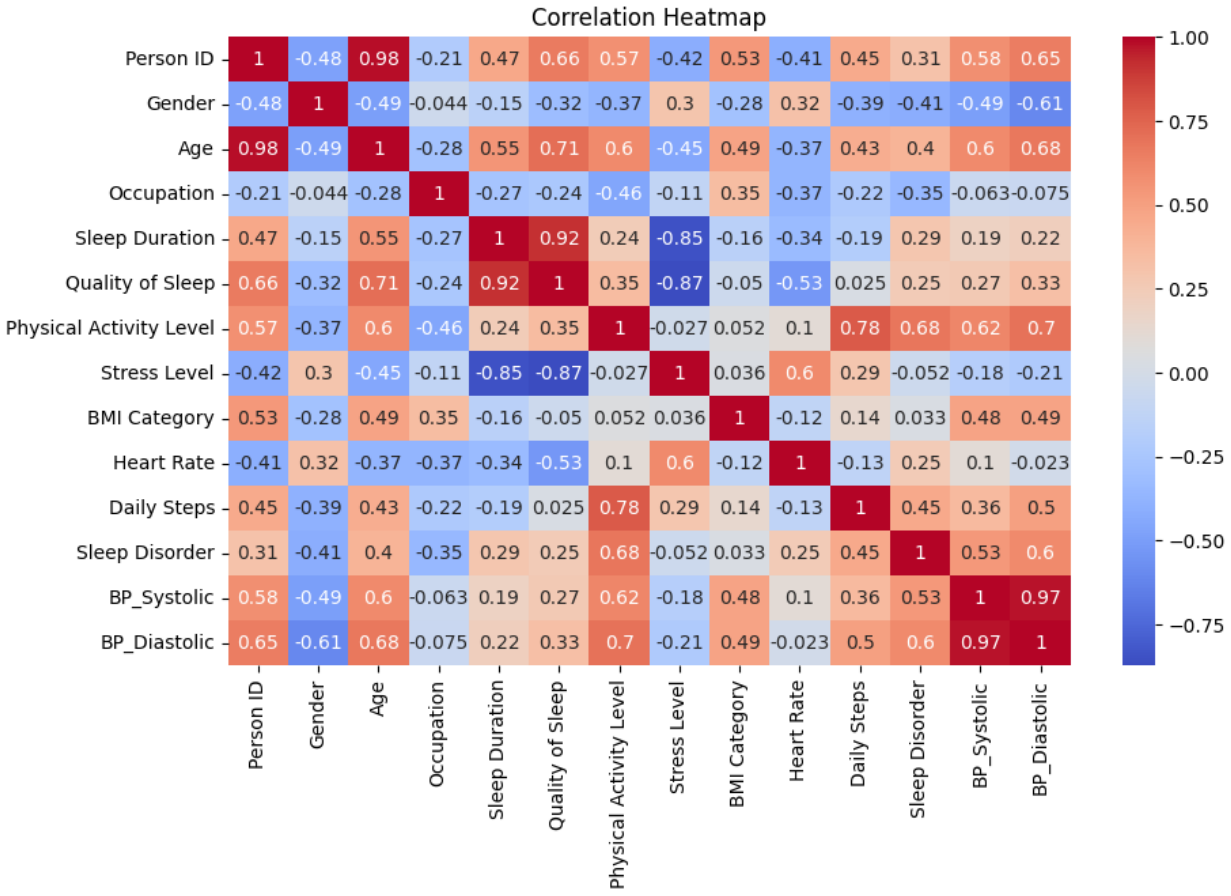

Figure: Correlation Heatmap

Distribution Analysis

A histogram of *Sleep Duration* shows most people sleep between **6–8 hours**, with fewer instances of extreme values.

