Sleep Quality Prediction

Objective

The objective of this project is to develop a machine learning model to predict sleep quality as "Good" or "Bad" based on various lifestyle and health-related factors. By analyzing these factors, we aim to provide insights into the key determinants of sleep quality and help individuals improve their sleep habits.

Dataset Used

The dataset used in this project is the **"Sleep Health and Lifestyle Dataset."** It contains multiple features related to sleep patterns, physical activity, and health indicators. The key attributes in the dataset include:

- **Age**
- **Gender**
- **Occupation**
- **Sleep Duration**
- **Physical Activity Level**
- **Stress Level**
- **BMI Category**
- **Heart Rate**
- **Daily Steps**
- **Sleep Disorder**
- **Quality of Sleep** (Target Variable)

Model Chosen

For this project, we selected **Logistic Regression** as the primary classification model due to its simplicity and effectiveness in binary classification problems. The steps involved in model training include:

1. **Data Preprocessing**: Handling missing values, encoding categorical variables, and scaling numerical features.
2. **Feature Selection**: Identifying key variables that contribute to sleep quality.
3. **Model Training**: Splitting the dataset into training and test sets and fitting the Logistic Regression model.

4. **Prediction & Evaluation**: Assessing the model's performance using various evaluation metrics.

Performance Metrics

The model was evaluated using the following metrics:

- **Accuracy Score**: Measures the overall correctness of predictions.
- **Classification Report**: Provides details on precision, recall, and F1-score.
- **Confusion Matrix**: Visual representation of correct and incorrect predictions.
- **Mean Squared Error (MSE) & Root Mean Squared Error (RMSE)**: Evaluates error in predictions.

Model Evaluation Results:

- **Accuracy:** 0.97
- **Precision, Recall, F1-score:** 0.96,0.98,0.98
- **Confusion Matrix:** Displayed as a heatmap for visualization.

Challenges & Learnings

Challenges:

- Handling **categorical variables** effectively without losing meaningful information.
- Managing **class imbalance**, ensuring fair representation of both "Good" and "Bad" sleep quality.
- **Feature scaling** to prevent biases due to varying numerical ranges.
- Identifying and addressing **outliers** in the dataset that might skew predictions.

Learnings:

- **Feature Engineering** plays a crucial role in improving model performance.
- **Scaling & Encoding** significantly affect the effectiveness of machine learning models.
- **Exploratory Data Analysis (EDA)** helps in understanding data patterns and correlations.
- Evaluating multiple models and metrics ensures better decision-making for predictive analysis.

Conclusion

The Logistic Regression model successfully predicts sleep quality with good accuracy. Further improvements can be made by exploring advanced models such as Decision Trees, Random Forest, or Neural Networks to enhance prediction accuracy. Future work can also involve incorporating real-time sleep tracking data for more robust analysis.