Sleep Quality Analysis Based on Logistical, SVC and Xgboost

Hanjun Gong*

Department of Computer Science, Tongda College of Nanjing University of Posts and Telecommunications, Yangzhou, China

*Corresponding author: ljones74718@student.napavalley.edu

Abstract. As a matter of fact, a healthy lifestyle requires sleep quality to be one of the main ingredients. It's a common practice to spend roughly a third of the lives sleeping, or at the very least, one should. Regrettably, that goal is not met by many of human beings in recent years. With this in mind, this study will investigate to build three models LogisticRegression Model, XGBClassifier Model, SVC Model using python technique. From the acquired data the model is used to find out some lifestyle effects on sleep. XGBClassifier Model is selected as the best model for the study. The model concluded that Occupation has the greatest effect on insomnia. Sleep apnea is most affected by BMI Category. BMI Category, Stress Level, Blood Pressure, Sleep Duration have more influence on sleep. These results show that the quality of sleep can be improved in terms of BMI Category and Blood Pressure.

Keywords: Sleep health; logistical model; SVC; Xgboost.

1. Introduction

There is a close relationship between sleep health and lifestyle. A healthy lifestyle can promote good sleep, while good sleep can support physical and mental well-being. Over the past few decades, people have gained a deeper understanding of the importance of sleep. Sleep is a process that is crucial for the normal functioning of the body and the brain. It helps the body repair and recover, consolidates learning and memory, regulates emotions, enhances immune function, and maintains overall health.

For most of its brief history, sleep medicine has defined itself in terms of sleep disorders and, more recently, sleep deficiency [1]. The established pattern has been followed by sleep medicine.Other medical disciplines are concentrating on disorders, diseases, and their treatment [2]. The relationships between health and the two components of sleep, quantity and quality, have been studied in some detail in clinical populations [3, 4]. For example, sleep duration has been linked to cardiovascular disease [5-7]. In addition, sleep itself has been shown to be a health risk in clinical populations, perhaps due to the physiological changes that occur during sleep [8, 9]. To assess subjective feelings of health, this study included The Cornell Medical Index (CMI) in the battery of surveys [10]. The CMI has been widely used as an aid in medical-history taking and significant correlations have been obtained between physicians' rating and self-rating of psychological and physical health . The CMI poses yes/no questions regarding physical and psychological health [11].

In order to examine the connections between sleep and indicators of health, well-being, and sleepiness, it is necessary to categorize sleep into two components, one being quantity and the other being the quality. While there is some overlap between these components of sleep, there is a qualitative difference distinguishing them. Sleep is made up of quantifiable elements such as the number of times you wake up during the night, sleep latency and sleep duration, but sleep quality also includes mainly subjective measures such as the depth of sleep, how refreshed you feel when you wake up and your overall satisfaction with your sleep.Subjective sleep quality indexes have been developed by researchers in order to better quantify the quality of sleep [12].

There are also different ways to measure sleep definition， The existing literature provides guidance in devising such tools by identifying thresholds for health risk associated with different dimensions of sleep. A self-report scale called SATED is a potential but unvalidated tool for measuring sleep health. The SATED scale evaluates five key dimensions of sleep that have been consistently linked to health outcomes, and it utilizes specific quantitative criteria for four of them.

The dimensions include satisfaction with sleep, alertness during waking hours, sleep timing, sleep efficiency, and sleep duration. The motivations for selecting the "sleep health and lifestyle" topic are to gain a deeper understanding of the impact of sleep on health, analyze how lifestyle and environmental factors influence sleep, and provide important information about health and quality of life to society. This can help individuals better manage their health, enhance their quality of life, and address relevant issues. This study first crawled some data from the website to import the data into the code to understand the statistical information. This study calculated the dimension of the data and calculated the Data Shape as well as the data type. Since different data types may affect the processing method, this study organized the data types for each item of data. Next, this study detected empty and type of values and made sense of the statistics of the data. After completing this Exploratory Data Analysis. Understanding a data set" can refer to many things including relationships between data sets, dealing with missing values or human error and identifying outliers. After that, this study performed EDA (EDA is a step in the data analysis process where techniques are used to better understand the dataset used) . In this process, this research visualized the age, sleep duration, daily steps, quality of sleep, stress level, heart rate, physical activity level of each of these people as a scatter plot. Then, percentage of persons have sleep disorder or not was calculated.Then data processing was done. After that, several models were selected and built and their accuracy was calculated and compared by training and testing. After this, the models were evaluated and predictions were made. Finally, the best model was applied to the data and visualized to draw conclusions.

2. Data and Method

This study used SQL technique to crawl some data on the data website and then conducted a questionnaire survey on the internet for 3 days. Finally, this paper integrated these data together to become the experimental data. Logistic Regression Model, XGB Classifier Model, and SVC Model are applied in this project. Among them, Logistic regression model Logistic regression is a general linear regression model., which is a linear regression model normalized by the Sigmoid function, and it is commonly used to solve the binary classification problem with strong interpretability. The text already adheres to the principles or lacks context, so the improved version of the text is. The basic principle is that its function expression is a log odds function, and the linear regression equation is transformed through the Sigmoid function, which maps any real number to between (0, 1). XGB Classifier Model is a gradient-based boosting decision tree, which simulates the principle of the training process is to initialize the model, set the predicted values, and then iteratively train multiple decision tree models, each iteration is Each iteration calculates the residuals from the gradient and second order gradient information. At each iteration step the best decision tree model is selected to fit the residuals by optimizing the loss function. SVC Model is a classification model based on Support Vector Machines it classifies a new unlabeled sample by calculating its distance from the optimal hyperplane given a new unlabeled sample. SVC model parameters are set as follows: regularization parameter float is 1.0, shrinking: bool, default=True, probability: bool, default=False, tol: float, default= 1e-3, Specify the size of the kernel cache (in MB) as float type 200 size. The logistic regression model parameters are set as follows: Tolerance for stopping criteria=1e-4, Inverse of regularization strength=1.0, Specifies if a constant=true, Dual or primal formulation=false.XGB Classifier Model parameters are set as follows: booster=gbtree, learning_rate=0.3, gamma=0. The other parameters in the three models are the default parameters of sklearn. The evaluation metrics of the three models are their accuracy as well as the application of the Confusion matrix, by calculating the accuracy rate to know in which sample this model performs well or poorly.

3. Results and Discussion

In order for the collected data to be applied to the code, this study imported the data into the program. The processing of the data to calculate the Percentage of persons have sleep disorder or not.

It is very clear to conclude that the percentage of normal people who have more disorders is greater. Through the data visualization technique, this study has studied the distribution of persons having sleep disorder or not. Fig. 1 shows the distribution, which gives a clear picture of the current sleep status of the respondents. Besides, one can be seen those normal men more than women, men who suffer from Insomnia more than women, and women have sleep apnea more than men. Fig. 2 shows the relationship between sex and sleep disorders. Immediately after that this study counted the sleep of different occupations. It can be seen that fewer people in the profession of doctor have sleep apnea. In addition, people who suffer from Insomnia, Salesperson more than others.sleep apnea nurses are more in number than other professions. The effect of work on sleep visualized through the data in Fig. 3. Finally, this study has visualized the last 3 items in the data blood pressure BMI pressure and Heart Rate in three dimensions which is Fig. 4. However, no significant features were found.

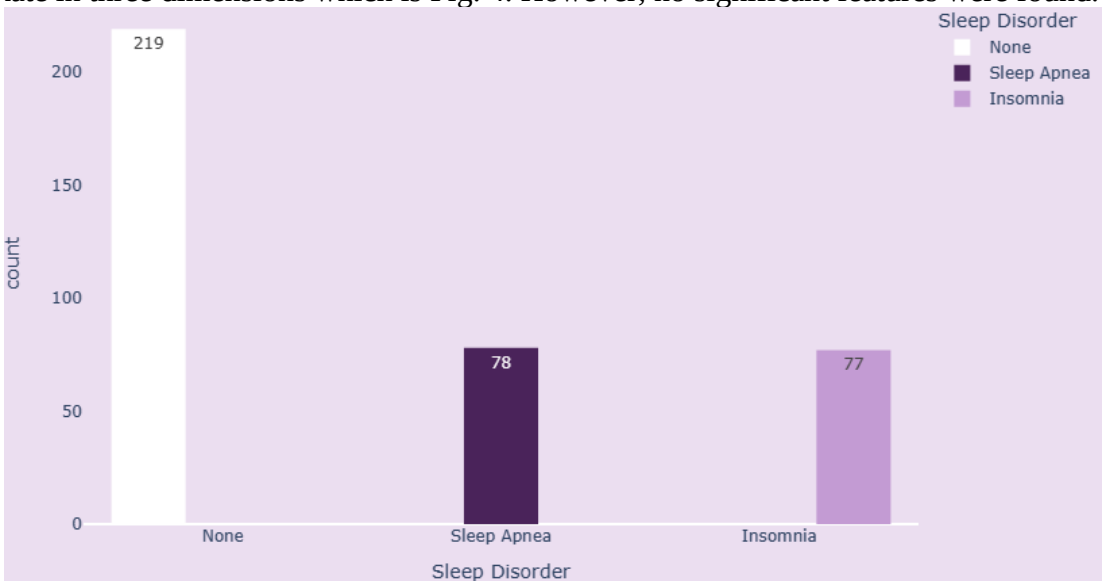


Fig. 1 Distribution of persons have sleep disorder.

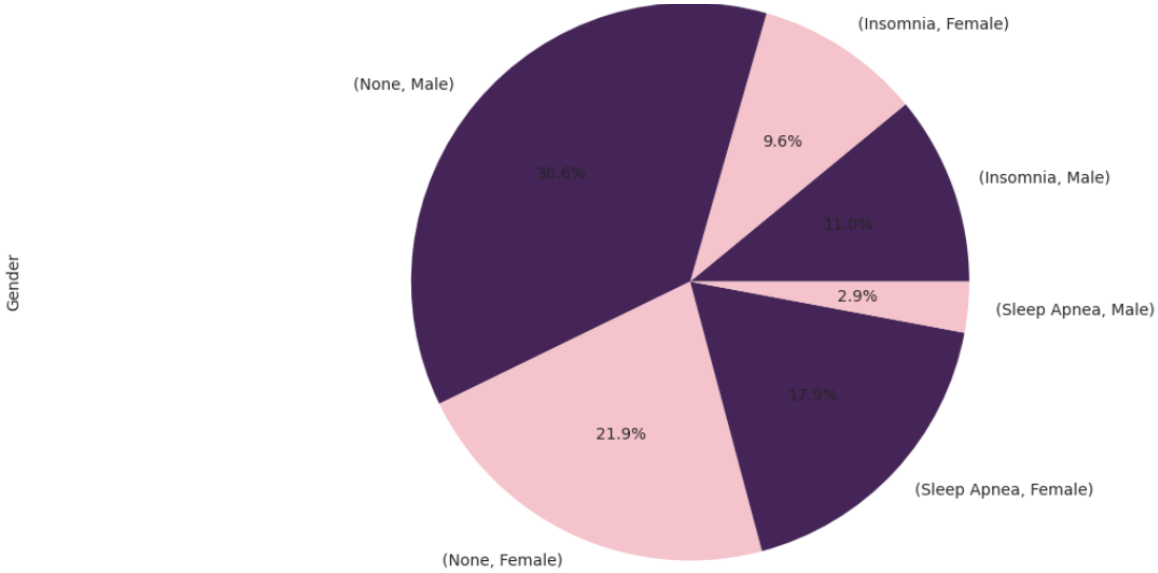


Fig. 2 The relationship between sex and sleep disorder.

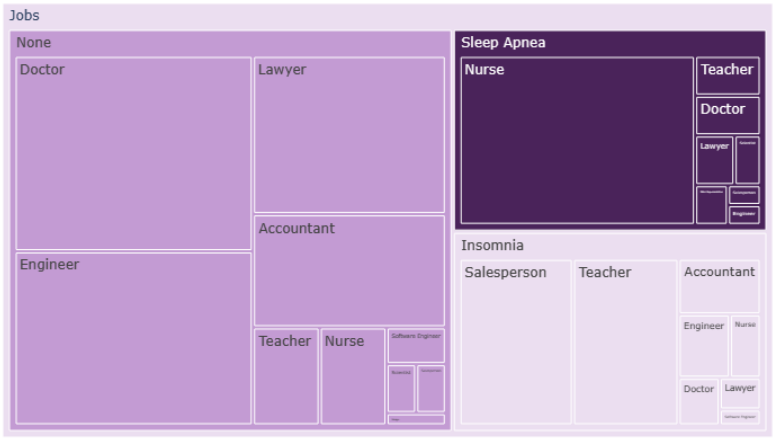


Fig. 3 The effect of job on sleep.

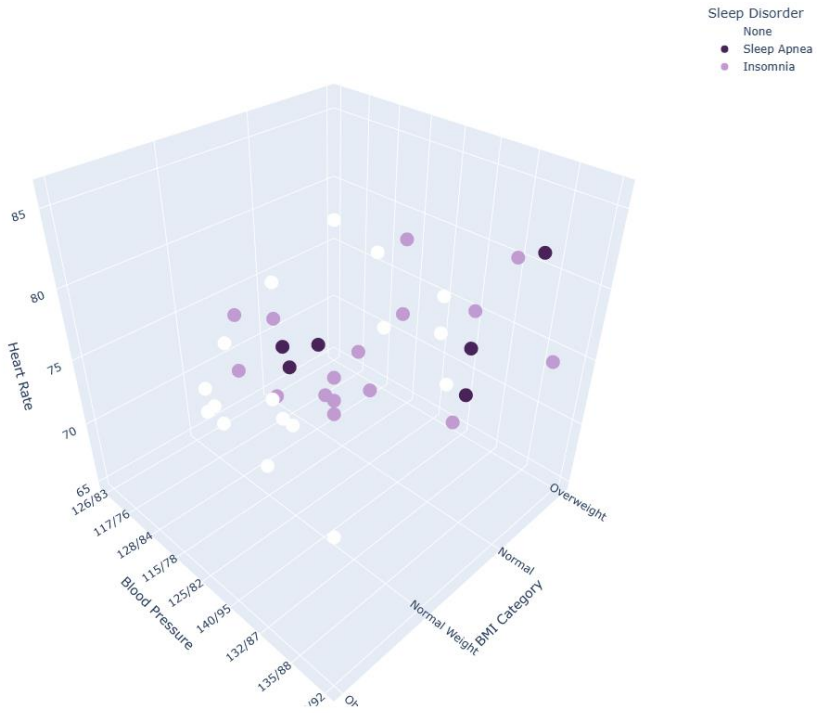


Fig. 4 The relationship between BMI and effects.

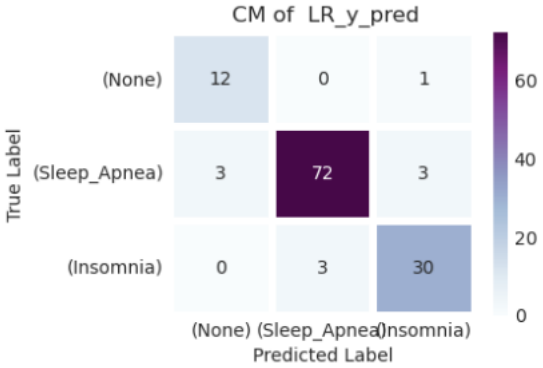


Fig. 5 LR's confusion matrix.

Fig. 5 shows the confusion matrix plot for Logistic Regression Model. Fig. 6 shows the confusion matrix of XGB Classifier model, SVC Model. Fig. 7 shows the results of the XGB Model applied to the data. The Logistic Regression Model training score is 90.8 and the test score is 91.94 which is generally similar. The XGB Classifier Model training score is 93.2 and the test score is 91.13 although the overall score is high there is a gap between the test and the training. The SVC Model training

score is 88.8 and the test score is SVC Model training score is 88.8 and test score is 87.9. It is the closest model between training and test among the three models. In terms of accuracy Logistic Regression Mode, SVC Mode outperforms the other model in terms of none as well. SVC Mode performs better in predicting insomnia. While XGB Classifier with Logistic Regression Model is better in predicting sleep apnea. So for predicting insomnia and sleep apnea one can use SVC Model but the accuracy of the model is a bit low. The XGB Classifier model can be used for none prediction and the overall performance of this model is good.

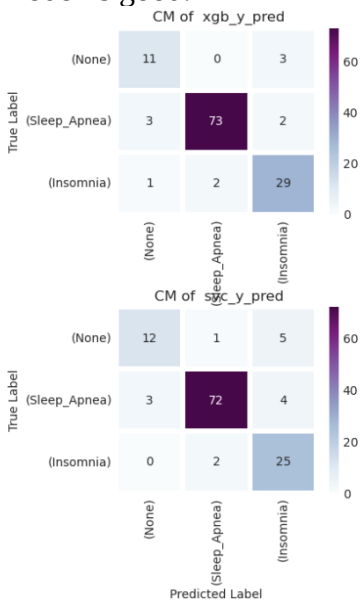


Fig. 6 XGB and SVC confusion matrix.

