PROJECT

ALEENA LIJU

IMSCCS(DA)

Roll.no:4

ABSTRACT

Sleep efficiency is defined as the percentage of time we spend asleep in bed. It is a sleep behaviour characteristic that can be used to identify a possible sleep disorder. If most ofthe time is spent on bed and can't fall asleep it shows a disturbed sleep patterns. A healthy sleepefficiency ranges between 85 and 95%. A value of 90 to 95% should be aimed for. This meansthat if you sleep for 8 hours, you will fall asleep in about 30 minutes. The aim of this project isto predict the sleep efficiency based on certain factors like alcohol consumption, caffeine consumption, exercise frequency etc. We can also analyse various sleep patterns followed by the test subjects. Project also aims to build various regression models and find the best model with greatest accuracy.

INTRODUCTION

Sleep efficiency is one of many metrics used to assess how well or poorly we sleep.When combined with other factors (duration, latency, and sleep quality), it can provide a more complete picture of our overall sleep health. Sleep efficiency is defined as the percentage of time we spend asleep in bed. Time spent in bed not trying to sleep, such as reading, does not count. You can calculate your sleep efficiency by dividing the amount of time you spend sleeping by the total amount of time you spend in bed. Some people are simply more adept at falling asleep than others. Many factors influence how well we sleep. Diet, exercise, stress, anxiety, and lifestyle are all obvious factors that influence our ability to sleep. Other factors are less obvious. An efficient sleep leads to deeper, higher-quality sleep with fewer interruptions. It may cause feelings of energy and well-being upon awakening, whereasinsufficient sleep may cause feelings of tiredness and restlessness. Extra time spent in bed is not recommended in order to achieve good sleep efficiency. If a person spends the majority oftheir time in bed actually sleeping, they are considered sleep efficient (or to have a high sleep efficiency). However, if a person spends the majority of their time in bed awake, this is not considered sleep efficient (or the person has a low sleep efficiency). This is common in cases of insomnia. Insomnia is defined by having difficulty falling or returning to sleep. As the condition worsens, more of the time spent lying in bed is spent awake. Examining sleep efficiency can help you understand sleep problems. This measurement may also prompt changes in sleep habits that improve sleep depth and quality. A sleep efficiency of 85 percent or higher is considered normal, while a sleep efficiency of 90 percent or higher is considered very good. A sleep efficiency of less than 85% is considered poor and indicates that an individual needs to sleep more efficiently.

A) Data Description

The dataset used in this project is obtained from the Kaggle. The details provided in the datasetare used to predict the sleep efficiency of each test subject. The dataset contains 452

instanceswith 14 features. Among them two are categorical. The target variable used is sleep efficiencywhich is a numeric variable. The features used in this dataset are:

• Sleep Efficiency-It is the percentage of time that a person spent asleep while in bed. It is calculated by dividing the time spent asleep by the total time in bed.

• Bed time-The time when each subject goes to bed.

• Wakeup time-The time each subject wakes up each day.

• Sleep duration-Total amount of time each test subject slept in hours.

• REM sleep percentage-The night's REM cycle begins about 90 minutes after you fall asleep and repeats every 90 minutes. Your eyes move quickly behind your eyelids, and your brainwaves resemble those of an awake person. Your heart rate, blood pressure, and breathingrate all increase to near-waking levels.

• Light sleep percentage-When you sleep, your body goes through several sleep cycles, each of which has four stages. The first of these is light sleep. These constitute roughly half of a typical night's sleep. This is usually the start of the sleep cycle, as we move from full wakefulness to light sleep.

• Deep sleep percentage-Deep sleep is particularly beneficial to brain health and function. Thisstage of sleep allows the brain to rest and recover while replenishing energy. It also aids in thereinforcement of declarative memory, or the recall of facts.

• ID-Unique ID to identify the tset subject.

• Gender-Gender of each test subject.

• Awakenings-Number of times each test subject wakes up during the night.

