Clustering Students based on Alcohol Consumption, Sleep Quality, and Academic Performance Using K-means Algorithm

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., S.I.R. Phase II, Matina, Davao City
jrfrancisco@usep.edu.ph

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

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Sleep quality and alcohol consumption contribute significant effects in cognitive skills of a person, specifically a student. Sleep is very crucial for student's learning specially the memory retention, as well as physical and mental health. Poor nighttime sleep quality and the consequent daytime sleepiness affect student's academic performance [1]. Inappropriate alcohol consumption among students is a major public concern due to its effects to psychological and mental state. Large consumption of alcoholic drinks will lead to deterioration of cognitive function and worsening student's academic performance [2]. Enough sleep and avoiding alcoholic drinks help in enhancing students' academic performance [1, 2] but, staying up late and binge drinking are the activities that become more popular nowadays which most students often do.

Various studies have been conducted about the relationship of sleep quality and academic performance and results revealed that insufficient sleep, increased frequency of short-term sleep, sleeping late and getting up early affect the learning capacity, academic performance, and neurobehavioral functions [3]. A related study stated that extreme alcohol consumption will lead to memory disorders; daily drinking was negatively related with academic performance; and problematic use of alcohol by the students was associated with low academic performance [2].

It is generally assumed from abovementioned results of related studies that sleep quality, alcohol consumption, and academic performance of a student is relational [2, 4]. However, this study sets to apply machine learning specifically K-means algorithm in building a model for clustering students based on their sleep

quality, alcohol consumption, and academic performance and to distinguish its relationship to each other. The computational results and testing of this study may use as a contribution in assessing students' academic performance problems.

2. REVIEW OF 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 [18]. According to the study of [19], 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 [20] 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, [21] 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. [22] 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 [23] 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 Alcohol Consumption on Students' Academic Performance

There are many reasons why people consume alcohol such as social reasons, peer pressure, family history of alcoholism, stress, and mental health issues. However, some people consume more alcohol than others and those who are indulged so much in alcohol can develop a dependency on it [24]. Based on the study of [25] regarding with association of alcohol consumption and mental health among young university students, students who were consuming alcohol at a hazardous levels were more likely to be doing low performance in school also being late for class, inability to think straight while listening to lectures, and inability to complete tasks or assignments. Moreover, the research of [26] about if alcohol consumption is associated with poor academic performance, by using an online questionnaire with seven socio-demographic characteristics (age, gender, year/discipline of study, accommodation type, being in intimate relationship, parental education, and income sufficiency), two perceived academic performance (students' subjective importance of achieving good grades and students appraisal of their academic performance compared to peers), and six alcohol consumption behaviors (length of time, amount consumed, frequency, heavy episodic drinking, problem drinking, and possible alcohol dependence) showed that, study discipline, income sufficiency, importance of achieving good grades, and academic performance compared to peers were not associated with any alcohol behaviors, Universities also need to help and prevent the students in to consuming alcohol. Furthermore, [27] stated in her research that there was a negative correlation between grade point average (GPA) and frequency of alcohol consumption of college students, meaning to say that as the nights per week that a college student consumption of alcohol went up, grade point average went down.

2.3 Clustering

Clustering is the process of dividing data points in to a number of groups and the goal is to separate groups with similar characteristics or traits and then assign them to clusters [28]. Moreover, clustering can be divided into two subgroups namely: hard clustering and soft clustering where hard clustering means every node may belong to only one cluster while soft clustering means every node may belong to several clusters with a fractional degree of membership in each [29]. The types of clustering algorithms that handles unique kind of data are: density-based, distribution-based, centroid-based, and hierarchical-based [30]. Then we have K-Means clustering algorithm that is a simple yet beneficial algorithm in data science. According to [31], the K-Means algorithm is a centroid-based algorithm or a distance-based algorithm that computing the distances to assign a point to a cluster is a need and each cluster is associated with a centroid. In addition, the main objective of K-Means algorithm is to minimize the sum of distances between the points and their respective cluster centroid. The study of [32] 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, [33] 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.

