Smt. Indira Gandhi College of Engineering
Computer Engineering Department
Ghansoli – Navi Mumbai
Academic Year 2023-24 (Even Sem)

Student Name: Khyati Garude **Roll No.:** 13 **Class:** BE **Sem:** VIII

Course Name: Applied Data Science Lab

Course Code: CSL8023

Experiment No. 01

Experiment Title:

Case Study Selection : "Sleep efficiency analysis" & Dataset Chosen : "Sleep efficiency Dataset". Illustrate The Data Science Life Cycle For The Same.

Date of Performance	Date of Submission	Marks (10)					Sign / Remark
		A	B	C	D	E	
		2	3	2	2	1	
11/1/24	18/1/24	2	3	2	2	0	20/1/24.
		Total Marks					
		09					

A: Prerequisite Knowledge
B: Implementation
C: Oral
D: Content
E: Punctuality & Discipline

Name - Khyati Garude
BE COMPS

DATE	EXPERIMENT - 01	SIGNATURE
18/1/24	Case Study Selection: "Prediction of Sleep Efficiency" And Dataset chosen: "Sleep Efficiency Dataset". Illustrate the Data Science Lifecycle for the same.	20/2/24

AIM: Case Study Selection: "Prediction of Sleep Efficiency" And Dataset chosen: "Sleep Efficiency Dataset". Illustrate the Data Science Lifecycle for the same.

THEORY

Sleep is a fundamental and complex physiological process that is essential for the overall health and well-being of individuals. It is a natural, recurring state of reduced consciousness, characterized by changes in brain activity, bodily functions, and responsiveness to the environment. While the exact purpose and mechanisms of sleep are not fully understood, it is universally recognized as a vital component of life.

The sleep-wake cycle is a dynamic and highly regulated phenomenon influenced by internal biological rhythms, external environmental factors, and individual lifestyle choices. During a typical night, individuals progress through different stages of sleep, including non-rapid eye movement (NREM) and rapid eye movement (REM) sleep. Each stage plays a unique role in physical restoration, memory consolidation, and emotional regulation.

The importance of sleep extends beyond mere rest, as it is intricately linked to cognitive functions, mood regulation, immune system function, and overall metabolic health. Chronic sleep deprivation or poor sleep quality can have significant consequences on physical and mental well-being, contributing to issues such as impaired concentration, mood disorders, compromised immune function, and an increased risk of chronic diseases.

Understanding the nuances of sleep is crucial for adopting healthy sleep habits and addressing potential sleep disorders. This involves recognizing the factors that influence sleep, implementing effective sleep hygiene practices, and seeking professional help when necessary. As a cornerstone of a balanced and healthy lifestyle, a good night's sleep is integral to promoting optimal daily functioning and long-term health.

Sleep efficiency is a measure that reflects the quality and effectiveness of your sleep during a specific period. A higher sleep efficiency percentage indicates that you are spending a significant portion of your time in bed asleep, suggesting better sleep quality. On the other hand, a lower sleep efficiency may be an indicator of sleep disturbances, such as difficulty falling asleep, frequent awakenings, or restless sleep.

While there is no universal benchmark for what constitutes good sleep efficiency, values around 85% or higher are generally considered healthy. Factors that can influence sleep efficiency include sleep disorders, lifestyle habits, stress, and environmental conditions. Monitoring sleep efficiency can be useful in assessing the overall effectiveness of your sleep and identifying areas for improvement in your sleep hygiene and routines.

A predictive model could be developed utilizing machine learning algorithms to forecast individuals' sleep efficiency based on a combination of demographic information, lifestyle factors, and sleep-related metrics. By training the model on the provided dataset, which includes features such as age, gender, caffeine and alcohol consumption, exercise frequency, smoking status, and sleep patterns (e.g., bedtime, wake-up time, sleep duration, REM sleep percentage), the model could learn complex relationships between these variables and sleep efficiency. The model could then be used to predict an individual's sleep efficiency score, providing valuable insights into factors influencing sleep quality and enabling targeted interventions to improve overall well-being. Through iterative refinement and validation, this predictive model has the potential to offer personalized recommendations for optimizing sleep habits and promoting better sleep health.

DATASET USED

This dataset encompasses a comprehensive set of attributes, combining demographic information, sleep-related metrics, lifestyle factors, and behaviors that can collectively contribute to a deeper understanding of sleep patterns and their influencing factors.

Subject ID

- Unique identifier for each test subject.

Age and Gender

- Recorded information about the age and gender of each test subject.

Bedtime and Wakeup Time

- Indicate the times when each subject goes to bed and wakes up each day, providing insights into their sleep schedule.