• Caffeine consumption-Records its consumption in the 24 hours prior to the bed time.

• Alcohol Consumption-Records the no.of times alcohol is consumed in the 24 hours proor tothe bed time.

• Exercise Frequency-Denotes the exercise frequency level of each test

subject.Smoking status-Denotes whether each test subject smoke or not.

• Age-Age of each test subject.

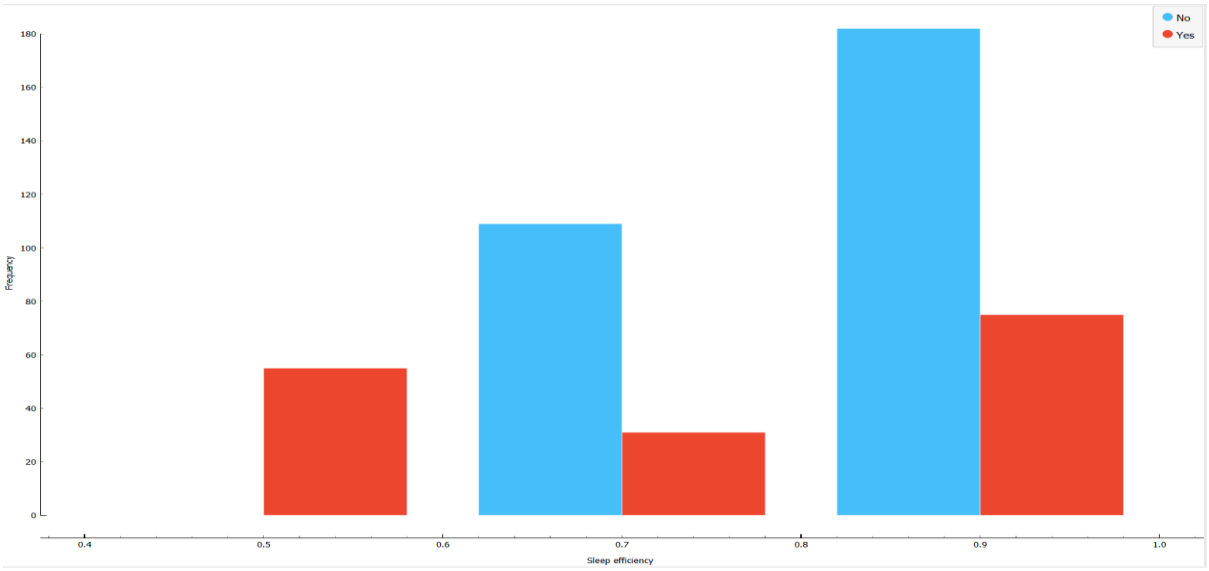
MATERIALS AND METHODS

Tool used: Orange

Orange is a data visualisation, machine learning, and data mining toolkit that is open source. It includes a visual programming front-end for exploratory data analysis and interactive data visualisation, as well as the ability to be used as a Python library.