2.4 Data Mining

According to [34] "Data mining, also called Knowledge Discovery in Databases (KDD), is the field of discovering novel and potentially useful information from large amounts of data. Data mining has been applied in a great number of fields, including retail sales, bioinformatics, and counter-terrorism. In recent years, there has been increasing interest in the use of data mining to investigate scientific questions within educational research, an area of inquiry termed educational data mining. Educational data mining (also referred to as "EDM") is defined as the area of scientific inquiry centered around the development of methods for making discoveries within the unique kinds of data that come from educational settings, and using those methods to better understand

students and the settings which they learn in." Based on the study of [35] 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. Moreover, 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 [36] 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, [37] 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.

2.5 Conceptual Framework

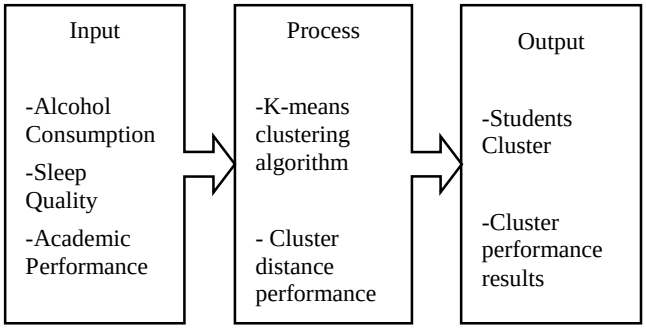

Figure 1. Conceptual Framework of the Study.

Figure shown above was developed to establish research limitations and seek individual factors influencing the successful clustering of students in the system. This will help the proponents in developing a model that can be used to find a solution. Moreover, the framework above has an input where data relating to alcohol consumption, sleep quality, and academic performance of students are gathered. Next, to make use of the datasets, the proponents applied K-means clustering algorithm to cluster each student according to their information and perform cluster distance performance to evaluate the cluster to further generate valid results. Finally, the output will be the clusters of the students based on their alcohol consumption, sleep quality, & academic performance data and result of the cluster performance.

3. METHODOLOGY

In this chapter, the process of the study is discussed. The processes includes the preparation and collection of dataset, data pre-processing, applying data mining, and evaluation of the results. Below show the methodological process of the study.

Figure 2. Methodological Process.

3.1 Dataset

The dataset was collected from Rdrr site, an accessible collection of R packages and documentation. The data was gathered on weekdays in November during the fall semester [5] and was consist of data on college students' sleep pattern and academic performance. The participants of the study were a total of two hundred fifty-five undergraduate college students but the data of the two participants were discarded as unusable. Overall, the remaining participants were two hundred fifty-three and filled out the questionnaires on sleep, behavior, mood, health, and substance use, as well as a sleep survey and cognitive function tests [5].

The questionnaire used in assessing the sleep quality of the students were Pittsburgh Sleep Quality Index. A lower total score of the questionnaire indicates wealthy sleep quality and a higher total score indicated poorer sleep quality. Participants were also tasked to complete a questionnaire adapted from [15] [14] on daytime sleepiness and sleep-wake behavior problems. This questionnaire asked participants whether they had struggled to remain awake or had fallen asleep or how often they arrived late to class, and felt tired during the day [5]. Another, the Depression, Anxiety, and Stress scale of [6] was used to assess mood of the participants, specifying higher values (α = .94) imply a negative mood. Participants were also asked whether they're taking a light, moderate, heavy, or avoid taking alcohol consumption. Participants also completed a brief health questionnaire, which asked about cold and flu occurrences, and course absences. Finally, participants calculated their overall GPA, with 80% of the study willing to have their current semester's GPA [5].

In conducting cognitive function, participants conducted three performance measures namely Memory Recall Task, Digit Symbol Coding Task [16], a measure of perceptual-motor speed, and Letter Cancellation Task [17], a measure of sustained attention to assess cognitive performance. The three performance tests' scores were normalized and combined to create a cognitive performance index [5].

3.2 Pre-processing

For data preparation, attributes such as GPA, NumOfDrinks, PoorSleepQuality were retain in the dataset and the rest of the attributes were removed. The removal of the unused attributes reduces the size of the dataset while maintaining the integrity of the original dataset. Decreasing of data volume aids in making analysis easier while producing the same or nearly identical results [7].Thus, this will make the dataset efficient and decent to work with.

Next, normalization is applied to the dataset in order to create useful results. It is a method for standardizing and weighting all of the attributes in a dataset in order to reduce redundant data, resulting in valid and reliable data that may be utilized to increase the accuracy of the output [13]. Data normalization can be done in a variety of ways, depending on the data to be normalized. In our case, Z-transformation will be used for normalization of the attributes.