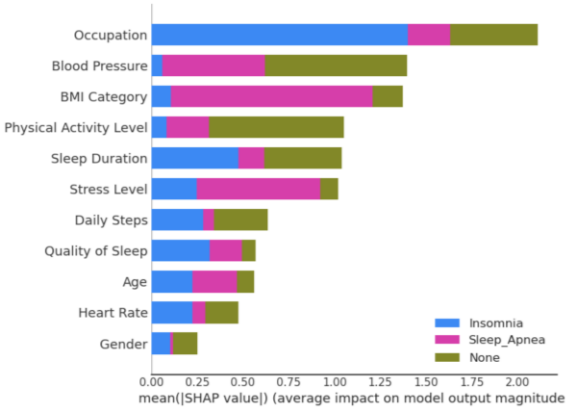


Fig. 7 XGB performance

4. Limitations and Prospects

The results shown in these three models are based on these data, but as the amount of information in the data increases these models may not be applicable to the data. In addition, most of the model parameters used in the model building are default parameters in sklearn, which can lead to bias in the real results. This study filtered out the better models and used relatively appropriate models to account for the effects of visualizing sleep and statistical data. However, the predictive effect of other models is unknown, so this is a limitation of this experiment. In addition, the data obtained in this experiment was not obtained using a rigorous survey method, which directly or indirectly affects the results of the experiment. The range of data in this experiment is not large, and the model can only find some general connections and patterns in the prediction process. Sleep is a hot issue in today's era, it is believed that the results of this experiment can play a role in guiding people's healthy sleep and have a new understanding of sleep. It is hoped that after seeing my research, people can realize some of

the aspects that affect sleep, and in the subsequent process of life can do a healthy work and rest, often insomnia can be based on the results of the experiment to adjust the work and rest to open up a new lifestyle and finally do not insomnia.

5. Conclusion

To sum up, this study is based on the theme of Sleep Health and Lifestyle. The data collected from the internet were analyzed, visualized, and the optimal model was selected and evaluated using three models: Logistic Regression Model, XGB Classifier Model, and SVC Model using python technology. It is concluded that XBG Classifier Model performs better in terms of accuracy, and confusion_matrix evaluation. It was concluded that Occupation had the greatest effect on Insomnia, and Sleep apnea was most affected by BMI Category.Age, Sleep Quality, Heart Rate, Physical Activity Level, Daily Steps, Sleep Duration, and Gender had little effect on Sleep apnea and Insomnia results. and Insomnia results. Blood pressure and stress indices also had a significant effect. However, the parameters used in this study are basically the default parameters in the sklearn library, which have an effect on the results. It is believed that this study will give people a new understanding of sleep. This experiment is of great significance to improve the quality of sleep and healthy life.

References

[1] National Center on Sleep Disorders Research. National institutes of health sleep disorders research plan. Bethesda, MD: National Institutes of Health, 2011.

[2] Czeisler C A. Impact of sleepiness and sleep deficiency on public health—utility of biomarkers. Journal of Clinical Sleep Medicine, 2011, 7(5 Suppl): S6-S8.

[3] Strollo P J, Badr M S, Coppola M P, Fleishman S A, Jacobowitz O, Kushida C A. The future of sleep medicine. Sleep, 2011, 34: 1613-9.

[4] Daniel J, Buysse M D. Sleep Medicine Institute and Department of Psychiatry, School of Medicine, University of Pittsburgh, Pittsburgh. PASLEEP 2014, 37(1): 9-17

[5] Ridner S L, Newton K S, Staten R R, et al. Predictors of well-being among college students. Journal of American college health, 2016, 64(2): 116-124.

[6] Fossey E, Fitzpatrick M F, Douglas N J, et al. Interaction and classification of sleep and health risk. Sleep and Health Risk. Berlin, Heidelberg: Springer Berlin Heidelberg, 1991: 174-179.

[7] Yang B, Qin Q, Han L, et al. Spa therapy (balneotherapy) relieves mental stress, sleep disorder, and general health problems in sub-healthy people. International journal of biometeorology, 2018, 62: 261-272.

[8] Brodman K, Erdmann A J Jr, Wolff H G, Miskovitz P F. Cornell Medical Index health questionnaire manual. New York: Cornell University Medical College, 1986

[9] Bowling A. Measuring health: a review of quality of life measurement scales. Philadelphia: Open University Press, 1991

[10] Lane N, Mohammod M, Lin M, et al. Bewell: A smartphone application to monitor, model and promote wellbeing. 5th international ICST conference on pervasive computing technologies for healthcare. 2012.

[11] Pilcher J J, Ginter D R, Sadowsky B. Sleep quality versus sleep quantity: relationships between sleep and measures of health, well-being and sleepiness in college students. Journal of psychosomatic research, 1997, 42(6): 583-596.

[12] Okano K, Kaczmarzyk J R, Dave N, et al. Sleep quality, duration, and consistency are associated with better academic performance in college students. NPJ science of learning, 2019, 4(1): 16.