Analysis of Sleep Disorder Classification Using Daily Habits

Buse Uzun
Undergraduate Student
Department of Statistics
Middle East Technical University
Ankara, Türkiye
buse.uzun@metu.edu.tr

***Abstract*—This paper aims to classify sleep disorders existence based on the daily habits of the participants. Since sleep health is an important topic for the physical and psychological well-being of society, it is valuable to grasp the primary factors for future public health strategies.**

For classification, Logistic Regression is used as a statistical model, while Artificial Neural Networks, Support Vector Machines, Random Forest, XGBoost, CatBoost, and LightGBM are used as machine learning models. Finding the best model for this classification is a great interest in this paper. Hence, several performance metrics are compared, such as F1 scores and accuracy. Feature importances of the best resulting model were also obtained.

Keywords— Machine Learning Models, Statistical Models, Exploratory Data Analysis, Feature Engineering, sleep

I. INTRODUCTION

The key to the well-being of an individual may come from many aspects of life, but it is no surprise that having a good night's sleep has a considerable effect on physical and psychological health. According to a study conducted on Portuguese adults, chronic sleep disorders have an effect on mental, social and physical problems as well as low life satisfaction. They have also found that concentration and memory problems, irritability, and decision-making difficulties are directly associated with sleep issues (Oliveira et al., 2024). Thus, understanding the underlying factors of sleep issues such as sleep apnea and insomnia is a crucial step for public health improvement.

This paper analyses the lifestyle choices of participants such as occupation, sleep duration, physical activity level, etc. to classify having a sleep disease or not. Several statistical and machine learning methods, such as Logistic Regression, Artificial Neural Networks (ANN), Support Vector Machines (SVM), Random Forest (RF), XGBoost, CatBoost, and LightGBM, are applied to achieve a suitable model. The performances of models are compared with metrics such as accuracy or F1 scores. Having several methods enables a comprehensive analysis of the classification of the target variable.

As well as machine learning analysis, exploratory data analysis (EDA) and confirmatory data analysis (CDA) are obtained for a deeper understanding of the data. Before analyses, data preprocessing and feature engineering are also handled. In the end, feature importances of the best model are obtained to see the weights of contributing factors. R-studio and Python are utilized for preprocessing and analysis.

II. METHODOLOGY

A. Dataset

The dataset of this analysis was obtained from Kaggle, and it contains variables related to daily habits and sleep. The dataset covers 4 main interests: comprehensive sleep metrics, lifestyle factors, cardiovascular health, and presence of sleep disorders. The data consists of 374 participants and 13 variables. The 'Person ID' variable is removed since participants will not be addressed individually. There are 481 NA values, which is around 10% of the whole data.

The features can be inspected more detailly below.

1. *Age: Age of the participants, continuous*
2. *Gender: Female = 1, Male = 2, categorical*
3. *Occupation: Occupation of the participant, categorical*
4. *Sleep_Duration: Sleep hours of participants per day, continuous*
5. *Sleep_Quality: Subjective rating of sleep quality, Low = 1, Moderate = 2, Good = 3, categorical (ordinal)*
6. *Physical_Activity_Level: The amount of physical activity of participants in minutes, continuous*
7. *Stress_Level: Subjective stress level of participants, Low = 1, Moderate = 2, High = 3, categorical (ordinal)*
8. *BMI_Category: BMI category of participants, Normal Weight = 1, Overweight = 2, categorical*
9. *Heart_Rate: Resting heart rate of the participants in beats per minute (bpm), continuous*
10. *Daily_Steps: The number of daily steps of participants, continuous*
11. *Sleep_Disorder: Shows whether the participant has insomnia, sleep apnea or none, No = 1, Yes = 2, binary*
12. *Blood_Pressure_C: Blood pressure of the participants. Computed by dividing systolic pressure by diastolic pressure, High Stage 1 = 1, High Stage 2 = 2, Moderate = 3, categorical*

B. Descriptive Statistics

For a deeper understanding of the features, descriptive statistics can be inspected. These statistics give several information such as minimum and maximum values, first and third quartiles, means, medians and number of NA values.

