Sleep Analysis Using Machine Learning

Kevin Liza Alex
PG Scholar Amal Jyothi College of Engineering,
Kanjirappally, Kerala
kevinlizaalex@mca.ajce.in

Gloriya Mathew
Assistant Professor, Amal Jyothi College of
Engineering, Kanjirappally, Kerala
gloriyamathew@amaljyothi.ac.in

***Abstract*—This paper analyses human sleep and evaluates how various factors affect the sleep a human being. For every single human being sleep is an essential factor. The lack of adequate sleep leads to various health issues. Using Machine learning technique with the help of linear regression algorithm, the model predicts whether the human being is having enough sleep based on the various factors that affects the quality of sleep.**

Keywords—sleep, analysis, enough sleep, machine learning, linear regression.

I. INTRODUCTION

Machine learning is an important branch of artificial Intelligence, that analyses the data input and train itself and thus enhances the purpose it is intended to perform using various learning algorithm available. Machine learning from the data its gets and they identify the pattern within the data and use this information in decision making. It helps the humans to handle the huge volume of data to interpret, evaluate and make decision.

Here linear regression algorithm which is a machine learning algorithm is used for the analysis. Linear regression is a well-known machine learning algorithm. As the name linear regression denotes it represents a linear relationship between and dependent variable (Y) and one or more independent variable (X).

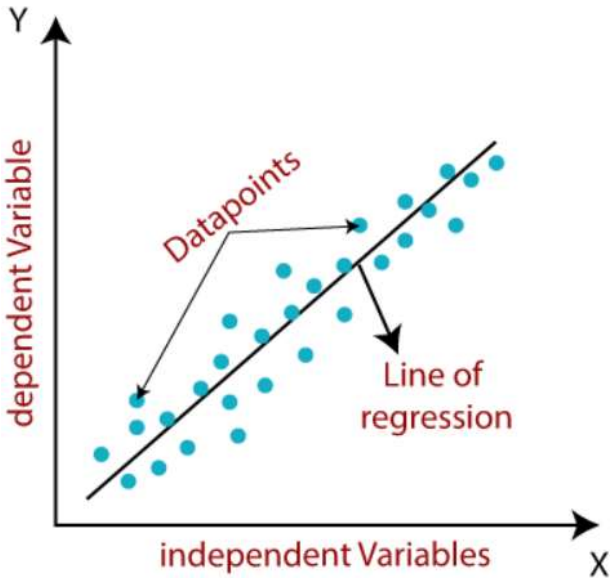

Figure 1: Linear regression graphical representation

That is,

Y= Dependent Variable (Target Variable)

X= Independent Variable (predictor Variable)

Using machine learning and linear regression algorithm the sleep data is analysis is performed here. Analysis is carried out based on factors that affect the sleep such as tiredness, hours of sleep, phone time, phone reach, breakfast and predict whether the sleep is enough. The proposed system analyses the influence of each of these factors and predicts whether the sleep is enough or not.

To train the model, a 10% data from the dataset is used and the rest 90% is used for testing.

A. REQUIRMENTS

- Jupyter Notebook - IDE
- Linear Regression - (Any appropriate algorithm)
- Dataset – Data related to the seminar domain

B. FEATURES

- Human sleep is analysed based on various factors that affect the rate and quality of sleep.

DOI: 10.5281/zenodo.6181505
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam

- The model is implemented based on evaluating only those factors which makes the model more accurate and yet which affects getting enough sleep.
- The model takes a dataset and pre-process it.
- Using linear regression algorithm, the model train and test the data.
- Finally demonstrate the result in a graphical manner.

Figure 2: Jupyter notebook, tool used for implementation

II. LITRATURE REVIEW

[1] "Analysis of Effective Sleep through Logistic Regression" uses frequency, percentage, chi-square test, analyzing factors related to quality of sleep and predicting sleep quality by using logistic regression with Spyder, an IDE for python language. The factors include the age spectrum related to the quality of deep sleep. Analysis of various sleep data of normal and sick people in different age groups found that sleep conditions and personal health were related.

