Comparative Analysis of Machine Learning Models for Sleep Disorder Prediction

Sarah Goodyear1, Abdallah Alsammani1,\*

1Department of Mathematics, Jacksonville University, Jacksonville, FL 32211

[sgoodye@jacksonville.edu](mailto:sgoodye@jacksonville.edu), [aalsamm@ju.edu](mailto:aalsamm@ju.edu)

# Abstract

Insomnia and sleep apnea represent sleep disorders that profoundly affect human health and well-being. This study analyzes multiple lifestyle and physiological parameters by examining how machine learning (ML) techniques predict sleep disorder risks. Our analysis used a dataset containing multiple health metrics, including sleep duration, sleep quality, physical activity levels, stress levels, and cardiovascular health indicators. Several classification frameworks were created and evaluated to enable machine automated detection of sleep disorders. Our analysis determined that decision tree, random forest models, and k-nearest neighbors (kNN) emerged as the most effective models after preprocessing data thoroughly, selecting features, and applying cross-validation. The decision tree model demonstrated superior performance with an AUC of 0.97 in our comparative evaluation. In contrast, the random forest model achieved an AUC of 0.96 and the kNN model attained an AUC of 0.95. Our feature importance analysis discovered that age combined with BMI category, blood pressure, and occupation emerged as the top predictors of sleep disorders. ML promises to improve early detection and treatment of sleep disorders in medical settings.

## Keywords

*Sleep disorders, machine learning, decision tree, random forest, k-nearest neighbors, prediction models.*

# 1. Introduction

Supervised Learning is the simplest type of machine learning where algorithms learn from labeled training data and map features to known output variables. The goal of these algorithms is to model a dataset to generalize well on unseen data. If these algorithsm can be implemented for decision support, elevating the breadth of what doctors can perform without these resources, ML could be a valuable tool for improving medical prognosis on a widescale. Specifically, exploring how ML methods can predict and diagnose the presence of a sleep disorder compared to current approaches, the potential efficacy of this method can be gauged.

# 2. Literature Review

Multiple prior studies have investigated how different ML techniques can be used for sleep disorder classification. One notable study uses the same Sleep Health and Lifestyle dataset from Kaggle which includes 374 synthetic samples on sleep patterns, health metrics, and lifestyle factors. Eight machine learning algorithms were used for classification including Light Gradient Boosting Machine (LGBM), AdaBoost, Random Forest (RF), Extra Trees, k-Nearest Neighbors (kNN), Logistic Regression (LR), Gaussian Naïve Bayes (GNB), and Gradient Boosting (GB). All features were used, and the target variable was sleep disorder which could have been either none, insomnia, or sleep apnea. Accuracy scores were used to evaluate and compare the models. A bivariate analysis was also conducted.

Based on the accuracy scores, the top performing model was the Gradient Boosting (GB) with an accuracy of 92%, as well as having the highest f1-score. This led to the conclusion that this model has the highest predicting power out of the eight tested models in predicting sleep disorders. Logistic regression and AdaBoost closely followed with a 91% accuracy and comparable f1-scores, indicating comparable predictive power. Random Forest with gini had an accuracy of 90% and a strong f1-score as well. The remaining models had weaker f1-scores and moderately comparable accuracies, with the exception of Gaussian Naïve Bayes with an accuracy of 53% which performed the worst. The most impactful features were age, BMI, and physical activity level.

The bivariate analysis revealed patterns in gender, BMI, age, and occupation with sleep disorder. Females, those will higher BMIs, and those who are older in age were revealed to be more prone to having a sleep disorder. Certain professions including professors, nurses, and those in sales display higher likelihood of having a sleep disorder compared to professions like doctors, engineers, and lawyers.

# 3. Methodology

Examining the contents of this dataset reveals all samples are unique and do not have any missing data. However, the python data frame interprets the “None” in “Sleep Disorder” column, representing no sleep disorder as clarified in the metadata, as a NaN value. Replacing these incorrectly labeled NaN values in the “Sleep Disorder” column with “No disorder” resolves this misinterpretation.

After cleaning the data, dummy variables were created for the categorical features so both numeric and categorical features could be normalized. They were normalized on a scale of 0 to 1. Normalization was more ideal than standardization for this study to limit the impact of the skew and any outliers present in the features, by placing the values in a fixed range. Normalization also allows the data to be more easily interpreted by a wider audience, as well as having the potential to perform better on a wider range of machine learning models. Particularly, k-Nearest-Neighbors (kNN) and many unsupervised techniques are more applicable when data is fixed on a bounded scale. Normalizing the data allows for a more thorough analysis of supervised techniques and prepares for a follow-up analysis and comparison with unsupervised techniques.