Descriptive statistics summaries of numerical and categorical data can be inspected separately for ease of understanding.

	Age	Sleep Duration	Activity Level	Heart Rate	Daily Steps
Min.	27	5.8	30	65	3000
1st Qu.	35	6.4	45	68	6000
Median	43	7.2	60	70	7000
Mean	42.27	7.128	59.49	70.23	6821
3rd Qu.	50	7.8	75	72	8000
Max	59	8.5	90	86	10000
Std. Dev.	8.78	0.8	20.95	4.18	1619.8
NA's	37	37	37	37	37

Table 1 Summary of Descriptive Statistics of Numerical Data

Standard deviations of the features are also calculated in addition to the summary statistics and included in the table above.

From *Table 1,* it is seen that the minimum age of the participants is 27 while the maximum age is 59, and the mean age is 42.27. Sleep duration of the participants ranges between 5.8 to 8.5 hours, the average sleeping hour is 7.13, and the median is 7.2. The closeness of the mean and the median can be an indicator of the normal distribution of the variable. Also, its standard deviation is 0.8. The lowest engaged physical activity is 30 minutes, and the greatest is 90 minutes. Mean and median are close in this feature as well. The heart rate variable suggests symmetry since the mean and median are around 70. Daily steps variable ranges between 3000 to 10000, with the median of 7000 and the average of 6821, there can be slight left skew since the median is greater, but the difference is modest. The standard deviation of daily steps is 1619.776. Lastly, all of the numerical variables have 37 NA values.

The tables below show summaries of descriptive statistics of categorical data.

Levels	Gender	BMI Category	Sleep Disorder
1	164	194	192
2	173	143	145
NA's	37	37	37

Table 2 Summary of Descriptive Statistics of Binary Data

Levels	Stress Level	Blood Pressure	Sleep Quality	Occupation
1	63	237	5	64
2	171	59	104	63
3	103	41	166	57
4			62	39
(Other)				114
NA's	37	37	37	37

Table 3 Summary of Descriptive Statistics of Categorical Data

Upon inspecting *Table 2*, the frequencies of the levels can be seen. There are 164 female and 173 male participants. BMI categories include 194 normal weight and 143 overweight people. Most importantly, the sleep disorder column, which is the dependent variable, has 192 people who do not have a sleep disorder and 145 people who do have a sleep disorder (either insomnia or sleep apnea). This can indicate that there is no class imbalance in the data.

Table 3 shows that the most chosen stress level by participants was moderate, with 171 people. It was followed by a high stress level with 103 people, and a low stress level with 63 people. For the blood pressure variable, most of the participants fall into the high stage 1 category. The sleep quality of the participants is mostly good, with 166 people. Finally, there are 11 different occupations of participants, with the top 3 being nurse, doctor, and engineer. There are 37 NAs in every feature.

C. Exploratory and Confirmatory Data Analysis

This section will answer some of the exploratory data analysis questions and analyze them with confirmatory data analysis for each question. There will be 3 questions in total.

1. *What is the relationship between sleep disorders and physical activity level?*

Violin plot is used since sleep disorders are categorical, while daily physical activity is continuous.

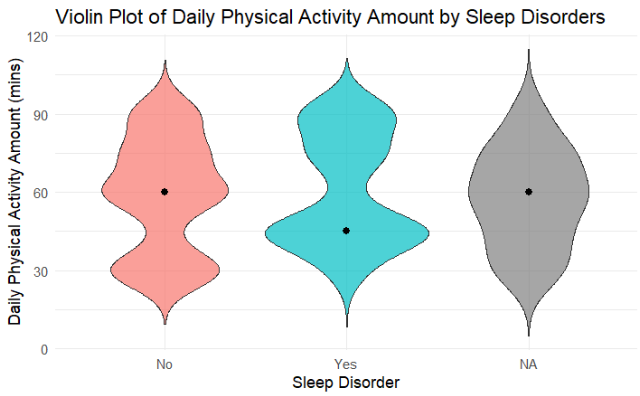

Fig. 1 Violin Plot of Daily Physical Activity and Sleep Disorder

The violin plot in *Fig. 1* shows that both not having a sleep disorder and having a sleep disorder with respect to the physical activities participants engaged in are not normal. Attendants who have a sleep disorder have a lower median in physical activity, around 45 minutes. The median of physical activity is 60 minutes for participants who do not have a sleep disorder. NA values in sleep disorders seem to have a unimodal distribution.