[2] Dilek Kara Yılmaz[1], Fatma Tanrıkulu[1], Yurdanur Dikmen[1] researched about the quality of sleep and conduct study on factors that are affecting the sleep of a normal human being. This is a descriptive research comprised of 223 students as volunteer who are studying at Uludağ University Faculty of Health Sciences.

[3] Nazish Rafique[1], Lubna Ibrahim Al-Asoom[1], Ahmed Abdulrahman Alsunni[1], Farhat Nadeem Saudagar[1], Latifah Almulhim[2], and Gaeda Alkaltham[2] conducted study analyses the association of mobile phone usage with poor sleep quaity. It finds out the mobile-related sleep risk factors (MRSRF) on humans.

III. METHEDOLOGY

By using the Jupyter notebook, an IDE, the model is implemented. Jupyter notebook is an open-source application contains analysis of dataset, training model and thus implements the predictive model. Jupyter notebook supports programming languages like Python, Julia and R.

Figure 3: Jupyter notebook interface

Jupyter notebook can be used to run linear regression, the machine learning algorithm used by the analysis.
Jupyter notebook is used for:

- Preprocessing
- Training
- Testing
- Load saved model
- Visualizations

Linear regression is a predictive model. This machine learning algorithm explains the relationship between a dependent variable (Y) and one or more independent variable (X).

A. DATASET

It is the collection of data to carry the analysis, to train and test the model. Columns in the table represents each factor that influence enough sleep rate. The dataset should be in the format .csv or arff file extension. If not, it should be converted to .csv format.

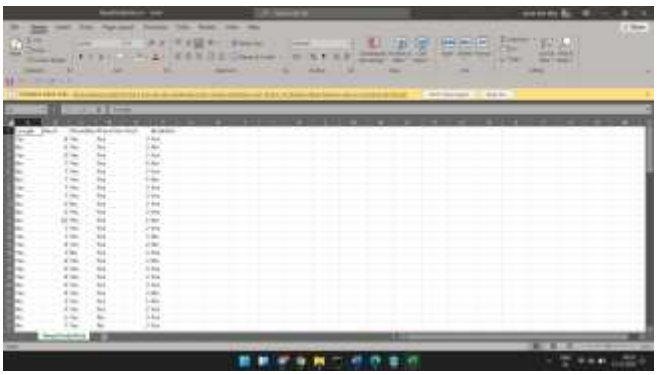

Figure 4: .csv dataset file

IV. IMPLEMENTATION

To implement the proposed analysis model version 3.7+ of python is required the local system. Download and install it in local machine.

Open the following directory, C -> users -> user-> appdata to install Jupyter notebook.

Take command prompt and type in the required command for opening Jupyter Notebook.

Type "jupyter notebook" in CMD to open Jupyter Notebook. The interface will open in the system's default browser.

Figure 5: Jupyter Notebook homepage

DOI: 10.5281/zenodo.6181505
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam

New -> Python 3 to open new file.

Import modules pandas, sklearn, matplotlib, pickle.

Then import the dataset by setting the location of the .csv file.

Figure 6: Code to import modules and loading dataset

Now encode target labels with 0 to n values using preprocessing.

```
le = preprocessing.LabelEncoder()

Enough = le.fit_transform(list(data["Enough"]))
Hours = le.fit_transform(list(data["Hours"]))
PhoneReach = le.fit_transform(list(data["PhoneReach"]))
PhoneTime = le.fit_transform(list(data["PhoneTime"]))
Tired = le.fit_transform(list(data["Tired"]))
Breakfast = le.fit_transform(list(data["Breakfast"]))
```

Figure 7: Encoding target labels

Assign the dependent variable (Y) and nondependent variable (X).

```
predict = "Tired"

X = list(zip(
    # Enough,
    Tired,
    Hours,
    # PhoneReach,
    PhoneTime,
    # Breakfast
))
y = list(Enough)
```

