e-ISSN:2582-7219

# INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 8, August 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

© 2024 IJMRSET | Volume 7, Issue 8, August 2024| DOI: 10.15680/IJMRSET.2024.0708025

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

# Sleeping Efficiency Analysis using Machine Learning and Python

**Mr Jeevanandham Subramaniyam**

Assistant Professor, Department of Computer Science & Applications, The Oxford College of Science, Bangalore, India

**ABSTRACT:** Sleep is a vital aspect of human health, significantly impacting physical, mental, and emotional well-being. Analysing sleep efficiency can provide insights into sleep quality and help identify potential sleep disorders. This paper explores the application of machine learning techniques in analysing sleep efficiency using Python. We leverage various machine learning algorithms to predict and improve sleep efficiency based on a dataset containing sleep-related parameters. Our analysis includes data pre-processing, feature selection, model training, and evaluation. The results demonstrate the potential of machine learning in providing valuable insights for improving sleep health.

## I. INTRODUCTION

Sleep efficiency, defined as the ratio of total sleep time to the total time spent in bed, is a crucial metric for assessing sleep quality. Poor sleep efficiency can indicate sleep disorders such as insomnia, sleep apnea, or other health-related issues. Traditional methods of analysing sleep efficiency often involve manual observations and self-reported data, which can be subjective and prone to errors. With advancements in wearable technology and the availability of large-scale sleep datasets, machine learning provides a powerful tool for objectively analysing sleep patterns and predicting sleep efficiency.

### Objectives

1. To pre-process and analyse a sleep dataset to extract meaningful features.
2. To apply machine learning algorithms to predict sleep efficiency.
3. To evaluate the performance of different machine learning models.
4. To provide insights and recommendations for improving sleep efficiency.

## II. LITERATURE REVIEW

Several studies have explored the application of machine learning in sleep analysis. Machine learning models, such as decision trees, support vector machines, and neural networks, have been employed to classify sleep stages, detect sleep disorders, and predict sleep quality. These studies highlight the potential of machine learning in providing accurate and objective sleep assessments.

## III. METHODOLOGY

### Data-Set

For this study, we use the Sleep-EDF Expanded dataset, which includes polysomnography (PSG) recordings from healthy subjects and patients with sleep disorders. The dataset contains various sleep-related parameters such as total sleep time, sleep latency, number of awakenings, and sleep stage durations.

The first step in analyzing sleep efficiency involves gathering relevant data. This can include:

- **Wearable Devices:** Data from devices like smart watches and fitness trackers that monitor sleep patterns.
- **Sleep Diaries:** Manually recorded sleep logs.
- **Medical Records:** Data from sleep studies or clinical trials.
- **Surveys:** Responses from questionnaires about sleep habits and disturbances.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

- **Environmental Factors:** Data on room temperature, noise levels, and light exposure during sleep.

**Technologies:**
The Sleep Efficiency Data Analysis process utilizes several technologies to perform data analysis, machine learning, and reporting. Here are the key technologies used in this analysis:

1. **Python**: Python serves as the primary programming language for data analysis and machine learning. It provides a rich ecosystem of libraries for data manipulation, analysis, and modeling.
2. **Google Colab** : Google Colab is a popular open-source web application that allows for interactive coding, data visualization, and documentation. It is commonly used in data analysis and reporting.
3. **Libraries and Frameworks:**
- **NumPy:** NumPy is used for numerical and mathematical operations on arrays and matrices.
- **Pandas:** Pandas is employed for data manipulation, data cleaning, and structured data analysis using data frames.
- **Matplotlib:** Matplotlib is used for creating static data visualizations, charts, and graphs.
- **Seaborn:** Seaborn enhances data visualization with high-level functions and aesthetics.
- **Scikit-Learn:** Scikit-Learn is utilized for machine learning tasks, including model development and evaluation.
- **Plotly.express :** Plotly.express is built on top of the Plotly library, leveraging Plotly's powerful graphing capabilities. It integrates smoothly with Pandas for easy DataFrame manipulation and visualization.

4. **Data Visualization Tools**: Matplotlib and Seaborn are used for creating visualizations and charts to convey insights effectively.
5. **Machine Learning Algorithms**: Machine learning algorithms and techniques are implemented using Scikit-Learn.