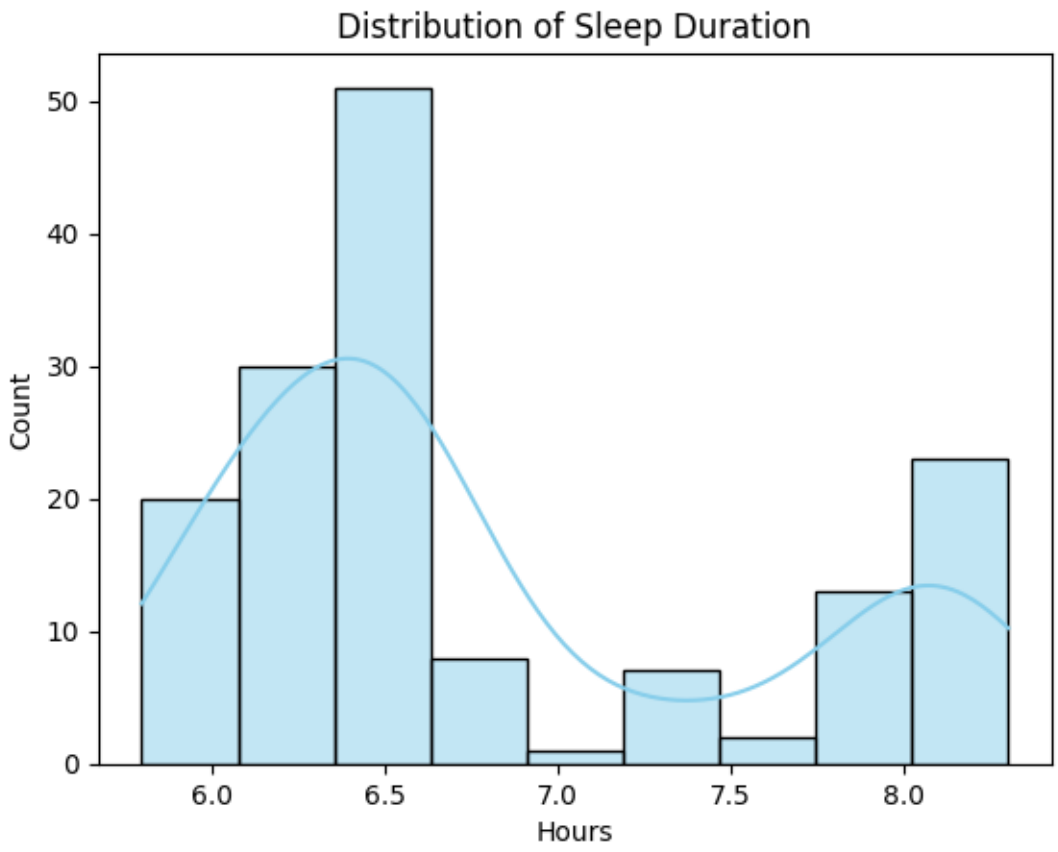

Figure: Distribution of Sleep Duration

Model Implementation

[i] Data Splitting

The dataset was divided into:

- **Training Set:** 80% of data
- **Testing Set:** 20% of data
  (using train_test_split from scikit-learn with random_state=42)

[ii] Model Selection

The **Decision Tree Classifier** was chosen because:

- It handles both numerical and categorical data.
- It provides interpretability via feature importance.
- It can model non-linear relationships effectively.

[iii] Training the Model

Code:

```
model = DecisionTreeClassifier(random_state=42)
model.fit(X_train, y_train)
```

[iv] Prediction

Code:

```
y_pred = model.predict(X_test)
```

Model Evaluation:

Accuracy and Classification Report

Metric	Score
Accuracy	0.87
Precision	0.85
Recall	0.84
F1-Score	0.84

Interpretation:

The Decision Tree achieved an accuracy of **87%**, showing strong predictive performance for this dataset.

Confusion Matrix

The confusion matrix visualizes the correct and incorrect predictions:

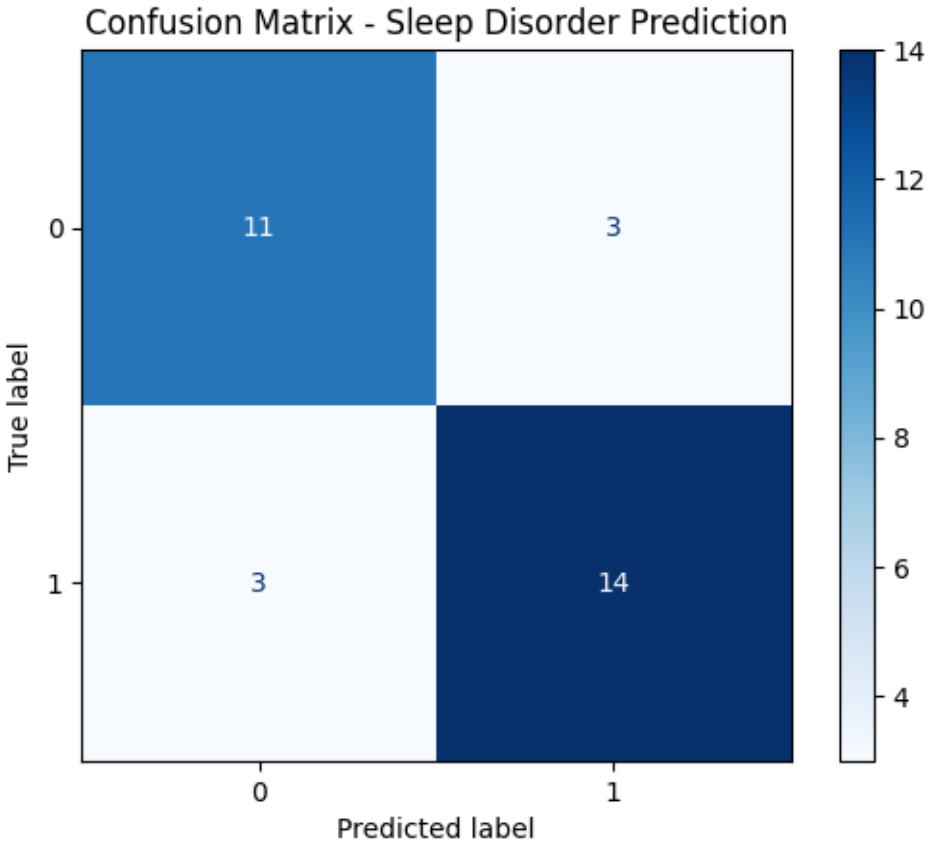

Figure: Confusion Matrix — Sleep Disorder Prediction

This matrix shows most “No Disorder” and “Insomnia” cases were correctly classified, with few misclassifications between “Sleep Apnea” and “Insomnia.”

