JOURNAL OF CRITICAL REVIEWS

ISSN- 2394-5125 VOL 7, ISSUE 08, 2020

STUDY THE RELATIONSHIP BETWEEN THE SLEEP QUALITY AND HAVING A BED-PARTNER USING DATA MINING TECHNIQUES

Soraya Gharavi*
Esfarayen Higher Education Complex, North Khorasan, Esfarayen
Instructor, Faculty of Computer and Electrical Engineering, Department of Computer
Bojnourd, Nader St., Daqiqi Alley, No. 32, Postal code: 9414965395, ORCID: 9065-7451-0001-0000
Iran, gharravi@esfarayen.ac.ir

ABSTRACT

Human body requires adequate rest and sleep. Sleep's high quality leads to maintain person's physical, spiritual, mental and emotional abilities as well as perceptual abilities such as memory, speaking, concentration and creative thinking at the optimum level. Improving sleep quality is significant for the improvement of life quality. The presence of clinical and laboratory features and symptoms in sleep disorders makes it difficult for physicians to diagnose the origin of the disorder. Data analysis makes it possible to analyze the clinical data of patients for medical decision making. Having a bed partner, while sleeping, may affect sleep quality. The purpose of the paper is to provide a model for the study of the relationship between sleep quality and having a bed partner using data analysis techniques.
Methodology: In this study, 500 people were examined. The data set was collected according to the standard Petersburg quality questionnaire in the villages around Golestan state. Data analysis techniques are used to provide the model for studying the relationship between sleep quality and having bed partner.
Findings: The prediction accuracy of the methods of Naïve Bayes, support vector machine and the nearest neighbor to study the relationship between sleep quality and having a bed partner was 60.0, 71.0 and 73.0, respectively. The nearest neighbor have had the highest accuracy with k = 11.
Conclusion: There is a significant relationship between sleep quality and having a bed partner. The findings indicate that having a desired bed partner increases sleep quality.

Keywords: data analysis, sleep quality, bed partner, support vector machine, the nearest neighbor algorithm, simple Bayes

1. INTRODUCTION

Sleep is one of the most significant factors in the circadian cycle to restructure the physical and emotional forces. Sleep disorders are the psychological problems which affect the life quality [1]. The most common of several types of sleep disorders is insomnia. Some diseases rooted in the quality of people's sleep [2]. Cardiovascular and neurological disease such as strokes, endocrine diseases such as diabetes, neurological diseases such as depression, occupational diseases and problems such as decreased performance and increased incidence of errors and accidents can be caused by or exacerbated by sleep disorders [3].

Research indicates that if the sleep quality is inadequate, it will reduce the maximum level ability of individuals' activity, their talent, decreased immune system, decreased hypothalamus function, decreased glucose tolerance and increases blood pressure. Aggressive behaviors, irritability and decreased social communication are most obvious and higher in those whose sleep quality is inadequate than those who sleep adequately [4].

It is possible to investigate the relationship between sleep quality and types of diseases using data mining techniques. In medicine, data mining is referred to the process of extracting valid, previously unknown, understandable and reliable information from a medical database to be used in order to predict, diagnose and help treat the disease. One of the applications of data mining in medicine is to detect useful patterns between sleep quality and patient's clinical and laboratory symptoms. Useful model means a model in the data which shows the relationship between a subset of patient data and diagnosis of the disease. Study the relationship between having a bed partner and sleep quality through data mining is the aim of this research [5].

Not having a bed partner may affect those who need more time to fall asleep than others or often wakes up in the middle of the night and it takes a long time to fall asleep again. The inadequate sleep quality will be observed very much in groups such as airline passengers, businessmen, pilots, flight attendants, and others who do not have a regular bed partner as well as drivers, nurses, security guards and those who mostly work in conditions without a bed partner. Therefore, study the relationship between sleep quality and having a bed partner is very important.

ISSN- 2394-5125 VOL 7, ISSUE 08, 2020

Based on the American Psychiatric Association, impaired sleep onset and early awakening, excessive drowsiness, disorders of sleep and wakefulness program as well as sleep disorders called "Parasomnia" are reported as sleep-related disorders. Drowsiness during the day has led to mistakes and sometimes human accidents and catastrophes in industrial societies and many cases have been reported in accidents and related deaths [6].

Pagal in study the pattern of sleep and wakefulness of a number of students, has found that 69.7% of students with low GPA had trouble falling asleep and 72.7% of those with poor sleep quality faced with difficulties in concentrating and paying attention during the day [8].

Mousavi et al. have done a cross-sectional study and studied the prevalence of sleep disorders and its related factors in 407 people. Accordingly, sleep disorders are more common in among interns and stagers [9]. In 2012, Narges Mirzaei et al. have done a cross-sectional study and selected 88 people using cluster sampling method and examined the prevalence of various sleep disorders in them. The PSQI questionnaire was used in order to examine the quality of students' sleep; the higher the scores are the sleep quality is poor and the lower the scores are the sleep quality is good; the scores range is 0 to 21; a score above 5 is regarded as sleep deprivation. 63.6% of the students were female. The correlation coefficient between GPA and inappropriate sleep quality score was negative and equaled to r = -0.139 [10].

A descriptive study was done by Marzieh Nojoumi et al., in which a questionnaire taken from Kaplan Psychiatry and Daily Habits Survey Questionnaire (S-DHQ) were used in order to examine the sleep pattern and the incidence of narcolepsy among 400 students. No narcolepsy was reported [11].

A descriptive cross-sectional study was conducted by Rasool Islami et al. to examine the prevalence of sleep disorders and address some of its-related causes and consequences. The data were collected using self-report through questionnaires whose validity and validity were confirmed. Data were analyzed using SPSS software and descriptive statistics and chi-square test. The results showed that having poor physical conditions was the most common cause of sleep disorders [12].

Reza Qanei'e et al. have evaluated sleep quality in a descriptive-analytical study, among 160 people. Statistical studies suggested that there is a significant relationship between sleep quality and total time of night's sleep [13].

Ali Aghajanloo et al. have used the two-part tools such as personal information questionnaire and Pittsburgh Sleep Quality Index (PSQI) to examine the sleep quality. Data were analyzed using SPSS software, descriptive statistics and frequency table, chi-square test and Pearson correlation coefficient. The results showed that most of the samples had poor sleep quality [14].

In this study, we have introduced a model to examine the relationship between having a bed partner and sleep quality by completing and collecting data sets using the Pittsburgh Questionnaire in the villages around the Golestan state.

2 Materials and methods

2.1 Collecting data sets

Sleep quality data set in the villages around the Golestan state has been collected and completed according to the Pittsburgh Standard Questionnaire (Table 1). The data relate to 2014-2015. This dataset contains 500 samples, 75 of which do not have complete information and so they were excluded from the study. The class names and the number of cases in each class are shown in Table 2.

2.2 Support Vector Machines

The basis of support vector machines is to minimize the structural error according to theories of statistical learning. Suitable generalizability using a limited number of educational data is one of the most important things in learning. In most diseases, adequate educational data don't exist for the learning algorithm or providing it is a time consuming and difficult task. The learning process for disease prediction is possible with a limited number of educational data through support vector machines [15].

In order to do the classification in the SVM method, first the data from the initial space Rn is transferred to Rm space using $\emptyset$ nonlinear conversion, which m>n. The following equation is used to find the optimal decision boundary:

$$\max_{a_i,\dots,a_N}\left[\frac{-1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}a_i a_j y_i y_j\left(\emptyset(x_i).\emptyset(x_j)\right)+\sum_{i=1}^{N}a_i\right]$$
$$0\le a_i\le C\ i=1,\dots,N$$
$$\sum_{i=1}^{N}a_i y_i=0, \quad (1)$$

Where, X, a_i and C are a point on the decision boundary, Lagrange coefficient and a constant number, respectively. In Equation (1), a Kernel Function is used instead of $\emptyset$ as follows:

$$k(x_i,x_j)=\emptyset(x_i)\emptyset(x_j), \quad (2)$$

By determining the value of $k(x_i,x_j)$ the optimization problem will be solved. The Sigmoid Function can be considered as the Kernel Function as follows:

ISSN- 2394-5125 VOL 7, ISSUE 08, 2020

$$k(x_i, x_j) = e^{(-\gamma\|x-x_i\|^2)}, \quad (4)$$

Table 1: Pittsburgh Sleep Quality Index (PSQI)

1	What time do you usually sleep at night?				
2	How many minutes does it take for you to fall asleep when you go to bed?				
3	What time do you usually wake up in the morning?				
4	How many hours do you actually sleep at night?				
5	How many times in the last month have you had trouble sleeping because of the followings?	None	Once a week	Twice a week	Three or more times a week
	After 30 minutes you could not sleep.				
	You would wake up in the middle of the night or early in the morning.				
	You had to get up to go to the bathroom.				
	You could not breathe easily.				
	You were coughing or snoring loudly.				
	You felt very cold.				
	You felt very hot.				
	You had a nightmare.				
	You had pain.				
	If there are reasons except the above mentioned, please name them and say how many timed you have had problems due to them. Your reasons:				
6	How many times have you taken a sedatives or sleeping pills to fall asleep over the past month?				
7	How many times during the last month have you had trouble staying awake to drive or eat or work?				
8	How many times in the last month have you had trouble maintaining your willingness to do things?				
9	How was your overall sleep quality over the past month?	Very good	Fairly good	Fairly bad	Very bad

Table 2: Names and number of instances of classes

Name of class	Number of samples
Has a suitable bed partner	211
Has no suitable bed partner	251

2.3 Support Vector Machines

In the nearest neighbor method, a group contains K records from a set of educational records will be selected which are the nearest records to the test record and the decision about the said experimental record class will be made based on the superiority of the class or the label associated with them. Simply put, this method selects a class which has the highest number of records assigned to that class in the selected neighborhood. Therefore, the class that is observed more than any other classes among the K nearest neighbors will be considered as the new record class [16].

In order to classify a record with an unspecified class, the followings will be done:

- The distance of the new record from all educational records will be calculated.
- Between K neighbor records will be voted and the class with the most views among these K records will be considered as the new record class.

Selecting the value of K in this classification method is very important and key. If the value of K is selected too small, the algorithm will be sensitive to the noise, if the If the value of K is selected too large, records from other classes may be placed among the nearest neighbors, which will lead to a classifier error in classifying the input record.

2.4 Naive Bayesian Classification

We will use Naive Bayesian Classification for problems in which each x is selected from a set of values of features and the objective function f (x) from a set such as V. The Bayesian method to classify the new

ISSN- 2394-5125 VOL 7, ISSUE 08, 2020

sample is to identify the most likely class or target value $< a_1, a_2, \dots, a_n >$ of by having the feature values that describe the new sample.

$$v_{MAP} = \arg\max_{v_j \in V} P(v_j | a_1, a_2, \dots, a_n), \quad (5)$$

Using the Bayesian theorem, the above expression can be rewritten as follows:

$$v_{MAP} = \arg\max_{v_j \in V} \frac{P(v_j | a_1, a_2, \dots, a_n) P(v_j)}{P(a_1, a_2, \dots, a_n)}$$
$$= \arg\max_{v_j \in V} P(a_1, a_2, \dots, a_n, | v_j) P(v_j), \quad (6)$$

Now, using the educational data the two terms of the above equation will be estimated. It is easy to calculate from the educational data as such "How much is the rate of v_j repetition in the data?" Each sample must be viewed several times to obtain a good estimate.

The Naive Bayesian Classification method assumption is based on this simplification that the values of features by having the values of the objective function of each other will be conditionally independent. Equation (7) shows the Naive Bayesian Classification method:

$$v_{NB} = \arg\max_{v_j \in V} P(v_j) \prod_i P(a_i | v_j), \quad (7)$$

Where, v_{NB} is the output of a Naive Bayesian Classification of the objective function. The number of the terms of $P(a_i | v_j)$ which must be calculated by this method is equal to the number of features multiplied by the number of output classes for the objective function, which is much less than the number of the terms of $P(a_1, a_2, \dots, a_n | v_j)$.

Consequently, the Naive Bayesian Learning tries to estimate different values of $P(v_j)$ and $P(a_i | v_j)$ using their repetition rate in the educational data [17].

3 Findings

Appropriate data format as data mining input will influence on the results and output. If the values of the dataset features are in a different range, the probability of error in the findings increases. Inserting the data of a statistical community in the same range is called "normalization" [17]. In the proposed model, the normalization method is Max / Min method and in the range of [1-0] [19, 18].

We have studied the relationship between sleep quality and having a bed partner using SVM, Naive Bayesian and the nearest neighbor methods. In order to examine the relationship between sleep quality and having a bed partner, we have first classified the data according to the classes listed in Table 1.

The purpose of data classification is to organize and assign data to separate classes. In this process, the initial model is created on the basis of the educational datasets, and then the model is used to classify new data so by applying the obtained model the belonging of new data to a certain class can be predicted.

Classification means examining the features of a new object and its assignment to one of the predefined sets. The accuracy rate in predicting determines the relationship between sleep quality and having a partner.

The Confusion matrix can be used to calculate the relationship between real classes and predicted ones. We have listed the required parameters of the Confusion matrix in Figure 1.

The following criteria such as Accuracy, Sensitivity, Specificity, Precision and F-Measure used according to Figure 5 and based on the following equations to compare the proposed model with other methods [20]:

$$\text{Accuracy} = (TP + TN)/All, \quad (8)$$
$$\text{Error Rate} = 1 - \text{Accuracy}, \quad (9)$$
$$\text{Sensitivity} = TP/(TP + FN), \quad (10)$$
$$\text{Specificity} = TN/(FP + TN), \quad (11)$$
$$\text{Precision} = TP/(TP + FP), \quad (12)$$
$$\text{Recall} = TP/(TP + FN), \quad (13)$$
$$\text{F Measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (14)$$

The diagram of the results of detecting different methods with the Accuracy criterion is shown in Figure 2. As it can be seen, the proposed model is more accurate than other methods. Also, the results of the e comparison of disease prognosis using Sensitivity, Specificity, Precision, and F-Measure criteria is shown in Table 3, Table 4, Table 5 and Table 6 respectively. Based on the comparison results, the performance of the proposed model is greater than others.

TP: The number of records that are correctly identified as positive.
TN: The number of records that are correctly identified as negative.
FP: The number of records that are not correctly identified as positive.
FN: The number of records that are not correctly identified as negative.

ISSN- 2394-5125 VOL 7, ISSUE 08, 2020

Figure 1: The required parameters for the relationship between the real classes and the predicted classes

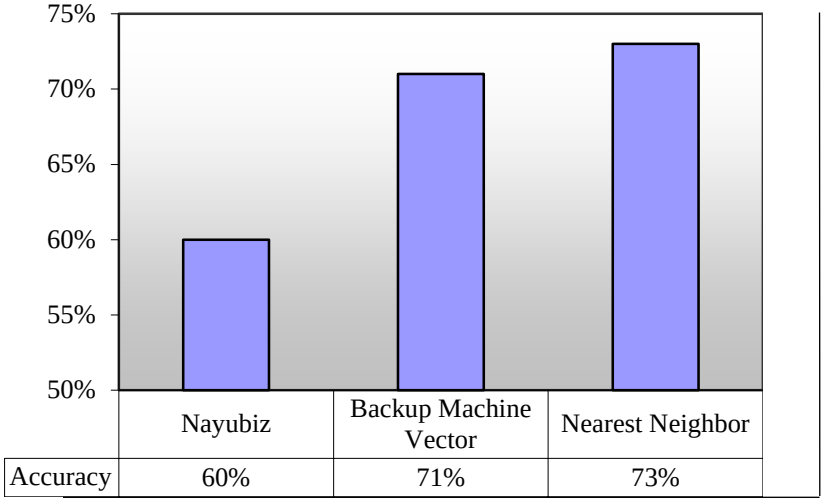

Figure 2: Graph the results with accuracy criteria

Table 3: Compare results with different criteria

	Nayubiz	Nearest Neighbor	Backup Machine Vector
Has a bed partner	0.758	0.979	0.968
No bed partner	0.191	0.064	0.045

Table 4: Comparison of results with specificity criteria

	Nayubiz	Nearest Neighbor	Backup Machine Vector
Has a bed partner	0.191	0.979	0.045
No bed partner	0.758	0.064	0.968

Table 5: Comparison of results with precision criteria

	Nayubiz	Nearest Neighbor	Backup Machine Vector
Has a bed partner	0.708	0.730	0.724
No bed partner	0.233	0.538	0.357

Table 6: Comparison of results with F.Measure criteria

	Nayubiz	Nearest Neighbor	Backup Machine Vector
Has a bed partner	0.732	0.837	0.829
No bed partner	0.210	0.114	0.081

4 Discussion

The accuracy or rate of classification is the most important criteria to determine the performance of a classifier algorithm. This criterion is the most famous and general criterion for calculating the performance of classification algorithms that shows "what percentage of the total set of test records is properly classified by the designed classifier?" The accuracy of the classification is obtained using Equation (8), which states that the two values of TP and TN are the most important values should be maximized in a two-class problem.

The Error Rate is exactly the opposite of classification accuracy criteria which is obtained using Equation (9). When we have the best performance, its minimum value is zero and similarly, when we have the least performance, its maximum value is one. The prediction accuracy of Naive Bayesian, SVM and the Nearest Neighbor is equal to 0.60, 0.71 and 0.73 to examine the relationship between sleep quality and having a bed partner, respectively. The nearest neighbor with the value k = 11 had the highest accuracy.

The accuracy rate obtained through the data mining algorithms in predicting the sleep quality using the bed partner feature shows a significant relationship between having a suitable bed partner and high sleep quality.

ISSN- 2394-5125 VOL 7, ISSUE 08, 2020

Overall sleep quality, adequate sleep and the time to fall asleep features had the greatest impact on predicting. Also, the waking up times in the morning and to go to the bathroom had the least impacts on predicting. Having pain, feeling cold, nightmares and bad sleep, coughing or snoring loudly, feeling warm, shortness of breath, and medicine consumption features had all impacts on sleep quality.

Mousavi et al. used multistage-random sampling method and the data were collected using a sleep and daily habits questionnaires. There is a statistical significant relationship between the person's sleep quality and the time to go to bed, the problems at the sleep onset, waking up due to the surrounding noise, consecutive night waking ups, nap duration in a day, unusual time to go to bed at night, waking up in the early hours of morning, feeling tired and drowsy while attending class.

Data obtained in the study of Marzieh Nojoumi et al. were analyzed through SPSS13 software using descriptive and analytic statistics. Among the types of insomnia, the broken sleep (49%) and the sleep latency period more than one hour (3.7%) had the most and the least prevalence, respectively. Among parasomnia the nightmare (32%) and the nocturnal eating habit (1%) had the most and the least prevalence, respectively. Types of sleep disorders were associated with extreme tiredness during the day and dissatisfaction with his performance, which was significantly higher than the prevalence of insomnia in women and the intern group. As a result, various sleep disorders cause dissatisfaction with performance and a feeling of tiredness during the day, especially in women.

A descriptive cross-sectional research was conducted by Mohammad Ali Mohammadi et al. and they studied the prevalence of various sleep disorders among 400 people. Data were collected using a regulatory questionnaire in demographic sections to assess sleep onset disorder, sleep continuity disorder and waking up disorder. The results showed that there was a statistically significant relationship between sleep disorders and medicine consumption, employment, place of residence and marital status [20].

Reza Qanei'e et al. have conducted a study and used convenience sampling. Pittsburgh Sleep Quality Index (PSQI) was used to collect data. There is statistically significant relationship between these sleeping pills and sleep quality. So that 93.3% of people who took the medicine had poor sleep quality.

5 Conclusion

Correct diagnosis of factors affecting sleep quality using artificial intelligence and machine learning will increase the chances of successful treatment of sleep-related diseases. In this article, we have used data mining algorithms to examine the relationship between sleep quality and having a bed partner. The simulation results show that the nearest neighbor algorithm with detection accuracy of 0.73 is higher in accuracy than the Naive Bayesian Classification and SVM. The rate of accuracy obtained using data mining algorithms in predicting the sleep quality using the feature of having a bed partner suggest that there is s relationship between having a suitable bed partner and high quality of sleep. The complexity and time consuming implementation are the weaknesses of these methods.

References

[1] Williams S. Mental health psychiatric nursing. St. Louise: Philadelphia: Mosby; 1993.
[2] Lima PF, Medeiros ALD, Araujo JF. Sleep-wake pattern of medical students: early versus late class.
[3] Ayas NT, White DP, Manson JE, et al. A prospective study of sleep duration and coronary heart disease in women: Arch Intern Med. 2003; 163(2): 205-9.
[4] Leproult R, Buxton O, Van Cauter E. Nocturnal sleep deprivation results in an elevation of cortisol levels the next evening. Sleep. 1997; 20(10): 865-870.
[5] Learthart S. Health effects of internal rotation of shifts. Nurs Stand. 2000; 14(47): 34-37
[6] Sathyanarayana A, Joty Sh, Fernandez-Luque L, Ofli F, et al. Sleep Quality Prediction From Wearable Data Using Deep Learning. JMIR Mhealth Uhealth. 2016 Oct-Dec; 4(4): e125.
[7] Kelman BB. The sleep needs of adolescents. J Sch Nurs. 1999; 15(3): 14-9.
[8] Pagal JF, Kwaitkowski CF. Sleep complaints affecting school performance at. different educational levels. Front Neurol 2010; 16: 125.
[9] Mosavi F,Golestan B,Mateini A, Study of prevalence of sleep disorders and its related factors in medical students of Islamic Azad University, Tehran, Medical Journal Quarterly. 1389; 20 (4): 278-284.
[10] Mirzaei N, Nasiri T, Fatemi M, Akbari A. Prevalence of sleep disturbances among students in North Khorasan University of Medical Sciences, Student Research Committee, Bojnourd, Iran, 2012.
[11] Nojomi M. Ghalenoi M,Akhbari R, Gorji R, Sleep Pattern and the Prevalence of Types of Sleep Disorders in Medical Students and Specialized Assistants. Journal of Medical Sciences, Islamic Azad University, Vol. 19, No. 1, Spring 88: Pages 55-59.
[12] Islamic A. Prophet Types of Sleep Disorders and Some Related Causes and Consequences in Students Residing in Dormitories of Jahrom University of Medical Sciences and Health Services, Journal of Jahrom University of Medical Sciences, Vol. 9, No. 4, Winter 90.

ISSN- 2394-5125 VOL 7, ISSUE 08, 2020

[13] Ghanei R . Sleep quality of nursing students living in dormitories of Urmia University of Medical Sciences, Urmia University of Medical Sciences, Nursing and Midwifery Faculty, Nursing, Faculty of Nursing and Midwifery, Urmia, Nineteenth, No. 4, pp. 33: October 2011.
[14] Ghajanlu A and aririan H. Sleep quality of students during the period of exams in Zanjan University of Medical Sciences. New care, Journal of Nursing and Midwifery, Birjand University of Medical Sciences. Volume 8, Issue 4, Year 2011.
[15] Vapnik, V. Support-vector networks. Machine Learning 20 (3). 1995: 273–297.
[16] Altman N. S. An introduction to kernel and nearest-neighbor nonparametric regression. The American Statistician 46 (3). 1992: 175–185.
[17] Russell S, Norvig P. Artificial Intelligence: A Modern Approach (2nd ed.). Prentice Hall. (2003) [1995].
[18] Dodge Y . The Oxford Dictionary of Statistical Terms, OUP; 2003.
[19] Han J, Kamber M, Pei j. Data Mining: Concepts and Techniques, 3rd edition, Morgan Kaufmann; 2011.
[20] Mohammadi M, Dadkhah B. Prevalence of sleep disorders in university students in Ardabil, Journal of Nursing and Midwifery Faculty, 11th year, Winter 2009.
[21] Lima P.F, Medeiros A.L.D, Araujo J.F. Sleep-wake pattern of medical students: early versus late class starting time. Braz J Med Biol Res. 2002; 35(11): 1373-1377.