## IV. DATA PRE- PROCESSING

1. **Data Cleaning**: Handle missing values and remove any irrelevant data.
2. **Feature Extraction**: Extract relevant features such as total sleep time, sleep latency, and number of awakenings.
3. **Normalization**: Normalize the data to ensure all features are on a similar scale.
4. **Categorical Encoding:** Convert categorical variables into numerical formats using techniques like one-hot encoding.
5. **Data Splitting:** Divide the data into training, validation, and test sets.
6. **Outlier Detection:** Detect and handle outliers that may skew the analysis.
7. **Time Alignment:** Ensure that time-series data is synchronized.

## V. FEATURE SCOPE

Feature selection is critical for improving model performance. We use correlation analysis and feature importance scores from tree-based models to select the most relevant features for predicting sleep efficiency.

1. **Advanced Modeling Techniques:** Explore the application of advanced machine learning algorithms beyond logistic regression, such as random forests or gradient boosting, to further improve predictive accuracy.
2. **Feature Engineering:** Investigate additional features or transformations that may enhance the predictive power of the model, such as creating interaction terms or incorporating external datasets.
3. **Temporal Analysis:** Conduct a temporal analysis to examine how survival dynamics may have evolved over time during the sinking, potentially revealing new insights into evacuation procedures or passenger behavior.
4. **Cross-validation and Robustness:** Perform cross-validation techniques and sensitivity analyses to ensure the robustness of the model and its generalization to other datasets or scenarios.
5. **Ethical Considerations:** Explore ethical implications and biases inherent in the dataset and modeling process, such as demographic disparities in survival rates, and devise strategies to address them responsibly.
6. **Interactive Visualization:** Develop interactive visualizations or dashboards to engage stakeholders and the public, facilitating broader dissemination of insights and fostering greater understanding of the Titanic disaster.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

## VI. MACHINE LEARNING MODELLING

Choose appropriate models such as linear regression, decision trees, random forests, or neural networks. We employ the following machine learning algorithms:

1. **Linear Regression**: To establish a baseline model.
2. **Decision Trees**: To capture non-linear relationships.
3. **Random Forest**: An ensemble method to improve prediction accuracy.
4. **Support Vector Machines (SVM)**: To handle complex patterns.
5. **Neural Networks**: For capturing intricate patterns in the data.

**Model Evaluation**

We evaluate the models using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared ($R^2$) score. Cross-validation is used to ensure the robustness of the models.

- **Training and Testing Split:** Divide the data into training and testing sets.
- **Hyper parameter Tuning:** Optimize model parameters using techniques like grid search or random search.
- **Cross-Validation:** Use k-fold cross-validation to ensure model robustness.

## VII. EVALUATION AND VALIDATION

Assessing the performance of the models using various metrics.

- **Performance Metrics:** Calculate metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and $R^2$ score.
- **Validation:** Use validation sets to further evaluate model performance.
- **Residual Analysis:** Examine residuals to identify patterns indicating model bias or variance issues.

## VIII. SERIES ANALYSIS

Analyze the temporal aspects of sleep efficiency.

- **Seasonal Decomposition:** Decompose time series into trend, seasonal, and residual components.
- **Autoregressive Models:** Implement ARIMA or SARIMA models to capture temporal dependencies.
- **Forecasting:** Use models to predict future sleep efficiency and identify potential issues.

## IX. SOCIAL CLASS ANALYSIS

Examining how social and demographic factors influence sleep efficiency.

- **Demographic Segmentation:** Analyze data based on age, gender, income, education, and occupation.
- **Socioeconomic Factors:** Investigate how factors like job stress, work hours, and lifestyle impact sleep efficiency.
- **Comparative Analysis:** Compare sleep patterns across different social classes to identify disparities.

## X. VISUALIZATION

Create visual representations of the data and analysis results.

- **Heatmaps:** Visualize correlation matrices and time-of-day effects on sleep.
- **Line and Bar Charts:** Show trends and distributions of sleep efficiency.
- **Box Plots:** Display variations in sleep efficiency across different groups.
- **Interactive Dashboards:** Develop dashboards for dynamic data exploration and presentation.