DATA VISUALISATION AND INTERPRETATION

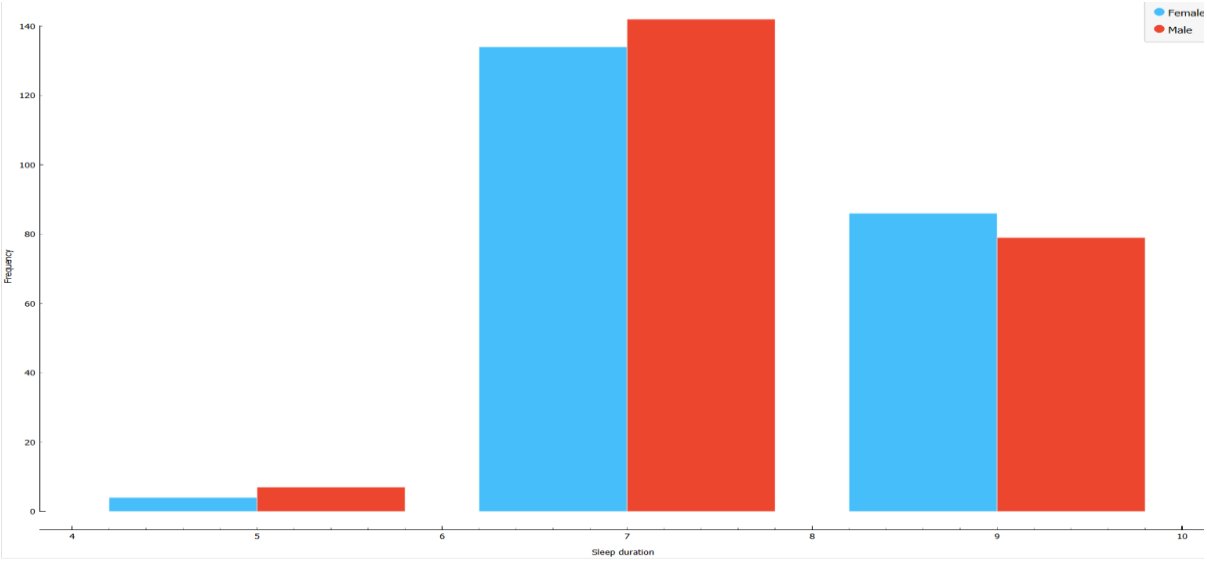
i) Smoking status v/s Sleep efficiency



Inference

Non smokers have better sleep efficiency than smokers.

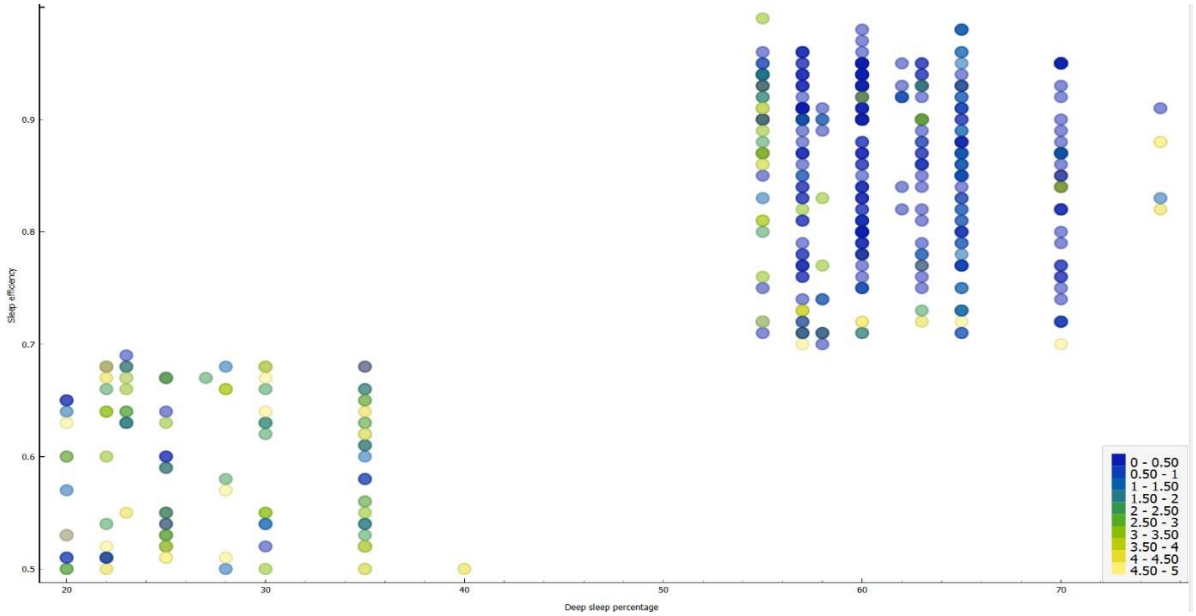
ii) Gender v/s Sleep duration



Inference

- The count of females with more sleep duration is greater than males.

