# Sleep Apnea Prediction Based on Health and Lifestyle Factors

# Objective

# To build a machine learning model that predicts whether an individual is likely to have sleep apnea using health, demographic, and lifestyle features. This can assist in early identification of individuals at risk, improving timely medical intervention

# Dataset Used

# Name: Sleep Apnea Sleep Health Dataset

**Key Features Include:**

* Age
* Alcohol/Smoking Addiction
* Daily Steps
* Heart Rate
* Stress Level
* Sleep Duration
* Sleep Quality
* Quality of Sleep
* Screen Time
* Screen Time (before Sleep)
* Physical Activity Level
* Systolic and Diastolic Blood Pressure
* BMI Category (Normal Weight, Overweight, Obese)
* Sleep Disorders (e.g., Insomnia, Parasomnia, Restless Leg Syndrome)
* Sleep Walking
* Occupation
* Gender

**Target Variable:** Sleep Disorder\_Sleep Apnea

# Model Chosen

For this project, we used:

1. **Logistic Regression** – As a baseline for binary classification.
2. **K-Nearest Neighbors (KNN)** – For comparison using distance-based classification

# Performance Metrics

1. **Accuracy**: The proportion of correctly predicted outcomes out of all predictions.
2. **Precision**: The proportion of true positives among all predicted positives.
3. **Recall**: The proportion of true positives identified out of all actual positives.
4. **F1-Score**: The harmonic means of precision and recall for balanced performance.
5. **Confusion Matrix**: A table showing true/false positives and negatives to evaluate model performance.

* **Model Evaluation Results:**

1. For Logistic regression
   * 1. Accuracy: 1.0
     2. Precision: 1.0
     3. Recall: 1.0
2. For K-Nearest Neighbors
   * 1. Accuracy: 0.7866666666666666
     2. Precision: 0.80
     3. Recall: 0.73

* **Challenges**
* Handling class imbalance, especially with the **Sleep Apnea** class.
* Selecting the most informative features from a mix of lifestyle, physiological, and categorical variables.
* Normalizing and encoding features like **Alcohol/Smoking Addiction**, **Physical Activity Level**, and **BMI Category**.
* Maintaining interpretability while using models like **KNN** that are sensitive to feature scaling.
* **Learnings:**
* Gained hands-on experience in preprocessing structured health data, including encoding categorical features like **BMI Category** and **Sleep Disorder** types.
* Learned how variables such as **Sleep Duration**, **Stress Level**, **Screen Time before Sleep**, and **Heart Rate** correlate with sleep apnea risk.
* Implemented and compared **Logistic Regression** and **KNN** for binary classification of **Sleep Disorder\_Sleep Apnea**.
* Discovered how scaling features such as **Daily Steps**, **Systolic/Diastolic BP**, and **Screen Time** improves model accuracy, especially for KNN.
* Strengthened evaluation skills using **accuracy**, **precision**, **recall**, **F1-score**, and **confusion matrix**.

# Conclusion

This project applied machine learning techniques to predict the presence of **sleep apnea** using a diverse set of health and behavioral features, such as **stress level**, **physical activity**, **sleep quality**, and **screen time before bed**. Both **Logistic Regression** and **KNN** models were trained and evaluated, with Logistic Regression providing more stable and interpretable results.