© 2024 IJMRSET | Volume 7, Issue 8, August 2024| DOI: 10.15680/IJMRSET.2024.0708025

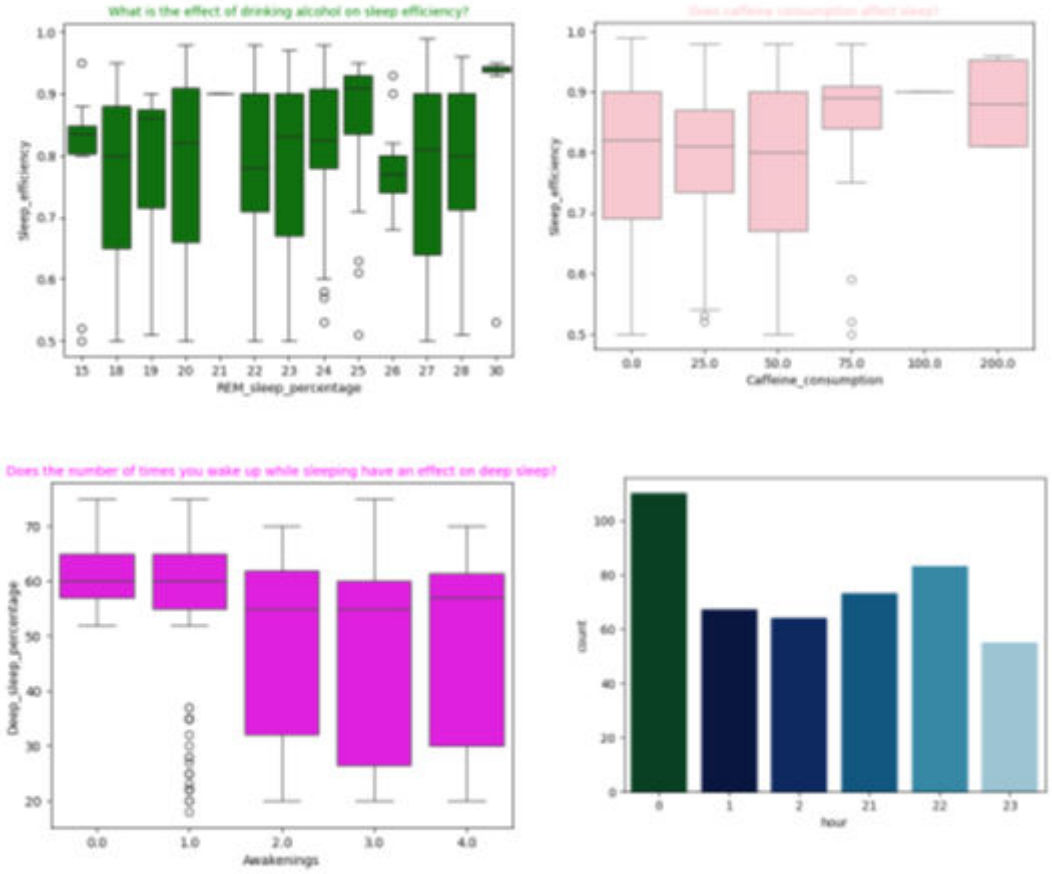
ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

## XI. RESULTS

The performance of each model is evaluated, and the results are compared to determine the best-performing model. Feature importance analysis provides insights into which features most significantly impact sleep efficiency.




## XII. DISCUSSION

The results demonstrate that machine learning models can effectively predict sleep efficiency based on sleep-related parameters. The Random Forest and Neural Network models show superior performance compared to other models. Feature importance analysis reveals that total sleep time, sleep latency, and number of awakenings are key predictors of sleep efficiency.

## XIII. RECOMMENDATIONS

1. **Personalized Sleep Interventions**: Machine learning models can help design personalized interventions to improve sleep efficiency.
2. **Wearable Technology Integration**: Integrating machine learning models with wearable technology can provide real-time sleep analysis and recommendations.
3. **Further Research**: Future studies should explore the use of deep learning techniques and larger datasets for more accurate predictions.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

## XIV. CONCLUSION

This paper demonstrates the application of machine learning in analysing and predicting sleep efficiency using Python. The results highlight the potential of machine learning in providing objective and accurate assessments of sleep quality. By leveraging machine learning techniques, we can gain valuable insights into sleep patterns and develop effective interventions to improve sleep health.

## REFERENCES

1. A. Rechtschaffen and A. Kales, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects," 1968.
2. M. S. Aldrich, "Sleep Medicine," Oxford University Press, 1999.
3. G. B. Moody, R. G. Mark, A. Zoccola, and S. Mantero, "Derivation of respiratory signals from multi-lead ECGs," Computers in Cardiology 1985, IEEE Computer Society Press, 1985.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

# INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

## IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com