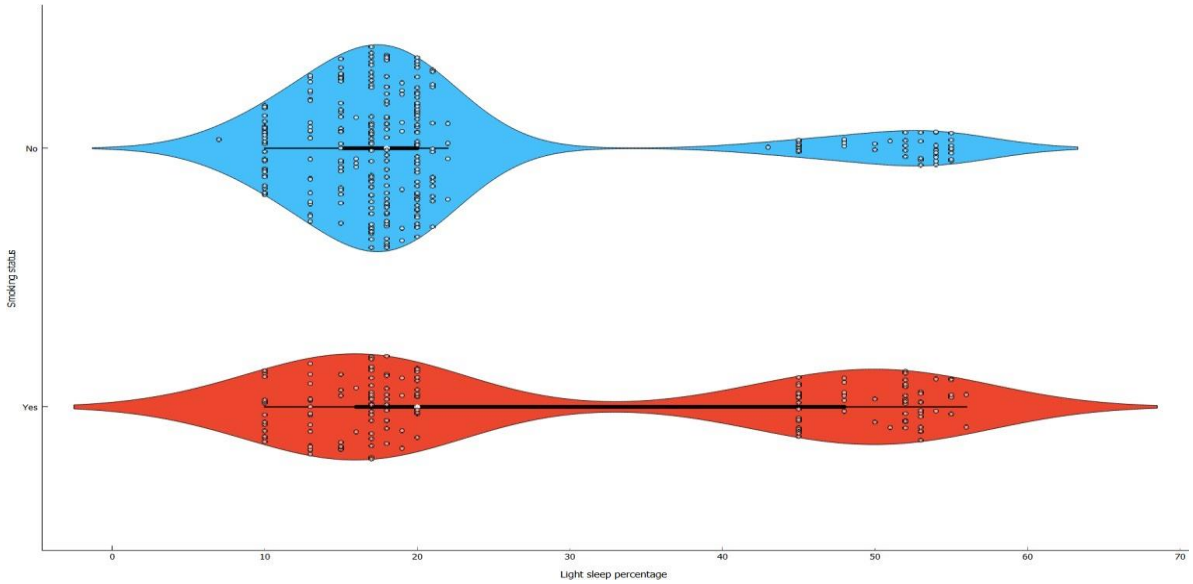
iii) Deep sleep percentage, sleep efficiency v/s alcohol consumption



Inference

- People with low alcohol consumption have more deep sleep percentage and sleepefficiency.