Before inspecting the difference between classes, some assumptions need to be checked to be able to use the two-sample independent t-test. Normality is checked with the Shapiro-Wilk Normality Test; normality is not satisfied since the p-value is smaller than 0.05. Non-parametric tests can be used in this case. Levene's test is used to check the homoscedasticity anyway, since the p-value is greater than 0.05 homogeneity of variance assumption is satisfied. The Wilcoxon Rank-Sum test is used because normality was not satisfied.

The output of the Wilcoxon-Rank-Sum test can be observed from *Table 4* below; the p-value is 0.046. Even though it is close to 0.05, it is still slightly smaller. Hence, the null hypothesis can be rejected. Therefore, there is a significant difference in the physical activity levels in a day between those with a sleep disorder or not. From the previous violin plot, it was already seen that participants who have sleep disorders have a lower median in physical activity in a day.

Wilcoxon rank sum test	
data: Physical_Activity_Level by Sleep_Disorder	
W = 15011	p-value = 0.04656

Table 4 Wilcoxon Rank-Sum Test Summary

2. *What is the relationship between sleep disorders and BMI categories?*

Since both variables are categorical, mosaic plot is used to show the relationship.

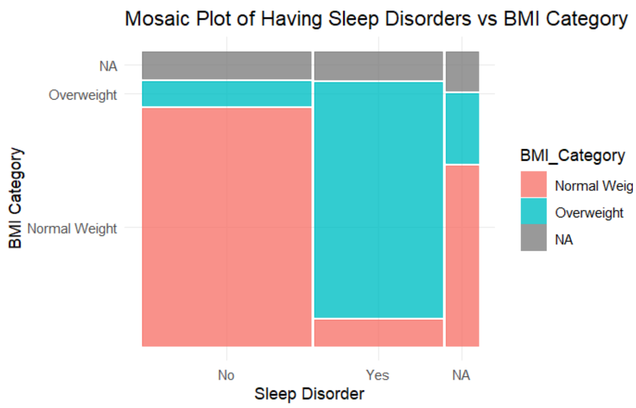

Fig. 2 Mosaic Plot of Having Sleep Disorders and BMI Category

From the mosaic plot in *Fig. 2*, it is seen that most of the overweight individuals tend to have sleep disorders (sleep apnea or insomnia) while normal-weight participants do not have sleep disorders for the most part. Chi-square test can be applied for the confirmatory analysis part.

Now, assumptions should be checked. The expected count table can be seen below.

	Sleep Disorder	
	Yes	**No**
Normal Weight	124.75	90.25
Overweight	92.25	66.74

Table 5 Expected Count Table

Since none of the expected counts are below 1 chi-square test is valid to use.

Pearson's Chi-squared test		
data: BMI_Category by Sleep_Disorder		
X-squared = 231.27	df = 1	p-value < 2.2e-16

Table 6 Pearson's Chi-squared Test Summary

From the chi-square test in *Table 6*, the null hypothesis is rejected since the p-value is less than 0.05. Which means there is a significant association between BMI categories and sleep disorders. This association was also seen in the visualization as the mosaic plot showed that while more overweight participants have sleep disorders than normal-weight participants, it is the opposite for normal-weight participants. This result is also seen in *Table 7*.

	Yes	**No**
Normal Weight	197	18
Overweight	20	139

Table 7 Frequency Table of Sleep Disorder and BMI Category

D. *Missingness*

The original data had no missing values. Thus, missing values were generated manually. After this step, around 10% of the data had missing values. In detail, there were 444 NA values, which means 37 missing values for each variable.

