Project Report

Sleep Disorder Prediction

Ammaar Naeem Laghari(20P-0180)

Muhammad Ahmad(20P-0033)

1)Libraries Used: For this project, we have used various Python libraries for data analysis and modeling. NumPy is used for numerical computations, Pandas for data manipulation and analysis, Matplotlib and Seaborn for data visualization and plotting and Scikit-learn for machine learning algorithms.

2)Data Acquisition: "The Sleep Health and Lifestyle Dataset.csv" has been taken from Kaggle datasets. It comprises 400 rows and 13 columns, covering a wide range of variables related to sleep and daily habits.

3)Data Pre-Processing: The dataset was first checked for any missing values. Since there were no missing values, it was not required to fill them. Then the Person ID column was dropped as it did not contain any relevant information. The blood pressure column was split into two columns of systolic and diastolic blood pressure for better analysis. The categorical variables like gender, occupation, BMI category and sleep disorder were label encoded to convert them into numerical format suitable for machine learning models. This allowed handling different categories as numbers. Further, the data was split into train and test sets for building and evaluating the models. Various preprocessing techniques helped clean the data and prepare it for analysis and building predictive models.

4)EDA (Exploratory Data Analysis): Exploratory data analysis was conducted to understand variable distributions and their relation to sleep disorders. Count and distribution plots created using Matplotlib and Seaborn showed that most people were aged 30-45 with good sleep quality. Insomnia was higher in males while females had more sleep apnoea cases. Count plots on stress indicated its link to increased sleep disorders. A joint plot revealed obese/overweight individuals suffered more sleep disorders than normal weight. This initial analysis provided insights on how factors like gender, age, stress and BMI influence sleep conditions, helping identify meaningful patterns in the data.

5)Feature engineering: Some feature engineering was performed to refine the dataset for building predictive models. Since blood pressure was recorded as a single column, it was split into two features - systolic and diastolic blood pressure for more informative analysis. Categorical variables like gender, occupation and BMI category which were nominal were label encoded to convert them into numerical format acceptable to machine learning algorithms. This allowed handling different categories as integer values. Additionally, a new feature called 'Age Group' was created by binning the continuous 'Age' variable into meaningful categories. These engineering steps helped extract more meaningful representations from raw data for better capturing relationships between variables. It resulted in a more optimized dataset for training classification models.

6)Model Building: Two classification algorithms - Decision Tree Classifier and Random Forest Classifier were used to predict sleep disorders. The dataset was split into train and test sets. The Decision Tree Classifier was first applied to the training split and evaluated on the test set. This helped analyze the baseline performance of a simple model. Next, a Random Forest Classifier with 100 estimators was trained on the same train split. Both models predicted the 'Sleep Disorder' column which had 3 classes.

7)Model Evaluation: The predictions were compared with actual values using various metrics like accuracy, precision, recall and F1-score. The random forest model showed improved performance over decision tree, achieving an accuracy of 89% and F1-score of 0.86, demonstrating its effectiveness in predicting sleep disorders.

8) conclusion : From the exploratory data analysis, we have concluded that the sleep orders depends upon three main factors that are gender, occupation and BMI of the patient. The males have more instance of Insomia whereas femlaes have more instances of Sleep Apnea. In addition the that people with occupation such as nurses are more prone to sleep disorders. The BMI of the patient also plays a vital role in the prediction of sleep disorders. The patients who are either Obese or overweight are more prone to sleep disorders.