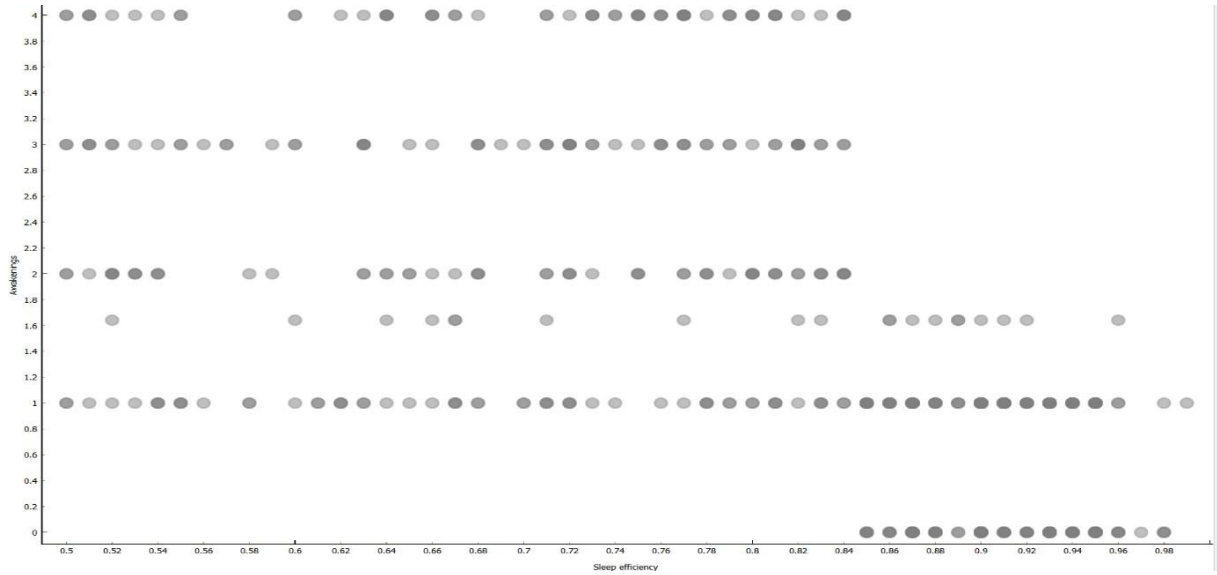
iv) Light sleep v/s smoking status



Inference

Smokers spend more time in light than deep sleep.

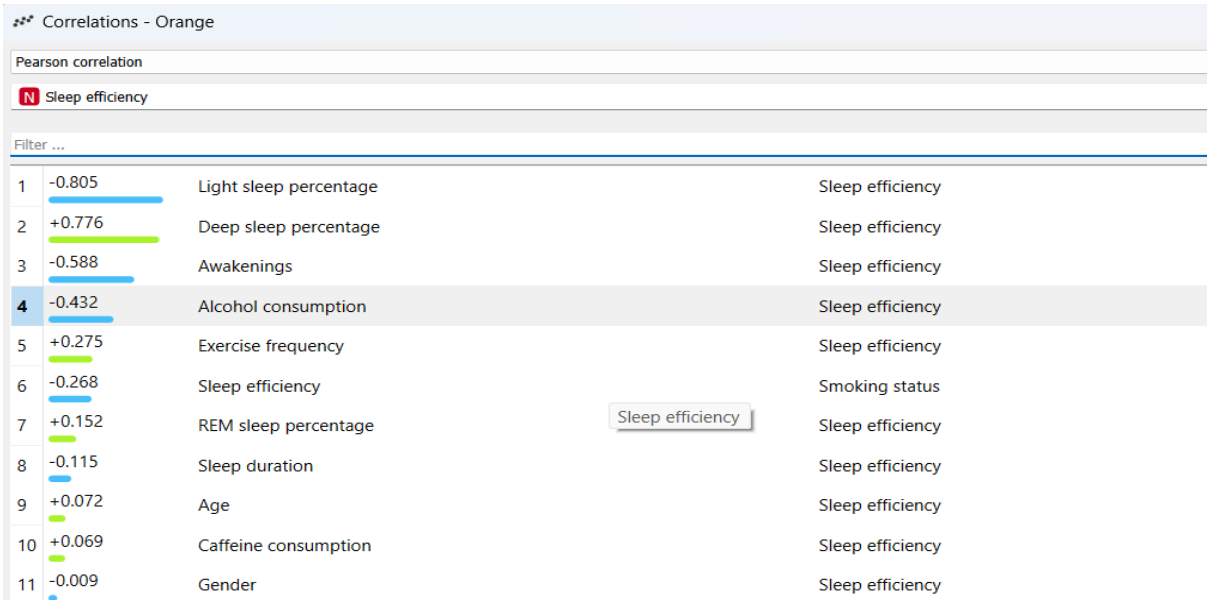
v) Awakenings v/s Sleep efficiency



Inference

- People with low level of awakenings have better sleep efficiency.

