Sleep Disorder Prediction

The project aims to identify and predict sleep disorders using machine learning techniques. It analyzes the persons lifestyle and medical variables such as age, BMI, physical activity, sleep duration, blood pressure to predict sleep disorder and its types.

Dataset

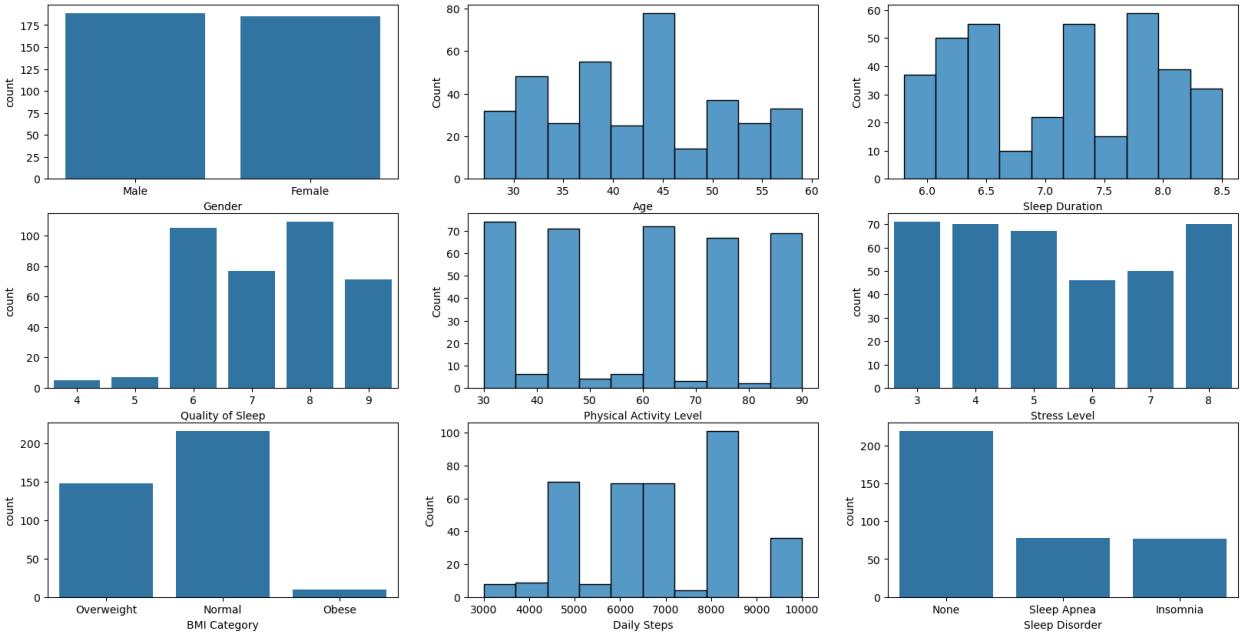
The dataset provides information related to sleep disorder and associated factors. It includes various attributes of individual such as:

- Demographic Information: Age and gender of each individual.
- Health Metrics: Body Mass Index (BMI), blood pressure and heart rate.
- Lifestyle Factors: Physical activity level, stress level and occupation.
- Sleep Characteristics: Sleep duration, quality of sleep, snoring frequency and presence of condition like insomnia, sleep apnea and normal sleep condition.

Data Dictionary

Column Name	Description
age	Age of the individual
gender	Gender of the individual
bmi	Body Mass Index (BMI) of the individual
snoring_frequency	Frequency of snoring during sleep
daytime_sleepiness	Level of daytime sleepiness reported
sleep_duration	Average sleep duration in hours
insomnia	Whether the individual experiences insomnia (Yes/No)
restless_legs	Whether the individual experiences restless legs (Yes/No)
sleep_apnea	Diagnosis of sleep apnea (Yes/No)
narcolepsy	Diagnosis of narcolepsy (Yes/No)

Exploratory Data Analysis



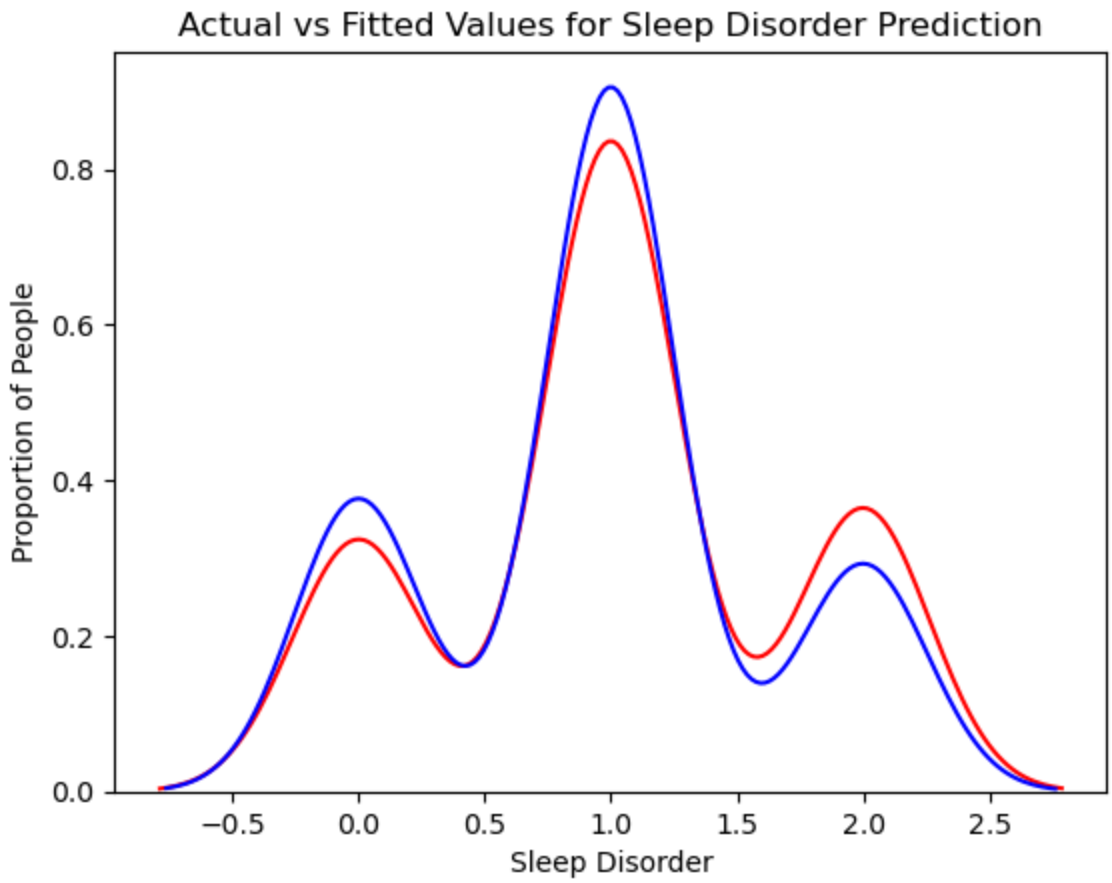
The number of males and females is almost equal, out of which majority of people have age between 30-45 years and most people have sleep duration of 6.5,7.5 and 8 hours. Most of the people have sleep quality greater than 5 which means there are getting sufficient sleep. Moreover, most of the people have normal BMI which directly relates with the distribution of sleep disorder which shows equal number of people with and without sleep disorder.

Model Building

For predicting sleep disorder through classification algorithm we used

- Decision Tree Classifier
- Random Forest Classifier

Distribution Plot for Predicted and Actual values



Classification Report

	precision	recall	f1-score	support
0	0.77	0.83	0.80	24
1	0.94	0.98	0.96	62
2	0.91	0.74	0.82	27
accuracy			0.89	113
macro avg	0.87	0.85	0.86	113
weighted avg	0.90	0.89	0.89	113

The Random Forest Classifier model has an accuracy of 89% and an avergae F1 score of 0.86. From the metrics it is quite clear that the model is able to predict the sleep disorder quite effectively, with increased accuracy than Decision Tree Classifer.

Conclusion

From the exploratory data analysis, we that the sleep orders depends upon three main factors that are gender, occupation and BMI of the patient. The males had more instance of Insomnia whereas females had more instances of Sleep Apnea. In addition the that nurses were prone to Sleep Apnea. The BMI of the patient also played a vital role in the prediction of sleep disorders. The patients who were either Obese were prone to sleep disorders.

Both decision tree and Random Forest models performed pretty good, however the Random Forest Classifier had excellent results with 89% accuracy.