Next, a correlation matrix was developed to aid in feature selection. Person ID is included in the matrix, but these metrics should be disregarded in analysis since this feature is an arbitrary value to distinguish uniqueness and holds no statistical significance. The warmer the tone, the more positive the correlation and the cooler the tone, the more negative the correlation. Age, occupation, physical activity level, stress level, BMI category, blood pressure, and heart rate are observed to have positive correlations with sleep disorder. Conversely, gender, sleep deprivation, quality of sleep and daily steps are observed to have negative correlations with sleep disorder.

A screenshot of a chart

AI-generated content may be incorrect.

# 4. Results

The three models selected for training based on cross validation results were the decision tree, random forest, and k-nearest-neighbors (kNN). For each of these models, parameter optimization was performed prior to training. Each model was trained using the same four aforementioned selected features, as well as the same train-test split. The test size was 20% of the data, leaving 80% for training, and the random state was set to 42. After each model was trained, predictions were made using the test set, the accuracy was computed, confusion matrices were produced, classification reports were printed, and auROC curves were created. These evaluation metrics were stored for each model for final comparison.

To evaluate the results of each model, a confusion matrix and area under the receiving operating characteristics curve (auROC) were developed. Classification reports showing precision, recall, and f1-scores were also created. These metrics give insight on how well each model can distinguish between classes and what classification types are accounting for the error in predicting the test set.

Based on the confusion matrix, the decision tree model had both high precision and recall, with slightly higher likelihood of a false negative result than a false positive result. This results in a high f1 score of 0.95, as well as a strong ROC curve with an AUC of 0.961.

Like the decision tree, the confusion matrix indicates that the random forest model had both high precision and recall, with false negatives being slightly more prone than false positives. An identical f1 score was achieved of 0.95, while a slightly inferior, though still strong, ROC AUC score of 0.958 was achieved.

Analysis of the kNN model indicates similar performance to the decision tree and random forest. Again, confusion matrix results show a slightly higher tendency to present false negatives than false positives, though precision and recall overall are high. The kNN model had an identical f1-score of 0.95 to the other models. However, the kNN model achieved a slightly lower ROC AUC score of 0.957.

# 5. Discussion

Given the statistically close performance of these models, the ROC AUC metric is the most suitable for comparison. This metric summarizes the auROC curve into a single metric, most amplifying the small differences between the other model metrics. The accuracy, f1 score, and ROC AUC score for the kNN, decision tree, and random forest models are shown in the table below. As mentioned, the accuracy and f1 scores for these models are so similar there is no statistically significant difference, if there is one at all. This is further highlighted in the bar plots below where the difference in bar length is not visually apparent for any of the metrics, including ROC AUC. Though the ROC AUC scores are close, they do highlight the minimal differences in performance as seen in the last column of the table. These scores indicate that the decision tree model performs the best, followed by random forest and then kNN.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Accuracy | F1 Score | ROC AUC |
| Decision Tree | 96 | 95.2 | 96.1 |
| Random Forest | 96 | 95.2 | 95.8 |
| kNN | 96 | 95.2 | 95.7 |

A graph of a curve

AI-generated content may be incorrect.

# 6. Conclusion

The results indicate that the decision tree, random forest, and kNN models all performed well in predicting the presence of a sleep disorder using lifestyle and physiological data. Given the close results, the ROC AUC score, being the most sensitive metric to variation in performance, is the best metric to distinguish between the models. Each had high accuracy and f1-scores, but the decision tree model performed the best based on ROC AUC scores with a score of 0.961, though the others closely follow. This makes the decision tree model the most optimal recommendation for predicting the preens of a sleep disorder with physiological and lifestyle data. Further study is required to determine how predictive power could be maximized with other model types or datasets that are different in size or features.

# References

M. Kumar, B. Ahmed, H. M. Mishra, A. K. Jha, P. K. Sikarwal and S. Rampal, "Predictive Sleep

Disorder Modelling: Using Machine Learning with Lifestyle and Sleep Health Data," 2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2024

Sleep health and lifestyle dataset, <https://www.kaggle.com/datasets/uom190346a/sleep->

health-and-lifestyle-dataset, (2023).