Before modelling, missing parts of the data should be inspected to see the missingness mechanism, and a suitable strategy to move forward should be developed accordingly. There are several missingness mechanisms, such as Missing at Random (MAR), Missing Completely at Random (MCAR), or Missing Not at Random (MNAR). These mechanisms have different patterns and statistical tests that enable one to differentiate them.

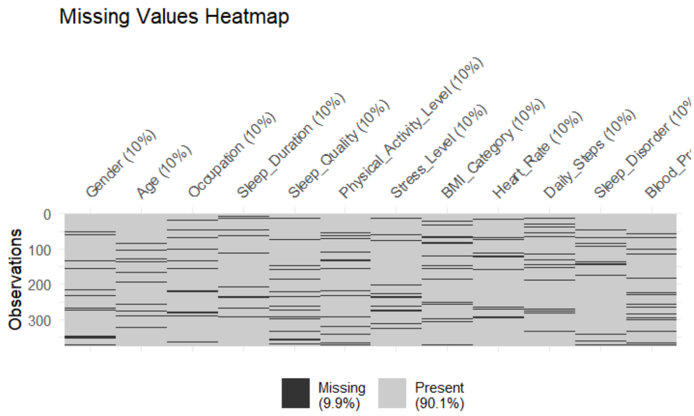

Fig. 3 Missing Values Heatmap

First, the missingness pattern can be inspected from *Fig. 3*. From the plot, there does not seem to be any patterns since the black parts, which show missing values, are scattered randomly across present values. In this case, MCAR can be suggested, but there is still a possibility of the mechanism being MAR if the missingness depends on another variable.

Statistical tests are used to see if the data has an MCAR pattern. First, the Hawkins test is conducted with the null hypothesis data have MCAR. Since its p-value (0.0006) is less than 0.05, the null hypothesis is rejected. Hawkins rejects MCAR, but it assumes normality. On the other hand, a non-parametric test is also applied since the dataset violates normality. P-value of the non-parametric test is 0.086; it fails to reject MCAR. Since Hawkins is not suitable due to assumptions, it can be concluded that the dataset has an MCAR pattern.

There are no visible missingness patterns for individual variables when their plots are inspected as well, an example can be seen below for the BMI category variable in *Fig. 4*.

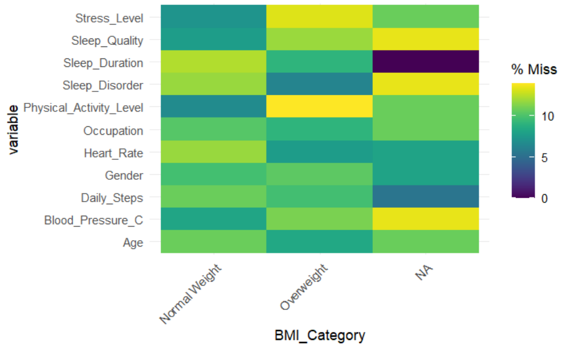

Fig. 4 Missingness Pattern of BMI Category

Now, as it is certain that the missingness comes from MCAR, they can be imputed. Non-missing part of the data will be used to impute the missing part of the data. Deletion methods are not favorable for our data since the data size itself is quite small. Hence, missing data was imputed from the "mice" package in RStudio. After the imputation, descriptive statistics of the original and imputed dataset were compared. Neither the ranges nor medians changed in continuous variables.

When kernel density plots are observed, it is seen that the imputed and original data overlays, which is good. An example is included in *Fig. 5*, similar results were obtained from other variables.

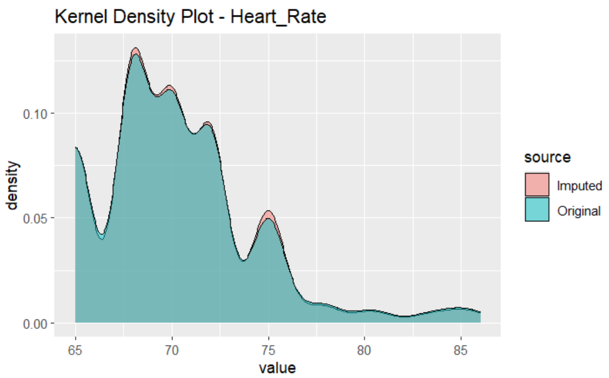

Fig. 5 Kernel Density Plot of Heart Rate

E. Modelling

a. Cross Validation

To enable testing the performance of the models, the data was split into train and test sets. 20% of the imputed data was taken as test and the remaining 80% was taken as train. The train set will be used to train the models, while the test set will be used to show performance metrics.

Class imbalance should be inspected and handled before modelling. Frequency tables of the target variable, which is sleep disorder, should be observed for this.

	Sleep Disorder	
	No	**Yes**
Train Set	58%	42%
Test Set	58.1%	41.9%

Table 8 Class Proportions of Train and Test Sets

Since class proportions of the target variable are quite close, it can be concluded that there is no class imbalance as seen in *Table 8*. The same proportions are reserved in the train and test sets.

1. Multiple Logistic Regression

Logistic regression is used for statistical modelling to classify the sleep disorder variable. First, correlations of the numerical features were checked to see if there is multicollinearity. It was seen that daily steps and physical activity level variables have a 0.74 correlation, which can cause multicollinearity. But there were only one pair close to 0.8, therefore PCA was not applied.

Then, a generalized linear model with binomial family was fitted. There was a problem since none of the variables were significant, VIF values were checked to see which variables were causing the problem and if there was multicollinearity. Variables with VIF values above 5 or 10 can be a candidate; in this case, all the VIF values were high, which ensured the multicollinearity existence. Sleep quality and stress level variables were removed since their VIF values were much greater than the others. In addition to this, daily steps were also removed since the correlation plot showed a 0.8 correlation with physical activity level (many combinations had been tried to get the best results). Finally, the model was fitted again, and the only occupation variable was removed due to a problematic VIF value. There was no longer a multicollinearity problem after removing the aforementioned 4 variables. The final model was conducted with only significant variables, which can be seen in *Table 9*.

	Estimate	Std. Error	Z value	Pr(>|z|)
(Intercept)	-13.72	3.82	-3.59	0.0003
BMI Cat. Overweight	4.49	0.42	10.75	< 2e-16
Heart Rate	0.16	0.05	2.95	0.003
Null deviance: 408.18, Residual deviance: 170.3, AIC: 176.3, Median: -0.26				

Table 9 Summary of Multiple Logistic Regression

From *Table 9,* it is seen that the intercept, BMI category, and heart rate variables are all significant since their p-values are less than 0.05. Median is close to zero, which is good.

It should be kept in mind that logistic regression estimates should not be considered directly, their exponential should be taken for direct interpretation. Hence, from the estimates, being overweight compared to normal weight increases log-odds by 4.48. Thus, overweight people are 88.5 times more likely to have a sleep disorder than normal-weight people, assuming all other factors are equal. One unit increase in heart rate causes a 17% increase in odds of having sleep disorders. These results were also observed in the EDA and CDA parts.

Performance Metrics	Train Set	Test Set
Accuracy	0.91	0.85
Precision	0.88	0.83
Recall	0.90	0.80
F1 Score	0.89	0.82
ROC AUC	0.93	0.86
Kappa	0.81	0.69

Table 10 Performance Metrics of Logistic Regression

From *Table 10,* when the accuracy of the train set is checked, it is seen that 91% of the overall predictions are correct, which is higher compared to the test set. This can indicate mild overfitting. Precision, which is the proportion of correctly classified positives is 88% on the train and 83% on the test set. Recall shows correctly classified true cases, it is good in both cases, but it is again high on the training set. The area under the curve (ROC AUC) shows the performance of how good the model can separate the two classes, it seems good. Lastly, the Kappa value shows that the model is good, the train set being a little higher might show overfitting. No information rate is also checked; it is observed that the cases which are no are 58% of the samples in both sets.

Generally, confusion matrix results are good in both cases. It should be kept in mind that higher train accuracy might indicate overfitting, but it does not seem severe.

2. Artificial Neural Networks (ANN)

ANN is used as a supervised machine learning approach for classification. Data was min-max scaled before feeding it to the model for more accurate results. Hyperparameter tuning was applied by grid search. For network architecture, different numbers and sizes of hidden layers are tried, such as (16,), (32,), (16,8), and (32, 16). ReLU and tanh are considered as activation functions. To minimize overfitting, dropout rates of 0.2 and 0.3 are considered. 16 and 32 are evaluated as batch sizes. 50, 100, and 300 epochs are applied. Finally, as learning rates, 0.01 and 0.001 are tested.

After training the model on the train set, the best parameters were obtained as such; *hidden_layers: (16,), activation: 'tanh', dropout_rate: 0.2, batch_size: 32, epoch: 300, learning_rate: 0.01*

Performance Metrics	Train Set	Test Set
Accuracy	0.93	0.919
Precision	0.91	0.93
Recall	0.93	0.87
F1 Score	0.92	0.90
ROC AUC	0.95	0.92

Table 11 Performance Metrics of ANN

From *Table 11,* the performance metrics of the ANN seem quite well. Accuracy is good in both sets and seems close, even though the train set is slightly greater. Precision and F1 score are greater in the train set, while precision is greater in the test set. Overall, all the metrics, including the area under the curve, are good. There can be mild overfitting, but metrics are quite close on train and test, so it is unlikely.

3. Support Vector Machine (SVM)

SVM is also applied with grid search to obtain the best-performing classification on sleep disorders. Again, min-max scaling was applied.

Kernels such as 'linear', 'poly', 'rbf', and 'sigmoid' are evaluated. Cost hyperparameter is considered with 0.01, 0.1, 1, and 10 values. The cost parameter was reduced due to overfitting. The gamma parameter had 'scale', 'auto', 0.001, 0.01, and 0.1 as candidates, and lastly, n was taken as 10, 50, and 100 as the number of SVM classifiers in bagging. Since bagging reduced the overfitting issue of SVM, it was applied. Also, 5-fold cross-validation was applied.

Best parameters are obtained as: *C: 0.01, gamma: 'scale', kernel: 'sigmoid', n: 10*

Performance Metrics	Train Set	Test Set
Accuracy	0.92	0.89
Precision	0.89	0.87
Recall	0.93	0.87
F1 Score	0.91	0.87
ROC AUC	0.93	0.91

Table 12 Performance Metrics of SVM

From *Table 12,* it is observed that accuracy is 92% on train and 89% on test, which is good but not the best. F1 score or accuracy can both be checked since there is no class imbalance. The F1 score is not good in the test set at all.

4. Random Forest Classifier

Similarly, Random Forest is used with different parameter combinations by grid search to classify using decision trees. Data was scaled prior to training. In grid search, the maximum depth of each decision tree was chosen as 'None', 10 or 20. For the number of considered features, 'sqrt' or 'log2' was evaluated. The minimum number of samples required at a leaf node was taken as 1 or 2. Considered values for the minimum number of splits were 2, 5, or 10. Trees get simpler as more samples are split. Lastly, 100 and 200 were evaluated as the number of decision trees.

Training the model with these parameter candidates led to the following best parameter combination: *max_depth: 'None', max_features: 'sqrt', min_samples_leaf: 2, min_samples_split: 10, n_estimators: 100*

Performance Metrics	Train Set	Test Set
Accuracy	0.94	0.919
Precision	0.92	0.93
Recall	0.93	0.87
F1 Score	0.92	0.90
ROC AUC	0.97	0.88

Table 13 Performance Metrics of Random Forest

When *Table 13* is observed, it is seen that performance metrics are good. Like ANN, except precision, all the other metrics, especially recall, are slightly greater on the train set. Metrics are better than models like SVM, accuracy and F1 scores do not have a big enough gap for serious overfitting.

5. XGBoost

XGBoost is applied for a supervised machine learning approach with the Optuna framework for model optimization, this approach enables searching across a range of parameters. Scaling was applied prior to classification. For this model, parameter candidates were between 100 to 500 for the number of boosting iterations. 3 to 6 for maximum tree depth. From 0.01 to 0.1 for the learning rate. 0.6 to 1 for subsampling ratio. Between 0.6 and 0.9 for the feature sampling ratio per tree. 1-10 for L1 regularization and 5-10 for L2 regularization. 5-fold cross-validation was applied.

Best parameters are obtained as such: *n_estimators: 447, max_depth: 5, learning_rate: 0.0737, subsample: 0.608, colsample_bytree: 0.891, reg_alpha: 8.492, reg_lambda: 6.062*

Performance Metrics	Train Set	Test Set
Accuracy	0.91	0.85
Precision	0.88	0.93
Recall	0.90	0.81
F1 Score	0.89	0.82
ROC AUC	0.94	0.89

Table 14 Performance Metrics of XGBoost

When *Table 14* is observed, the performance metrics seem moderate. The train set is better than the test set, which suggests a mild overfitting.

6. LightGBM

LightGBM is applied in a similar fashion to XGBoost modelling. Again, several parameters were considered within a range, and scaling was applied before training. Detailed parameter explanations can be found under XGBoost modelling. Besides them, the maximum number of leaves in one tree (num_leaves) and the minimum number of samples for each leaf (min_child_samples) are also optimized for controlling model complexity and overfitting. Also, 5-fold cross-validation was used again.

Best parameters for LightGBM: *n_estimators: 406, max_depth: 5, learning_rate: 0.012, subsample: 0.721, colsample_bytree: 0.651, reg_alpha: 1.585, reg_lambda: 14.488, num_leaves: 30, min_child_samples: 82*

Performance Metrics	Train Set	Test Set
Accuracy	0.91	0.85
Precision	0.88	0.83
Recall	0.90	0.81
F1 Score	0.89	0.82
ROC AUC	0.94	0.88

Table 15 Performance Metrics of LightGBM

As seen in *Table 15*, the performance metrics of LightGBM are very similar to the XGBoost model. Accuracy and F1 scores are moderate, there seems to be overfitting since train performance is greater than test performance.

7. CatBoost

Similar to XGBoost and LightGBM, CatBoost was applied as supervised machine learning classification with the Optuna framework. 5-fold cross-validation and scaling was applied. Iterations were considered from 100 to 500. Minimum depth of each tree ranged from 3 to 10. Learning rate ranged from 0.01 to 0.3. Parts of training samples in each iteration were between 0.6 to 1. Randomly chosen features for each tree ranged between 0.6 to 1. Lastly, values from 0 to 10.

Best parameter for CatBoost: *iterations: 106, depth: 7, learning_rate: 0.016, subsample: 0.996, colsample_bylevel: 0.864, reg_lambda: 6.972*

Performance Metrics	Train Set	Test Set
Accuracy	0.93	0.88
Precision	0.91	0.84
Recall	0.93	0.87
F1 Score	0.92	0.86
ROC AUC	0.96	0.89

Table 16 Performance Metrics for CatBoost

Table 16 shows better metrics for CatBoost compared to XGBoost and LightGBM. There might be overfitting. Still accuracy and F1 scores are not better than ANN or Random Forest. Performance did not get better when different parameter ranges were evaluated.

b. Model Performance Comparison and Feature Importance

Performance metric tables of the previously mentioned models are compared to decide the outperforming model. There is one crucial point in this step, only comparing train or test metrics would be misleading since overfitting and underfitting cannot be inspected in this case. Either accuracy or f1 score can be evaluated since there is no class imbalance. When train accuracy and F1 scores are compared, it seems like Random Forest outperforms ANN since Random Forest accuracy is 0.94 while ANN accuracy is 0.93. However, when test accuracies and F1 scores are compared, which is more important for new datapoints, it is seen that they are exactly the same. Both have 0.919 test accuracy. Thus, Random Forest overfits the data slightly more than ANN. For these reasons, ANN is chosen as the best-performing model.

The feature importance of the best-performing model is obtained as seen in *Fig. 6*.

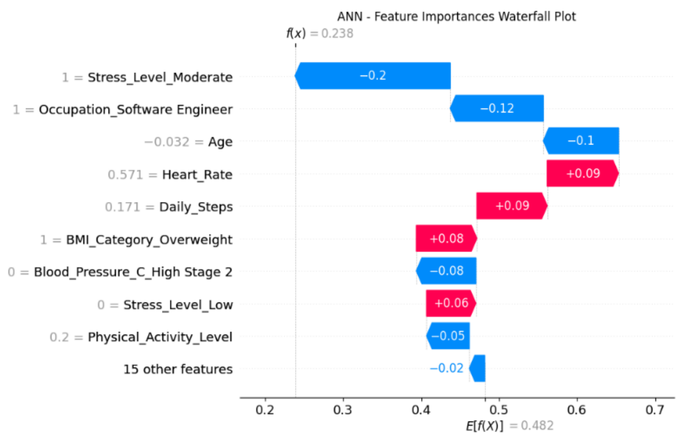

Fig. 6 Feature Importance Plot of ANN

Shapley Additive Explanations (SHAP) was used for obtaining feature importances rather than the commonly used weight-based feature importance since it fails to capture non-linear activation functions and complex interactions and does not work well with ANN architecture. Thus, SHAP is applied. When *Fig.6* is observed, it shows how each feature contributes to the model output. Blue bars represent negative SHAP contributions, which reduce the predicted value, while red bars represent positive contributions. Moderate stress level and software engineering occupation have the largest negative effect on the prediction, which means they decrease the model's positive class confidence. On the contrary, heart rate and daily steps variables have a positive contribution, making the prediction higher towards the positive class. In the plot, each variable's SHAP value is either added to or subtracted from the cumulative prediction.

III. Conclusion

This paper aimed to classify sleep disorders (having sleep apnea or insomnia) based on the lifestyles of participants. Several research questions are answered with exploratory data analysis and then analyzed with confirmatory data analysis. Feature engineering is applied, as well as understanding the missing data mechanism, to impute missing data. Train-test split is used for cross-validation to be able to see performance metrics. Statistical and machine learning approaches are used for the classification of the data. For the statistical method, logistic regression is used. For machine learning methods, ANN, SVM, RF, XGBoost, LightGBM, and CatBoost are used with hyperparameter tuning to get the best performance from models. Models are compared with several performance metrics such as accuracy, F1 score, recall, etc. The variable importance of the best-performing model is given. In conclusion, ANN gave the best results. Moderate stress level, software engineering as occupation, age, heart rate, and daily steps variables have the most importance in the model for classifying sleep disorder. With these kept in mind, new public health strategies can be considered.

IV. Discussion

Some suggestions can be made to improve the analysis. Due to compute unit limitations, some parameter values could not be evaluated, as well as limited epoch numbers.

Dataset size can be expanded since 374 is not big enough for training the models with the best performance. In future, some ensemble models can also be considered. Different cross-validation approaches can also be used, which can improve results. For logistic regression, PCA can be applied instead of removing high VIF values. Some new features can be considered, or the existing features can be reduced, depending on the aim.

This study enables further research to be done in the field of sleep health in individuals.

References

[1] Oliveira, S., Agostinis Sobrinho, C., Martins, S., Augusto, C., Araújo, O., Vieira, T., Macedo, A. P., Silva, M. J., & Rosário, R. (2024). Sleep disorders and subjective well-being in Portuguese adults: Evidence from a representative study. *Sleep Science*, *17*(01). https://doi.org/10.1055/s-0043-1772812

[2] Siddiqui, A. (2019, August 9). *Neural feature importance*. Medium. https://medium.com/data-science/neural-feature-importance-1c1868a4bf53

[3] Tharmalingam, L., (2023), *Sleep health and lifestyle dataset* [Dataset]. Kaggle. https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset/data