Figure 7: Code to assign values to X and Y variables

Train the model inside a loop to get the best accuracy.

```
# Training model in a loop

best_accuracy = 0
for _ in range(500):
    x_train, x_test, y_train, y_test = sklearn.model_selection.train_test_split(
        X, y, test_size=0.1)
    linear = linear_model.LinearRegression()
    linear.fit(x_train, y_train)
    acc = linear.score(x_test, y_test)  # acc stands for accuracy
    if(acc > best_accuracy):
        best_accuracy = acc
        with open("savedmodel.pickle", "wb") as f:
            pickle.dump(linear, f)

print("Accuracy = ",best_accuracy)
```

Figure 8: Code to train model in loop

Loading the saved model with best accuracy.

```
# Loading saved model and using it

saved_model = open("savedmodel.pickle", "rb")
linear = pickle.load(saved_model)
x_train, x_test, y_train, y_test = sklearn.model_selection.train_test_split(
    X, y, test_size=0.1)
predictions = linear.predict(x_test)
for x in range(len(predictions)):
    print(predictions[x], y_test[x])
```

Figure 9: Code to load the saved model with best accuracy.

Visualization of the output

```
# Visualization

x = list(Hours)
y = list(Enough)

plt.plot(x, y, color='g', linestyle='dashed',
         marker='o', label="Mood Graph")

plt.xticks(rotation=25)
plt.xlabel('Hours')
plt.ylabel('Enough')
plt.title('How phone usage affect sleep', fontsize=20)
plt.grid()
plt.legend()
plt.show()
```

Figure 10: Code for graphical representation of result

Thus, the final output is produced.

V. RESULT

```
Accuracy =  0.7040366368198803
0.27688379647181605 0
0.4604208843063746 1
0.08918819504155567 0
0.5348342968342823 0
0.27688379647181605 1
0.44306575291700295 1
0.27688379647181605 0
0.27688379647181605 1
0.7927847971967483 1
0.6266028407515616 1
0.20247038394390843 0
```

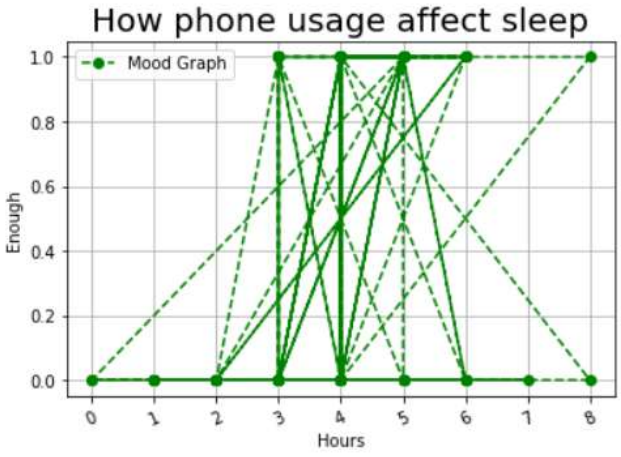

Figure 11: Final output

DOI: 10.5281/zenodo.6181505
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam

The result produced the best accuracy value. It also represents the prediction value and the test data to evaluate the model. A graph representation is produced as an output in the end to find how the factors such as Tired, Hours, PhoneTime affect Enough.

VI. CONCLUSION

Sleep is an important factor for every human being. Various factors define whether the sleep is enough or not. Among the factors evaluated such as Tired, Hours, PhoneTime, PhoneReach, Breakfast, while considering the only Tired, Hours and PhoneTime make the model more accurate. Thus, the model can be used to analyze and predict whether the sleep is enough or not.

VII. REFERENCE

[1] Jakrith Intha, Part Pramokchon, Paween Khoenkaw and Kitisak Osathanunkul. Analysis of Effective Sleep through Logistic Regression.

[2] Dilek Kara Yılmaz, Fatma Tanrıkulu, Yurdanur Dikmen, Research on Sleep Quality and the Factors Affecting the Sleep Quality of the Nursing Students.

[3] Nazish Rafique, Lubna Ibrahim Al-Asoom, Ahmed Abdulrahman Alsunni, Farhat Nadeem Saudagar, Latifah Almulhim, and Gaeda Alkaltham. Effects of Mobile Use on Subjective Sleep Quality.

DOI: 10.5281/zenodo.6181505
ISBN: *978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam*