# **Capstone Two: Predicting Sleep Disorders Final Project Report**

## **Introduction**

This project aimed to develop and evaluate a predictive model for identifying sleep disorders based on health and lifestyle factors. Sleep disorders, such as insomnia, significantly impact the quality of life and overall health. This project seeks to aid healthcare professionals in the early detection and management of sleep disorders by leveraging machine learning techniques.

## **Data Exploration and Preprocessing**

The dataset used in this project includes information on sleep duration, physical activity levels, stress levels, BMI categories, and other relevant metrics. Key preprocessing steps included handling missing values, encoding categorical variables (e.g., one-hot encoding), and scaling numerical features using StandardScaler. Exploratory Data Analysis (EDA) provided insights into the distribution of sleep disorders and correlations between variables, guiding subsequent modeling decisions.

## **Model Selection and Evaluation**

A Random Forest classifier was chosen for its ability to handle complex relationships and provide robust predictions. The model was trained and evaluated using the following metrics on a test dataset:

* **Accuracy:** Measures overall correctness of predictions.
* **Precision:** Indicates the proportion of true positive predictions out of all positive predictions.
* **Recall:** Measures proportion of actual positives correctly identified.
* **F1-score:** Harmonic mean of precision and recall, providing a balanced measure.
* **ROC-AUC score:** Evaluates the model's ability to distinguish between classes.

## **Hyperparameter Tuning**

GridSearchCV was employed to optimize the Random Forest model by tuning parameters such as the number of estimators, maximum depth, minimum samples split, and minimum samples leaf. This process identified the best set of hyperparameters that improved model performance.

## **Conclusion**

In conclusion, the Random Forest model achieved promising results with an accuracy of approximately 87%, indicating its effectiveness in identifying sleep disorders based on health and lifestyle factors. Precision, recall, and F1-score metrics of around 85.7% suggest balanced performance in classifying insomnia cases. The ROC-AUC score of 0.87 confirms the model's ability to distinguish between classes. Future work may include further refinement of features, exploring additional algorithms, and validating the model on diverse datasets to enhance generalizability and robustness.