Sleep Duration

- Represents the total amount of time each subject spent asleep, measured in hours.

Sleep Efficiency

- A measure that represents the proportion of time spent in bed that is actually spent asleep. Calculated as the ratio of total sleep time to total time in bed.

REM Sleep Percentage, Deep Sleep Percentage, and Light Sleep Percentage

- Provide information about the percentage of time spent in each stage of sleep (REM, Deep, Light). These stages are crucial for understanding the quality and depth of sleep.

Awakenings

- Records the number of times each subject wakes up during the night, offering insights into the frequency of disturbances in their sleep.

Caffeine and Alcohol Consumption

- Information about the amount of caffeine and alcohol consumed by each subject in the 24 hours prior to bedtime. These factors can influence sleep quality.

Smoking Status

- Indicates whether a subject is a smoker or a non-smoker, as smoking habits can have implications for sleep health.

Exercise Frequency

- Provides details about how often each subject engages in physical exercise in a week. Regular exercise is known to impact sleep patterns positively.

The provided dataset exhibits missing values across various attributes, a common occurrence in real-world datasets. These missing values pose challenges to the reliability and accuracy of subsequent analyses. However, through effective data preparation techniques, such as imputation or removal of incomplete records, these missing values can be addressed. By systematically handling missing data, the dataset can be refined to ensure completeness and integrity, thus facilitating more robust and insightful analyses. This preparatory step is essential for deriving accurate conclusions and actionable insights from the data.

SAMPLE DATA

Sleep_Efficiency

ID	Age	Gender	Bedtime	Wakeup time	Sleep duration	Sleep efficiency	REM sleep percentage	Deep sleep percentage	Light sleep percentage	Awakenings	Caffeine consumption	Alcohol consumption	Smoking status	Exercise frequency
1	65	Female	2021-03-06 01:00:00	2021-03-06 07:00:00	6.0	0.88	18	70	12	0.0	0.0	0.0	Yes	3.0
2	69	Male	2021-12-05 02:00:00	2021-12-05 09:00:00	7.0	0.66	19	28	53	3.0	0.0	3.0	Yes	3.0
3	40	Female	2021-05-25 21:30:00	2021-05-25 05:30:00	8.0	0.89	20	70	10	1.0	0.0	0.0	No	3.0
4	40	Female	2021-11-03 02:30:00	2021-11-03 08:30:00	6.0	0.51	23	25	52	3.0	50.0	5.0	Yes	1.0
5	57	Male	2021-03-13 01:00:00	2021-03-13 09:00:00	8.0	0.76	27	55	18	3.0	0.0	3.0	No	3.0
6	36	Female	2021-07-01 21:00:00	2021-07-01 04:30:00	7.5	0.9	23	60	17	0.0		0.0	No	1.0
7	27	Female	2021-07-21 21:00:00	2021-07-21 03:00:00	6.0	0.54	28	25	47	2.0	50.0	0.0	Yes	1.0
8	53	Male	2021-08-16 00:30:00	2021-08-16 10:30:00	10.0	0.9	28	52	20	0.0	50.0	0.0	Yes	3.0
9	41	Female	2021-04-05 02:30:00	2021-04-05 08:30:00	6.0	0.79	28	55	17	3.0	50.0	0.0	No	1.0
10	11	Female	2021-09-16 01:00:00	2021-09-16 10:00:00	9.0	0.55	18	37	45	4.0	0.0	0.0	No	0.0
11	50	Male	2021-02-28 00:30:00	2021-02-28 08:30:00	8.0	0.92	23	57	20	1.0	50.0	0.0	Yes	3.0
12	55	Male	2021-03-10 22:30:00	2021-03-10 08:00:00	7.5	0.93	18	60	22	0.0	0.0	0.0	No	3.0
13	30	Female	2021-08-07 02:30:00	2021-08-07 11:30:00	9.0	0.93	24	58	18	0.0	50.0	0.0	No	1.0
14	28	Male	2021-02-25 01:00:00	2021-02-25 09:30:00	8.5	0.64	28	25	47	4.0	0.0	0.0	No	3.0
15	36	Female	2021-07-03 01:30:00	2021-07-03 10:00:00	8.5	0.54	20	32	48	2.0	25.0	1.0	Yes	0.0
16	32	Female	2021-12-06 22:00:00	2021-12-06 05:30:00	7.5	0.92	25	55	20	0.0	50.0	2.0	No	5.0
17	21	Female	2021-03-24 01:00:00	2021-03-24 08:00:00	7.0	0.54	26	22	52	4.0	0.0	2.0	Yes	0.0
18	40	Female	2021-08-27 23:00:00	2021-08-27 06:30:00	7.5	0.5	18	20	62	3.0	50.0	2.0	Yes	3.0
19	43	Female	2021-07-27 00:00:00	2021-07-27 08:00:00	9.0	0.98	20	67	13	0.0	25.0	1.0	No	0.0
20	52	Male	2021-12-03 00:30:00	2021-12-03 07:30:00	7.0	0.89	26	52	20		50.0	0.0	Yes	3.0
21	24	Male	2021-05-02 00:00:00	2021-05-02 08:00:00	8.0	0.83	15	75	10	3.0	0.0		No	2.0
22	32	Male	2021-06-14 02:30:00	2021-06-14 10:00:00	7.5	0.71	23	58	19	3.0	50.0	4.0	No	1.0
23	29	Female	2021-12-23 22:00:00	2021-12-23 08:00:00	8.0	0.84	23	60	17	2.0	75.0	0.0	No	2.0
24	63	Female	2021-05-21 02:30:00	2021-05-21 10:30:00	8.0	0.98	22	65	13	0.0	50.0	0.0	No	1.0
25	24	Male	2021-09-16 00:00:00	2021-09-16 07:00:00	7.0	0.88	15	75	10	1.0		5.0	No	2.0
26	52	Male	2021-03-26 00:00:00	2021-03-26 08:00:00	8.0	0.91	18	72	10	1.0	0.0	0.0	No	2.0
27	36	Female	2021-06-18 00:00:00	2021-06-18 07:00:00	7.0	0.95	28	55	17	0.0		0.0	No	1.0
28	35	Male	2021-12-15 02:30:00	2021-12-15 09:30:00	7.0	0.84	24	60	16	2.0	0.0	0.0	Yes	2.0
29	23	Male	2021-11-08 21:30:00	2021-11-08 08:30:00	9.0	0.85	27	20	53	3.0	0.0	0.0	No	3.0

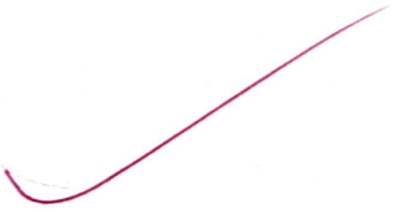
DATA SCIENCE LIFECYCLE

1. **Data Preprocessing**
   - Clean the data: Handle missing values, remove duplicates, and correct any inconsistencies.
   - Convert categorical variables into numerical ones using techniques like one-hot encoding or label encoding.
   - Scale or normalize numerical features to ensure they're on a similar scale.
   - Split the data into features (X) and the target variable (y), where y represents sleep efficiency.
2. **Model Training**
   - Train a Random Forest Regressor model on the preprocessed data.
   - Random Forest is chosen for its ability to handle complex relationships and feature interactions, and it's robust to overfitting.
3. **Model Evaluation**
   - Evaluate the trained Random Forest model using appropriate evaluation metrics for regression tasks, such as Mean Absolute Error (MAE), Mean Squared Error (MSE), or R-squared (R2) score.
   - Split the dataset into training and testing sets, and use the testing set to assess the model's performance on unseen data.
4. **User Input**
   - Set up a user interface where users can input their sleep-related data, such as bedtime, wake time, sleep duration, and potentially other factors.
   - Validate and preprocess the user input data similarly to the training data.
5. **Selection of Most Important Features using Random Forest**
   - Extract feature importances from the trained Random Forest model.
   - Identify the most important features contributing to sleep efficiency.
   - This step helps to understand which factors have the most significant impact on sleep quality.
6. **Random Forest Implementation**
   - Implement the Random Forest Regressor algorithm using a suitable library like scikit-learn in Python.
   - Tune hyperparameters such as the number of trees, maximum depth, and minimum samples per leaf to optimize model performance.
7. **Deployment**
   - Deploy the trained Random Forest model into a production environment, such as a web application or mobile app, where users can input their sleep data and receive predictions on sleep efficiency.
   - Ensure scalability, reliability, and security of the deployment environment.

8. **Monitoring and Maintenance**
    - Set up monitoring tools to track the performance of the deployed model in real time.
    - Monitor key metrics such as prediction accuracy, response time, and system health.
    - Implement a maintenance plan to periodically retrain the model with new data to keep it up-to-date and improve performance over time.
    - Address any issues or drift in model performance promptly through updates or retraining.

CONCLUSION

In summary, examining a sleep efficiency dataset offers valuable insights into the factors influencing sleep quality. The data helps identify patterns and correlations, informing personalized interventions for improved sleep. However, it's crucial to acknowledge and address potential biases in the dataset. Ongoing research and collaboration are essential for refining our understanding of sleep patterns and developing effective interventions to enhance overall well-being.



20/3/24.