Clustering Students based Sleep Quality and Academic Performance using Educational Data Mining

Meichell Jynein J. Managing
BSCS – 3rd Year Student
093 Dama de Noche, Poblacion,
Kalamansig, Sultan Kudarat
mjjmanaging@usep.edu.ph

Joseven R. Francisco
BSCS – 3rd Year Student
Blk E12 Lot 18, M. Ponce St.
Phase II, Matina, Davao City
jrfrancisco@usep.edu.ph

Bezalel O. Delos Reyes
BSCS – 3rd Year Student
Prk. Villarosal, Magugpo
South, Tagum City
bodelosreyes@usep.edu.ph

ABSTRACT

Cognitive skills of a person can be affected by their sleep quality. Sleep is very crucial for every person's learning specially the memory retention, as well as physical and mental health. Various studies have been conducted about how sleep quality affects student's academic performance. In this study, the proponents aim to apply machine learning specifically K-means algorithm in clustering students based on sleep quality data and academic performance. The proponents are successful in utilizing EDM methods in mining and analysing the students' sleep pattern and academic performance data. The findings of such an analysis may lead to the discovery of the relationship between students' sleep quality and academic performance and aid advisors to conduct plans in order to help students improve academic performance. Results shows that majority of the students tend to have better sleep quality and academic performance, while minority of the students have poor sleep quality and academic performance. Therefore, sleep quality has a difference between students with high academic performance and those with low academic performance. Result also shows that mental illnesses affects the student's sleep quality. The study showed that minority of the students who struggled in mental health are also having difficulty in sleeping. All in all, these results can be improve if the sample size of the dataset is increase and to use multiple sources of within the same study such as sleep laboratory test, additional personal information like parent data.

Keywords

Data Mining, Sleep Quality, EDM, K-means Algorithm, Mental Illness, Clustering, RapidMiner

1. INTRODUCTION

Cognitive skills of a person can be affected by their sleep quality. Sleep is very crucial for every person's learning specially the memory retention, as well as physical and mental health. Poor night-time sleep quality and the consequent daytime sleepiness affect student's academic performance [1]. Academic performance is considered as the center where the whole educational system is revolving. One of factors that helps in enhancing students' academic performance is enough sleep every night [2] But, staying up late become more popular nowadays which most students often do.

Various studies have been conducted about the relationship of sleep quality and academic performance. Results revealed that insufficient sleep, increased frequency of short-term sleep, and getting up early affect the learning capacity, academic performance, and neurobehavioral functions of a student [3]. Also, a related study stated that quantity of sleep as delayed or inappropriate sleep, waking up too late, especially at weekends and daytime sleepiness is associated with compromised academic performance in children and adults [4].

Results from related studies presents that student's sleep quality and academic performance is relational [3]. This study aims to apply machine learning specifically K-means algorithm in clustering students based on the gathered data on their sleep quality, psychological and emotional signs, and academic performance. The study helps in distinguishing the relationship of students' sleep quality, psychological and emotional signs, and academic performance to each other. The computational results of this study may use as a contribution in assessing students' problems in sleeping and in their academic performance.

1.1 OBJECTIVES OF THE STUDY

The purpose of this study is to classify college students through educational data mining with regards to their sleep quality and academic performance. Specifically this study aims to:

1. To develop a clustering model to assess the data of the student's sleep quality and academic performance.

2. To determine if mental illnesses (stress, anxiety, and depression) have any relationship in student's sleep quality.

3. To distinguish the relationship between students' sleep quality and academic performance.

1.2 SIGNIFICANCE OF THE STUDY

This research will provide insights on how sleep quality contributes to mental issues and to academic performance of college students through machine learning computations.

Community – Community will be aware that frequent low sleep quality can be a threat to mental health and how mental issues can affect the academic performance of the students.

Academic Institutions and administrators – Academic institutions and administrators will be informed about the impact of sleep quality to students' mental health and academic performance and may promote programs and advocacies for it.

Mental health advocates – the result of this research will provide valuable information for them to widen their advocacies and programs on how sleep play apart in mental health awareness.

Students – findings of this research can make students be aware of advantages and disadvantages of sleep quality to their academic performance.

Future researchers – Future researchers can use this research as source, additional evidence, and for future discussion on the relationship of sleep quality and academic performance.

2. REVIEW RELATED LITERATURE

2.1 SLEEP QUALITY ON STUDENTS' ACADEMIC PERFORMANCE

Problems with sleeping are very common among young adults and has a high significance when related to their quality of life [5]. According to the study of [6], students whom reported having sleep problems with insufficient sleep as the most common complaint, taking naps and adjusting sleeping schedules were the coping strategies to attain better sleep quality. While on the other hand, students whom reported trying to do a sleep-encouraging activity, not keeping in mind their sleep problems altogether, or failing to find a way of coping with their sleep problems outputs a poorer sleep quality. Moreover, based on the study of [7] about sleep quality among pharmacy students, students who does have higher GPA tends to have a good and enough quality of sleep, while most of those within the lower GPA mark resulted from lack of sleep or poor quality of sleep. In addition, [8] study towards senior high school adolescents in Taiwan about sleep problems indicated that, sleep problems, psychosocial impacts, and medical disorders has something to deal with fatigue and daytime sleepiness in adolescents. Furthermore, being an adolescent requires to have similar duration of sleep to younger children, up to 9-9.25 hours per night. [9] research about the impact of poor sleep quality on the academic performance of medical students showed that, using Pittsburgh Sleep Quality Index (PSQI) to assess the students sleep quality and analyzing the data using the Statistical Package for the Social Sciences (SPSS), students who had lower sleep duration experienced daytime dysfunction almost daily and leads to poor academic performance and those who are well rested refreshes their mind and does perform good towards their academic performance. Since there still remains a lacking of quantitative data using objective measures to pursue the relatedness of sleep and academic performance the research of [10] lends the surveyed students a wearable activity tracker enabling to measure multiple sleep patterns and to be correlated with in class performance such as test, quizzes, and midterm examinations. Better quality, longer duration and the consistency of sleep correlated with good academic performance resulting to high grades.

2.2 MENTAL ILLNESS ON STUDENTS' ACADEMIC PERFORMANCE

Individuals who suffer from health conditions including changes in emotion, thinking or behaviour (or combined) and associated with distress or problems doing well in work, school or family activities may have what the society called as "Mental Illness" [11]. [12] stated that, mental illness can turn your enjoyable life in to a miserable one by disrupting your motivation in school or work or in relationships, examples of mental illnesses are depression, anxiety disorders, schizophrenia, eating disorders and addictive behaviours. In addition, mental illness comes from not a single cause but many factors such as: genes & family history, an individual's life experiences from the past relating to history of abuse, especially if they happened in childhood, brain injury, use of alcohol or illegal drugs, having a serious medical condition like cancer, and feeling lonely or isolated [13]. According to [14], the types of anxiety disorders are: Generalized Anxiety Disorder (GAD), Panic Disorder, Social Anxiety Disorder, Specific Phobias, Obsessive-Compulsive Disorder (OCD), and Post-traumatic Stress-Disorder (PTSD). Moreover, all these types of disorders have a significant relationship between sleep. A common example of symptom of anxiety disorder in sleep is insomnia. In addition, [15] stated that, having a mental illness can lead to sleep problems such as: sleep apnea, narcolepsy, and insomnia. Based on [16], people experiencing poor sleep can create conditions for negative thinking and emotional vulnerability, while having a good night's sleep helps a person to strengthen his/her physical and mental state. When you ask what comes first either anxiety disorder or sleep disorder the answer would be both since anxiety causes sleeping problems and sleep deprivation causes anxiety disorder. Furthermore, the risk of lacking sleep extend way beyond just being tired, people with anxiety and sleep disorders tends to get heart disease, heart failure, irregular heartbeat, heart attack, high blood pressure, stroke, diabetes, and obesity [27].

2.3 CLUSTERING

The study of [21] involving K-means clustering to study how reasoning lines can be modified by a learning activity based on Feynman's Unifying Approach resulted to students initially highlight the use of lines of reasoning based on the use of memory of past studies and on an application of mathematics, without a search for a proper mechanism of functioning. In some cases, everyday-like reasoning is also highlighted. Moreover, Feynman Unifying Approach affected the students' reasoning lines to have clearly changed in to explicative ones. Furthermore, [22] found that by using K-Means to determine learner typologies for students' project-based learning, it helped the instructors in grouping similar traits learners according to their strength and weaknesses and showed that K-means unsupervised learning algorithm did show significant results and is useful in grouping learners with similar concept score. Based on [20] about clustering insomnia patterns by using wearable devices such as smart bands, found out that unsupervised learning enables medical practitioners to focus on precise and accurate interventions at the level of data-guided user clusters that would lead to a novel answer in treating insomnia and other mental disorders. According to [19] research about applying cluster analysis of sleeping patterns of infants and the association with maternal health, the researchers found out that after gathering four distinct sleep profiles, two infant sleep profiles are significantly associated with maternal stress, depression, and poorer self-report of health. Furthermore, sleep pattern characterization was done via cluster analysis of audio data showed that using self-organizing neural-network based approach is feasible in analyzing sleep disorders [18].

2.4 DATA MINING

According to [17] study about data mining in medical field, data-mining technology has been an innovative change in medical research, also data mining has unique advantages in clinical big-data research, specifically in large-scale medical

databases. Moreover, the emergence of mining big data in educational perspective has led researchers in to a new data-driven approaches and help improve educational effectiveness [23]. In addition, [28] compared learning analytics and educational data mining by using a topic model approach and found out that even if there appears a disciplinary differences in both disciplines, still there is a little support for a clear distinction of the two approach and also both fields have converged on an ascending focus on student behaviour over the last years. Also, a bibliometric analysis was done with regards to educational data mining by [29] to show that interest in educational data mining has increased in recent years. Based on the study of [24] about the use of data mining and analytics in education, there is a potential and many possibilities in enhancing the quality of educational resources, finding at-risk students, and providing a better support to professors and students from increasing amount of data that is now available to gather. Hopefully, students and instructors will be innovative and not staying the traditional pattern in terms of education because of data mining and analytics. The study of [25] inclined to predicting students' academic performance using machine learning algorithms, showed that, students' midterm exam grades play a vital role in predicting their final exam grades and data mining can help predict students' academic performance. Furthermore, [26] found that demographic such as gender, age, economic status, number of courses attended, internet access, and geographic data of students have significant effect on students' academic performance.

3. METHODOLOGY

This chapter presents the process of the study. The methodological process of the study can be seen in the following figure. This incorporates dataset, pre-processing, educational data mining, and evaluation of the study.

Figure 1. Methodological Process.

3.1 DATASET

The dataset was obtained from R package called Lock5withR. The dataset was consist of data from a study pertaining to college students' sleeping patterns. The data was collected on weekdays in November during the fall semester [27]. Below discussed on how the data was gathered and what type of data gathering method was employed.

3.1.1 PARTICIPANTS

Participant were undergraduate college students at a liberal arts college in the northeastern United States. Participants gave their approval for the study and received credit for their participation as well as for accessing their GPA. A total of 255 participants took part in the study, however, two participants' data were discarded as unusable. Overall, the remaining participants completed the questionnaires on sleep, behaviour, mood, health, and substance use, as well as a sleep survey and cognitive function tests [27].

3.1.2 QUESTIONNAIRE

On assessing the sleep quality of the study, Pittsburgh Sleep Quality Index were implemented. A higher total score indicates a poorer quality of sleep [28]. Participants completed a measure of circadian preference (the Owl-Lark scale) adapted by [29] in which they selected preferred schedules for daily activities. A higher scores indicate morning preference; α =.81 [27]. Another, participants completed a questionnaire adapted from [31] [30] on daytime sleepiness and sleep-wake behaviour problems. The participants were asked whether they had struggled to remain awake or had fallen asleep in 12 different situations in the previous two weeks (e.g., while studying, watching television, etc.). Participants were also asked how often they arrived late to class, felt tired during the day, stayed up all night, had trouble falling asleep, or had sleep problems (e.g., snored, had nightmares, woke up during the night) in the previous two weeks [27]. Another, the Depression, Anxiety, and Stress scale of [34] was used to assess mood of the participants, specifying higher values (α = .94) imply a negative mood.

Student indicated whether they avoid drinking alcohol or taking a light, moderate or heavy alcohol consumption. Participants also filled out a brief health questionnaire, which inquired about cold and flu occurrence (since the beginning of the semester), doctor, hospital, or health center visits, and the number of courses missed. Finally, participants determined their aggregate GPA, and 80% of the example consented to have the present semester's GPA acquired from source records once the semester finished [27].

3.1.3 EVALUATION OF COGNITIVE FUNCTION

Participants conducted a delayed memory recall task (recalling a list of 30 common words 10 minutes after they were studied), Digit Symbol Coding Task [32], a measure of perceptual-motor speed, and Letter Cancellation Task [33], a measure of sustained attention to assess cognitive performance. Scores on the three performance measures were standardized and merged together to construct an index of cognitive performance [27].

3.2 PRE-PROCESSING

In preparing the data, missing values inside the data were removed since this missing data will hinder the procedure and may result to unreliable and poor output. Unnecessary attributes presented at the dataset such as Gender, ClassYear, LarkOwl, NumEarlyClass, AlcoholUse, etc. were removed. This reduces the size of the dataset while maintaining the integrity of the original dataset. Thus, this will make the dataset efficient and decent to work with.

The implementation of college students' data on data mining using K-means clustering method uses 2 attributes namely Student's Grade (GPA), and Sleep Quality (PoorSleepQuality). These data will be processed in 2 clusters (Cluster 1) Poor-Sleep Quality-Grades and (Cluster 2) Rich-Sleep Quality-Grades. Aside from that, another 2 attributes such as Depression Anxiety Stress Scale (DASS) and Sleep Quality (PoorSleepQuality) are used in determining the relationship between students' mental illness and sleep quality.

3.3 EDUCATIONAL DATA MINING

3.3.1 RAPIDMINER

The RapidMiner tool is used for this experiment. RapidMiner is a software data mining application developed by the company of the same name that can handle everything from data mining through model deployment and model operations. It can also provide data science platform that includes data preparation, validation, optimization, and machine learning capabilities [35]. It's utilized for research, teaching, training, fast prototyping, and application development, as well as for business and commercial applications [36]. Here, procedures in this study such as data cleaning, data transformation, visualization, designing the model are all performed in this tool.

3.3.2 NORMALIZATION

In order to generate effective findings, normalization is performed to the dataset. It is a method for standardizing and weighting all of the attributes in a dataset so that redundant data can be removed which creates valid and reliable data and this data can be used to improve the accuracy of the output [38]. There are several data normalization to be used and it depends on the data to be normalized. In our case, we utilized Z-transformation method for normalization in the attributes assigned for data mining.

3.3.3 K-MEANS CLUSTERING

Cluster analysis is an effective method for gathering big clusters of existing research topics [37]. In the clustering of college students' data, K-means algorithm is used to cluster the students based on sleep quality and academic performance attributes. A study proposed a data mining method based on K-means clustering algorithm in determining the relationship between social media activities and a rare event [40]. The said algorithm will also be used in determining the relationship between student's mental illness data and sleep quality data. The number of K assigned on the clustering method parameter is two since we only want to distinguish two clusters in each objective. Table 1 below shows the attributes that were pre-processed throughout the study.

Table 1. Attributes of the data and its description.

Attributes	Description
GPA	Students' grade point average (0-4 scale)
PoorSleep Quality	Measure of students sleep quality (higher values are poorer sleep)
Depression Score	Measure of degree of depression (high number indicate severe depression)
AnxietyScore	Measure of amount of anxiety (high number indicate severe anxiety)
StressScore	Measure of amount of stress (high number indicate severe stress)
DASScore	Combined score of depression, anxiety and stress

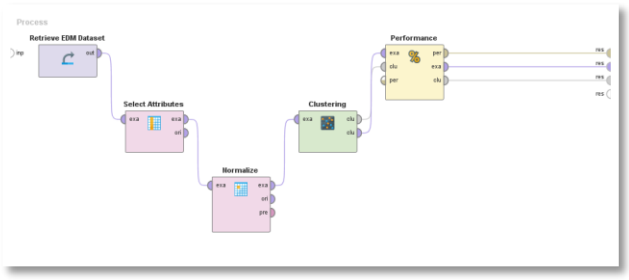

Figure 2. Design of K-means clustering algorithm.

Figure 2 describes the process of the study. First, the retrieval of the dataset called "EDM Dataset" is initialized. All the attributes are now set to be used by the model. We then applied "Select Attributes" since we only utilized the attributes shown in Table 1. After selecting the attributes needed for applying the model, we first normalize the attributes by using "Normalize" operator. The method used in normalizing the attributes was Z-transformation. This eliminates redundant data and developed data integrity. Next, the attributes are then clustered with the used of "Clustering" operators. The k parameter are set into two because we only have to determine two groups. Lastly, Davies Bouldin Index are used to evaluate the cluster distance performance of K-means clustering algorithm.

3.4 EVALUATION

Here, clusters are assessed based on some measure of similarity, such as the distance between cluster points. In this case, we utilized Davies-Bouldin score as clustering validation. The lowest score is zero, and a model with a lower Davies-Bouldin index score indicating better clustering [39].

4. RESULTS AND DISCUSSIONS

4.1 CLUSTER PERFORMANCE ON STUDENTS' SLEEP QUALITY AND DASS DATA.

In finding relationship between sleep quality and DASS data, we utilized the design of K-mean algorithm shown in Figure 2. In the "Select Attributes" operator shown in Figure 2, sleep quality and DASS data was selected to apply for this objective. Below shows the result of K-means clustering method.

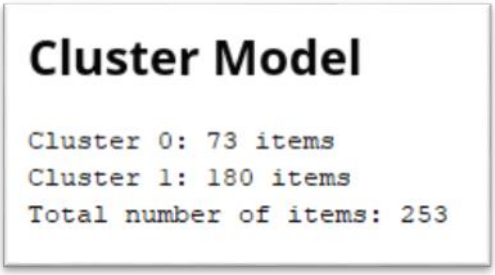

Figure 3. Cluster Model Description on Sleep Quality and DASS data.

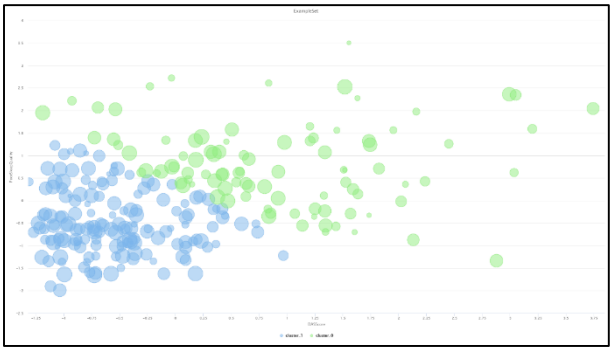

Figure 4. Result of K-means Clustering Method on Sleep Quality and DASS data.

In figure 4, result of K-means clustering algorithm using sleep quality and DASS score was presented. Poor sleep quality score were set as Y-axis while DASS score were assigned as X-axis. Here, cluster 0 denoted by green dot contains 73 students who have high DASS scores and have high poor sleep quality scores which indicates that they have poor sleep. This specifies that they struggled in mental health and have difficulty in sleeping. Whereas, cluster 1 assigned by blue dot contains 180 students with low DASS scores and low poor sleep quality scores. This cluster indicates that they are mentally stable and have pleasant sleep. Based on the result, this defines that mental illness such as depression, anxiety and stress have a significant impact on students' sleep quality.

4.2 CLUSTER PERFORMANCE ON STUDENTS' SLEEP QUALITY AND GPA DATA.

Cluster Model

```
Cluster 0: 92 items
Cluster 1: 161 items
Total number of items: 253
```

Figure 5. Cluster Model Description on Sleep Quality and Academic Performance data.

In Figure 5, it shows the description of the cluster model of students' sleep quality and academic performance data produced with RapidMiner tools. In this case, students' sleep quality and GPA were selected in "Select Attributes" operator shown in Figure 1. Here, cluster 0 which is symbolized by the green dots, contains 92 students while cluster 1 denoted by blue dots, contains 161 students with a total of 253 students. In this case, cluster 1 has the highest number of students clustered than cluster 0.

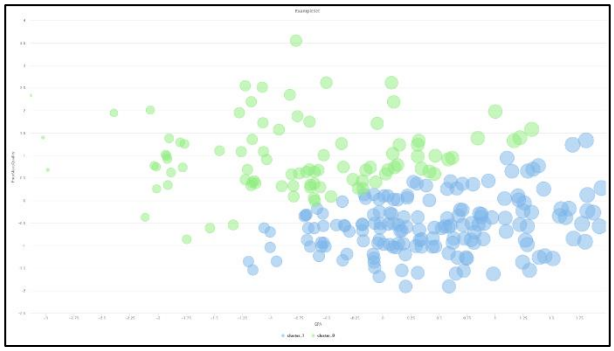

Figure 6. Result of K-means clustering method on Sleep Quality and Academic Performance Data.

Figure 6 above displays the results of K-means clustering method with RapidMiner. Shown in the plot, Y-axis were assigned as poor sleep quality scores of the students while X-axis of the plot were conveyed as student's grade. In this figure, Cluster 0 indicates students who have high poor sleep quality scores and has low to moderate grades. This shows that students who has difficulty in sleeping tend to have low to moderate grades. On the other hand, Cluster 1 indicates students who have low poor sleep quality scores and has moderate to high grades compared to the grades of Cluster 0 students. In other words, students were incline to have high grades when they have efficient sleep. Overall, results show that sleep quality affects students' academic performance.

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 1.122
Avg. within centroid distance_cluster_0: 1.503
Avg. within centroid distance_cluster_1: 0.905
Davies Bouldin: 1.001
```

Figure 7. Performance accuracy of the study.

In Figure 7, this shows the performance accuracy using % Performance operator. Davies Bouldin Index is used to validate the K-means model. It's defined as an average similarity measure of each cluster with its most comparable cluster. In other words, clusters that are farther apart and less dispersed will score better [30]. Here, the study of [31] stated that when the value of DBI score is lesser than K, that value of K is better for clustering. The results obtained on Davies-Bouldin Index for clustering students sleep quality and academic performance data is 1.001. Hence, K = 2 achieved a significant result producing 2 clusters in this study shown in Figure 6.

5. CONCLUSIONS AND RECOMMENDATIONS

The aim of this study was to categorize students based on sleep quality and academic performance. The researchers have been successful in using K-means to categorize students based on sleep quality and academic performance. Based on the results, majority of the students tend to have better sleep quality and academic performance, while minority of the students have poor sleep quality and academic performance. Therefore, sleep quality has a difference between students with high academic

performance and those with low academic performance. The result is similar to the study of [32] wherein their study stated that sleep quality correlated with students' exam grades. The authors expressed that poor sleep quality implied poor academic performance. Another result of the study of [34] is associated with our findings, where better sleep quality is associated with higher academic performance. In determining the relationship between students' mental illness and sleep quality, minority of the students struggled in mental health and having difficulty in sleeping. Another study of [33] showed an association between sleep quality and mental illness. The study showed that high levels of stress among medical students influenced their quality of sleep. Hence, mental illness has an effect on students' sleep quality.

Overall, the method provides sufficient results which acknowledged the objectives of the study. However, there are some recommendations for this study. One is to increase the sample size of the dataset. Having bigger sample size provide more accurate findings. Second is to use multiple sources of within the same study such as sleep laboratory test, additional personal information like parent data, even though the current data used a variety of measurement in acquiring students' sleep and academic performance data. A multi-measure method will provide a more thorough and potentially more accurate assessment [35]. All in all, these recommendations can be used for future research which can draw conclusions with a valuable findings.

6. ACKNOWLEDGMENTS

First and foremost, praises and thanks to the God, the Almighty, for His showers of blessings throughout our research work to complete the research successfully. The researchers would like to express their deep and sincere gratitude to our subject advisor Ms. Maureen Villamor for providing vital guidance and comments during this process. The proponents are extending our thanks to our friends and family for their support and sacrifices. Additionally, the researchers are extremely grateful to their fellow researchers for their constant encouragement throughout this research work. Finally, the researcher's thanks go to all the people who have supported us throughout the duration of the study.

7. REFERENCES

[1] Maheshwari G, Shaukat F (April 01, 2019) Impact of Poor Sleep Quality on the Academic Performance of Medical Students. Cureus 11(4): e4357. doi:10.7759/cureus.4357

[2] Tran, T.(ND). IBD's Student Attitude About Staying Up Late. Retrieved from https://www.academia.edu/31876349/IBDs_Student_Attitude_About_Staying_Up_Late

[3] Jalali, Rostam, et al. "The Effect of Sleep Quality on Students' Academic Achievement." *Advances in Medical Education and Practice*, vol. Volume 11, July 2020, pp. 497–502, www.ncbi.nlm.nih.gov/pmc/articles/PMC7381801/pdf/amep-11-497.pdf, 10.2147/amep.s261525.

[4] Cort-Blackson, Maureen. *ScholarWorks the Effects of Sleep Deprivation on Online University Students' Performance*. 2018.

[5] Yilmaz, D., Tanrikulu, F., & Dikmen, Y. (2017). Research on Sleep Quality and the Factors Affecting the Sleep Quality of the Nursing Students. Current health sciences journal, 43(1), 20–24. https://doi.org/10.12865/CHSJ.43.01.03

[6] Yang M, Wu H, Hsieh M, Liu M, Lu F. Coping with sleep disturbances among young adults: a survey of first-year college students in Taiwan. Behavioral Medicine.

[7] Marshall E. Cates, Andraya Clark, Thomas W. Woolley and Amy Saunders American Journal of Pharmaceutical Education February 2015, 79 (1) 09; DOI: https://doi.org/10.5688/ajpe79109

[8] Chen, T. Y., Chou, Y. C., Tzeng, N. S., Chang, H. A., Kuo, S. C., Pan, P. Y., Yeh, Y. W., Yeh, C. B., & Mao, W. C. (2015). Effects of a selective educational system on fatigue, sleep problems, daytime sleepiness, and depression among senior high school adolescents in Taiwan. Neuropsychiatric disease and treatment, 11, 741–750. https://doi.org/10.2147/NDT.S77179

[9] Maheshwari, G., & Shaukat, F. (2019). Impact of Poor Sleep Quality on the Academic Performance of Medical Students. Cureus, 11(4), e4357. https://doi.org/10.7759/cureus.4357

[10] Okano, K., Kaczmarzyk, J.R., Dave, N. et al. Sleep quality, duration, and consistency are associated with better academic performance in college students. npj Sci. Learn. 4, 16 (2019). https://doi.org/10.1038/s41539-019-0055-z

[11] Mayo Clinic. (2019). Mental Illness. Retrieved April 23, from https://www.mayoclinic.org/diseases-conditions/mental-illness/symptoms-causes/syc-20374968

[12]American Psychiatric Association. (2018). What is Mental Illness? Retrieved April 23, from https://www.psychiatry.org/patients-families/what-is-mental-illness

[13] MedlinePlus. (2022, March 15). Mental Disorders. https://medlineplus.gov/mentaldisorders.html

[14] Agnafors, S., Barmark, M. & Sydsjö, G. Mental health and academic performance: a study on selection and causation effects from childhood to early adulthood. Soc Psychiatry Psychiatr Epidemiol 56, 857–866 (2021). https://doi.org/10.1007/s00127-020-01934-5

[15] Mental Health Foundation. (2021, September 14). Sleep and mental health. https://www.mentalhealth.org.uk/a-to-z/s/sleep-and-mental-health

[16] Health Navigator New Zealand. (2021, April 27). How does sleep affect mental wellbeing? https://www.healthnavigator.org.nz/healthy-living/s/sleep-and-mental-health/

[17] Wu, WT., Li, YJ., Feng, AZ. et al. Data mining in clinical big data: the frequently used databases, steps, and methodological models. Military Med Res 8, 44 (2021). https://doi.org/10.1186/s40779-021-00338-z

[18]] Hongle WU, Ken-ichi FUKUI, Takafumi KATO, Masayuki NUMAO, Sleep Pattern Characterization via Cluster Analysis of Audio Data, JSAI Technical Report, SIG-KBS, 2015, Volume 106, 106th (Nov, 2015), Pages 08-, Released on J-STAGE July 14, 2021, Online ISSN 2436-4592, https://doi.org/10.11517/jsaikbs.106.0_08

[19] Hughes, A., Gallagher, S., & Hannigan, A. (2015). A Cluster Analysis of Reported Sleeping Patterns of 9-Month Old Infants and the Association with Maternal Health: Results from a Population Based Cohort Study. Maternal and child health journal, 19(8), 1881–1889. https://doi.org/10.1007/s10995-015-1701-6

[20] Park, S., Lee, S. W., Han, S., & Cha, M. (2019). Clustering Insomnia Patterns by Data From Wearable Devices: Algorithm Development and Validation Study. JMIR mHealth and uHealth, 7(12), e14473. https://doi.org/10.2196/14473

[21] Battaglia, O. R., Di Paola, B., & Claudio, F. (2017). K-means Clustering to Study How Student. EURASIA Journal of Mathematics Science and Technology Education.

[22] Dake, D. K., & Gyimah, E. (2019). Using K-Means to Determine Learner Typologies for Project-based Learning: A Case Study of the University. International Journal of Computer Applications, 29-33.

[23] Fischer, C., Pardos, Z. A., Baker, R. S., Williams, J. J., Smyth, P., Yu, R., Slater, S., Baker, R., & Warschauer, M. (2020). Mining Big Data in Education: Affordances and Challenges. Review of Research in Education, 44(1), 130–160. https://doi.org/10.3102/0091732X20903304

[24] Baker, S., & Inventado, P. S. (2016). Educational data mining and learning analytics: Potentials and possibilities for online education. In G. Veletsianos (Ed.), Emergence and Innovation in Digital Learning (83–98). doi:10.15215/aupress/9781771991490.01

[25] Yağcı, M. Educational data mining: prediction of students' academic performance using machine learning algorithms. Smart Learn. Environ. 9, 11 (2022). https://doi.org/10.1186/s40561-022-00192-z

[26] Waheed, H., Hassan, S. U., Aljohani, N. R., Hardman, J., Alelyani, S., & Nawaz, R. (2020). Predicting academic performance of students from VLE big data using deep learning models. Computers in Human Behavior, 104(October 2019), 106189. https://doi.org/10.1016/j.chb.2019.106189

[27] Anxiety and Depression Association of America. (2021, October 19). Sleep Disorders. https://adaa.org/understanding-anxiety/related-illnesses/sleep-disorders

[28] David J. Lemay, Clare Baek, Tenzin Doleck, Comparison of learning analytics and educational data mining: A topic modeling approach, Computers and Education: Artificial Intelligence, Volume 2, 2021, 100016, ISSN 2666-920X, https://doi.org/10.1016/j.caeai.2021.100016.

[29] Baek, C., & Tenzin, D. (2022). Educational Data Mining: A Bibliometric Analysis of an Emerging Field. IEEE Access, 31289 - 31296.

[30] Zuccarelli, E. (2021, December 29). Performance Metrics in ML - Part 3: Clustering | Towards Data Science. Medium. https://towardsdatascience.com/performance-metrics-in-machine-learning-part-3-clustering-d69550662dc6

[31] Singh, A. K., Mittal, S., Malhotra, P., & Srivastava, Y. V. (2020). Clustering Evaluation by Davies-Bouldin Index(DBI) in Cereal data using K-Means. 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC). https://doi.org/10.1109/iccmc48092.2020.iccmc-00057

[32] Ahrberg, K., Dresler, M., Niedermaier, S., Steiger, A., & Genzel, L. (2012). The interaction between sleep quality and academic performance. Journal of psychiatric research, 46(12), 1618-1622.

[33] Ahrberg, K., Dresler, M., Niedermaier, S., Steiger, A., & Genzel, L. (2012). The interaction between sleep quality and academic performance. Journal of psychiatric research, 46(12), 1618-1622.

[34] Tonetti, L., Fabbri, M., Filardi, M., Martoni, M., & Natale, V. (2015). Effects of sleep timing, sleep quality and sleep duration on school achievement in adolescents. Sleep medicine, 16(8), 936-940.

[35] Wolfson, A. R., & Carskadon, M. A. (2003). Understanding adolescent's sleep patterns and school performance: a critical appraisal. Sleep medicine reviews, 7(6), 491-506.