Correlation Analysis

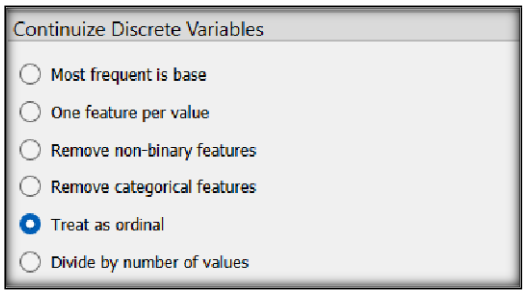


From this we can understand the light sleep percentage and deep sleep percentage have more correlation with sleep efficiency. Light sleep is negatively correlated. When light sleep percentage increases sleep efficiency decreases.

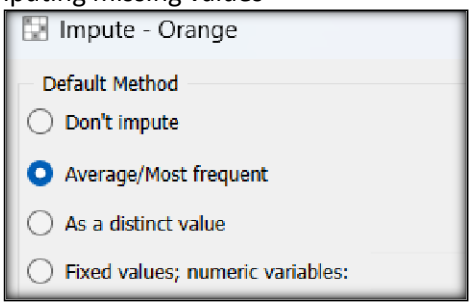
PREPROCESSING

The dataset consists of 1.1 percent missing data. Missing values are present in the attributeslike awakenings, alcohol consumption, and caffeine consumption. These missing values arereplaced using the average values. Categorical variables present in the attributes like genderand smoking status are converted to binary variables. The features like bedtime and wakeuptime which are not necessary for the model building are dropped using select columns widget.

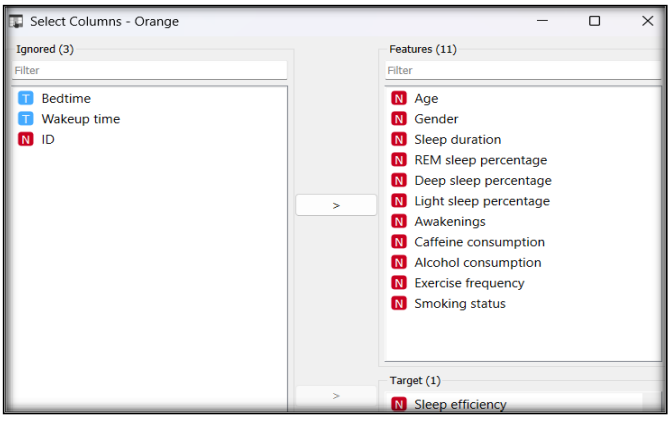
i) Converting to binary variables



ii) Imputing missing values



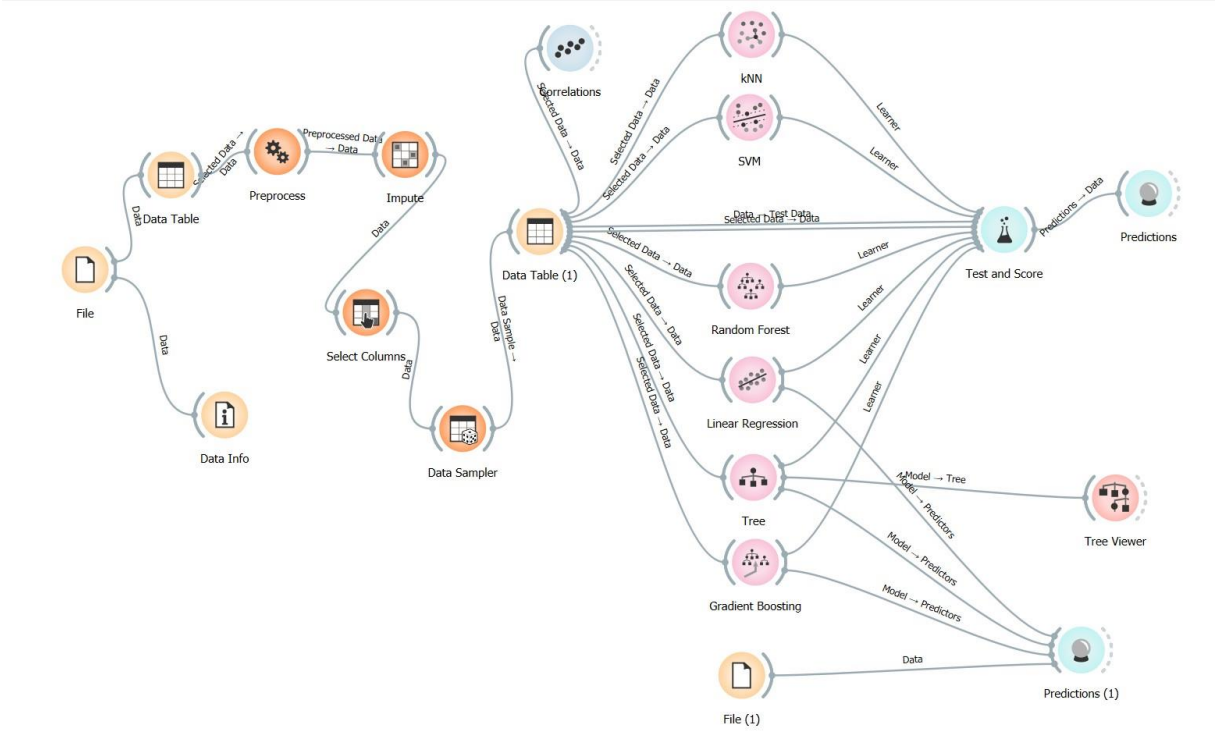
iii) Dropping the columns



MODEL CONSTRUCTION

The project uses various models like K-Nearest Neighbour, Random forest, decision tree, Linear regression, Gradient Boosting, and SVM. Best model is evaluated from

these models.The project uses various models like K-Nearest Neighbour, Randomforest, decision tree, Linear regression, Gradient Boosting, and SVM. Best model is evaluated from these models.



PERFORMANCE

. The performance value obtained for each model is given below:

Model	MSE	RMSE	MAE	R2
kNN	0.003	0.058	0.046	0.817
Tree	0.001	0.022	0.014	0.973
SVM	0.004	0.065	0.056	0.773
Random Forest	0.001	0.024	0.018	0.968
Linear Regression	0.003	0.056	0.045	0.829
Gradient Boosting	0.001	0.030	0.022	0.951

	Sleep efficiency
1	0.88
2	0.66
3	0.89
4	0.51
5	0.76
6	0.90

Original values

	Tree
1	0.88
2	0.52
3	0.88
4	0.52
5	0.74
6	0.95

Predicted values

Predicted values of all models:

Sleep efficiency	Selected	Tree	kNN
0.84	No	0.806667	0.82
0.52	No	0.52	0.552
0.94	No	0.9525	0.93
0.93	No	0.89	0.85
0.78	No	0.7975	0.84

SVM	Random Forest	Gradient Boosting	Linear Regression
0.811126	0.802176	0.793877	0.827319
0.620261	0.530143	0.552512	0.589434
0.839874	0.919875	0.90454	0.846781
0.862029	0.918181	0.889997	0.893956
0.817228	0.784833	0.783479	0.835148

RESULTS AND CONCLUSION

Among the models tree has highest value for co-efficient of determination(R squared), whichis interpreted as the proportion of the variance in the dependent variable that is predictable from the independent variable. It is value between 0 and 1 that measures how well a statistical model can predict a outcome. So, tree is the best model among others.

CONCLUSION

Sleep is a important factor which determines our health. A healthy sleep efficiency is approximately between 85%-95%.From this project it is clear that alcohol consumption ,smoking status have impact on the sleep efficiency. People who consume more amount of alcohol and smokers have less sleep efficiency .They have more light sleep percentage than deep sleep.

REFERENCES

[1]. Keanu Paul B.Sygaco, The Correlation of Sleep And Academic Performance-

2021

[2]. Luiz Fernandez -luque,"Sleep quality prediction from wearable data using

deep
learning,"

[3]https://www.researchgate.net/publication/30
135
0658_Sleep_quality_and_stress_a_literature_re
vi w