Academic Year	Module	Assessment Number	Assessment Type
2025	**Concepts and Technologies of AI**		Report

Classification Report

Student Id : 2407743
Student Name : Utkarsha – Aryal
Section : L5CG7
Module Leader : Siman Giri
Tutor : Siman Giri
Submitted on : 11-02-2025

UNIVERSITY OF WOLVERHAMPTON

HERALD COLLEGE KATHMANDU

Table of Contents

Abstract

Purpose: TO predict of someone has sleep disorder or not using categorical varables

Approach: The dataset called “Sleep Health and Lifestyle Dataset” is used which contains data related to sleep habits and lifestyle like a person profession and daily habits. Methodology includes data preprocessing , EDA ,building logistic regression model from scratch and making logistic regression from sickit learn and random forest from sicikt learn and making a final modal based on performance.

Key Result: The logistic regression done from scratch gave an accuracy: 0.45 the logistic regression done with sickit learn gave an accuracy of 0.806 and accuracy of random forest is 0.838

Conclusion: It can be seen that random forest has a far greater performance compared to other models.

Introduction:

Problem statements:

The goal to predict if someone has sleeping disorder according to their lifestyle and habits.

Dataset:

This dataset has 374 rows and 13 columns has it is related to the overall health of a person this is related to SDG 3: Good Health and Well

Objectives:

The primary objective of this models is to build a final classification model to predict if someone has a sleeping disorder.

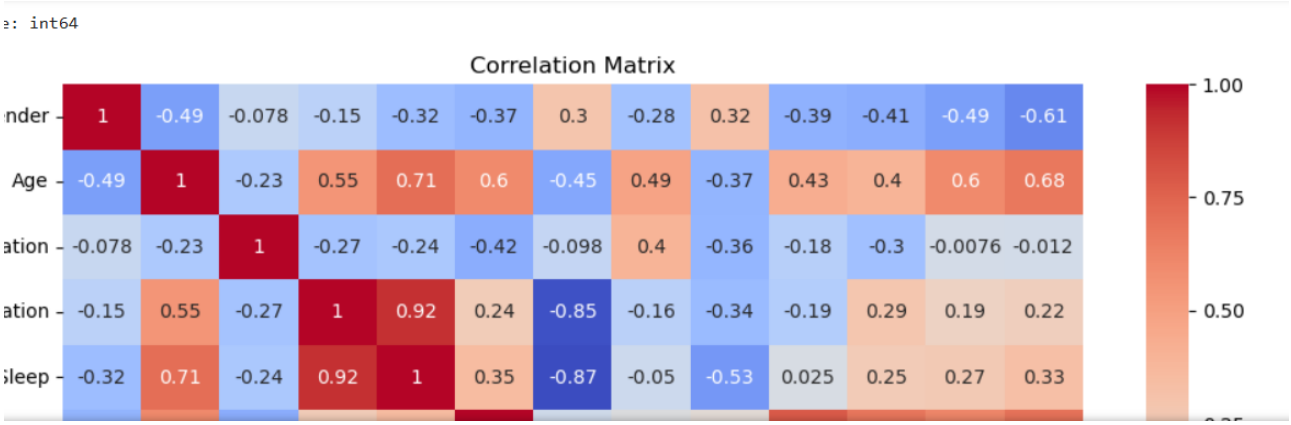
Methodology:

Data Pre-processing:

- There weren’t any missing information.
- Encoding: Some categorical values were encoded
- Train-test split: The data was separated into train and test
- The person Id was dropped
- The blood pressure was separated into two column systolic and diastolic bp.

Exploratory Data Analysis(EDA):

IT was found through EDA that sleep disorders across categories and correlations between sleep quality ,duration and stress levels.



Model Building:

Performance metrics include:

Model	Precision	Recall	F1-score	Accuracy
Logistic Regressiom	0.81	0.81	0.81	0.80
Random forest	0.84	0.84	0.84	0.84

Best Model: the best model is random forest.

Hyper parameter optimization:

GridSearchCv was used for hyperparameter optimization.

Feature selection:

Selected Features: ['Age', 'Physical Activity Level', 'Daily Steps', 'Systolic_BP', 'Diastolic_BP', 'Gender_1', 'Occupation_5', 'Occupation_7'] for both

Conclusion:

Key Findings:
Both the model performed well but random forest was the best.

Final Model:

The final model give a accuracy of 0.87, precision of 0.87, and recall of 0.87
And f1 score of 0.87. The final model performed far better than the previous model.