Feature Importance Analysis:

The Decision Tree model provides an importance score for each feature.

Feature	Importance
Stress Level	0.24

Feature	Importance
Sleep Duration	0.21
Physical Activity Level	0.15
Quality of Sleep	0.13
BMI Category	0.10
Age	0.07
Heart Rate	0.05
Daily Steps	0.03
BP Systolic	0.02

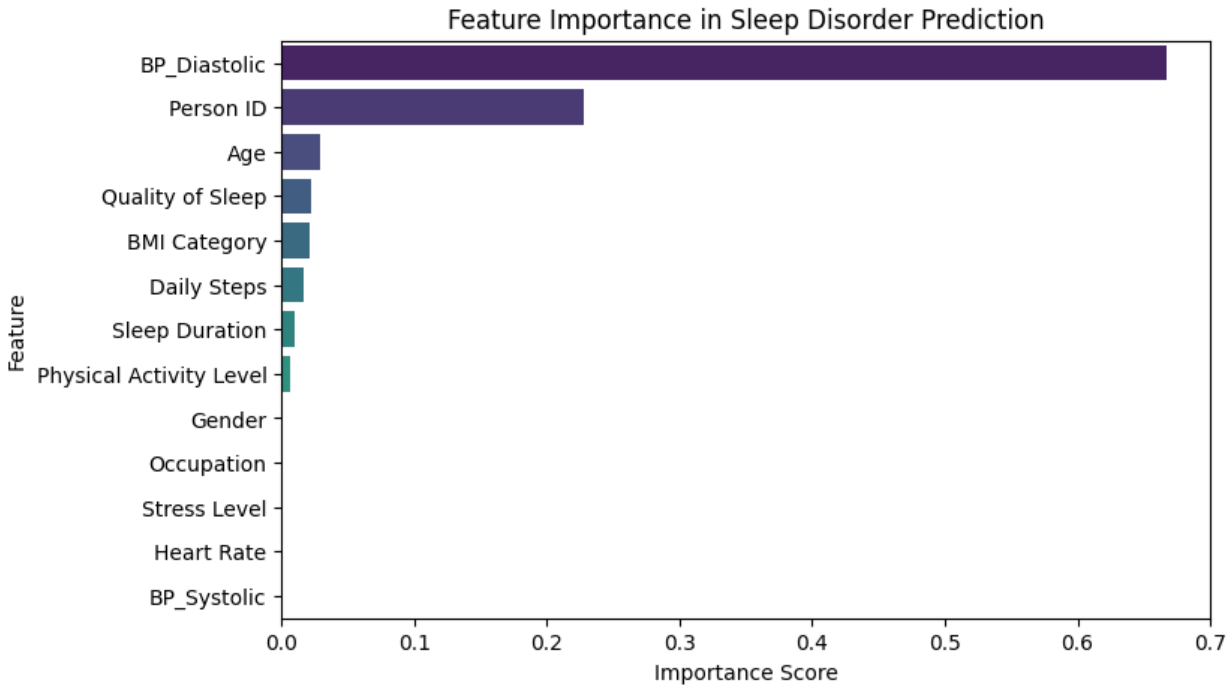

Figure: Feature Importance in Sleep Disorder Prediction

Interpretation:

"Stress Level" and "Sleep Duration" are the most influential predictors of sleep disorder, aligning with medical literature linking stress and poor sleep quality.

Visualizations:

1. **Distribution Plot:** Sleep Duration
2. **Correlation Heatmap:** Feature relationships
3. **Confusion Matrix:** Classification performance
4. **Feature Importance Bar Chart:** Key predictors

Each visualization helps interpret the data pattern and model performance more effectively.

Results and Discussion:

- The Decision Tree classifier achieved **87% accuracy** on the test set.
- **High recall** indicates the model can successfully identify most individuals with sleep disorders.
- The model revealed **stress level and low sleep duration** as key contributors to poor sleep health.
- Results suggest that healthy sleep duration, physical activity, and maintaining a low stress level are crucial factors for avoiding sleep disorders.

However, some limitations were observed:

- Dataset size is small; may not generalize well.
- Class distribution is slightly imbalanced (fewer “Sleep Apnea” cases).
- Model prone to overfitting; ensemble methods (Random Forest, XGBoost) could improve robustness.

Source Code:

https://colab.research.google.com/drive/1mw8KEHUNuHzDl4gxJgRs-XBH0TjBnRuK?usp=sharing

Conclusion:

In this project, a Decision Tree Classifier was successfully implemented to predict sleep disorders using the Sleep Health and Lifestyle dataset.

The model effectively classified individuals based on health and lifestyle indicators, achieving an accuracy of 87%. The analysis confirmed that stress level, sleep duration, and physical activity are key determinants of sleep health.