3.3 Educational Data Mining

The tool used in this experiment is RapidMiner. It is a robust data mining application that can handle everything from data mining through model deployment and model operations [8]. This allows data mining, text mining, and predictive analytics. The user can enter raw data, such as databases and text, into the program, which is then automatically and intelligently analyzed on a huge scale [9]. The procedures in this study such as data preparation, data transformation, and applying machine learning method are all performed in this tool.

In clustering the students' data, the researchers utilized the K-means clustering algorithm. It's a method used for clustering analysis. It uses a simple process to divide a given data set into a number of clusters, each of which is designated by the letter "k,"

which is predetermined [10]. 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 [11]. In our case, K-means clustering will be used in finding relationship between students' alcohol consumption, sleep quality, and academic performance.

3.4 Evaluation

Clusters are evaluated using a similarity measure, such as the distance between cluster points. Here, Davies-Bouldin index score was used to validate clustering of the study. A model with a lower Davies-Bouldin index score indicates better clustering [12].

4. RESULTS AND DISCUSSIONS

K-means clustering algorithm was used to determine the clusters of students in students sleep pattern data. All the attributes used in the dataset was pre-processed to make it more usable for clustering student data. Below shows the pre-processed data of the study.

Table 1. Pre-processed Data.

	GPA	PoorSleepQuality	Drinks
1	0.881	-0.773	1.082
2	-0.009	-0.088	0.105
3	-0.677	4.022	-0.627
4	1.277	0.939	-0.872
5	-0.108	0.939	-0.383
6	0.634	-0.088	-1.360
7	0.263	-1.458	0.105
8	-0.603	1.282	-0.627
9	1.870	-0.430	-0.627
10	-0.850	-1.458	0.105
...	...	...	...
250	-0.603	0.939	-0.139
251	0.634	-1.458	-0.139
252	-1.592	-0.430	0.349
253	-1.840	-1.115	1.815

The tool used in this experiment was RapidMiner. First, the dataset is imported to the software. Second, the selection of attributes were applied. Here, the GPA, PoorSleepQuality, and Drinks were the selected attributes used in the study. Next, data normalization is implemented to the selected attributes to improve data integrity and ensures all the attributes behaves the same manner. After that, the implementation of the clustering model were performed. The data were clustered with a k value of 2. Below shows the cluster process of the study.

Figure 3. Design of Cluster Process with value K = 2.

4.1 A. Cluster Performance on alcohol consumption and students' academic performance

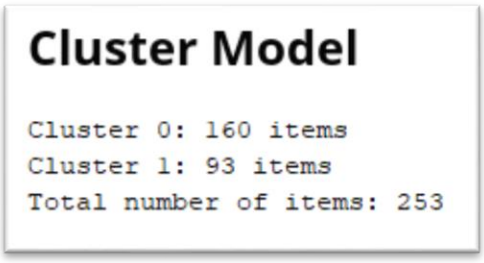

Figure 4. Clustering Results of Alcohol Consumption and Academic Performance.

Figure 4 shows the results of the clustering of students data by alcohol consumption and academic performance. Here, Cluster 0 contains 160 students and Cluster 1 contains 93 items. Cluster 0 produces the highest number of students clustered.

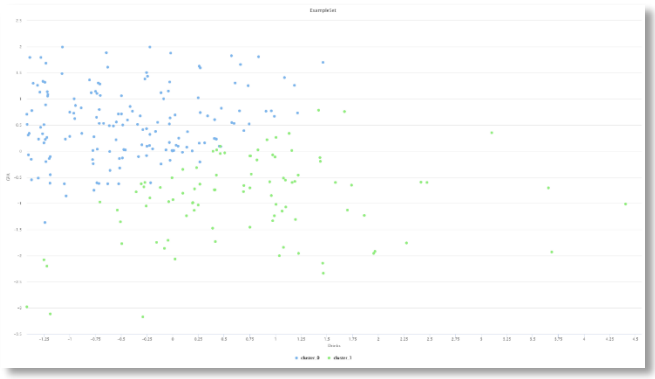

Figure 5. Results of K-means Algorithm on Alcohol Consumption and Academic Performance

In figure 5, the result of K-means clustering on alcohol consumption and academic performance was presented. Here, cluster 0 assigned by blue dot, indicates student who has high grades and low alcohol consumption. Meanwhile, cluster 1 denoted by green dot, indicates students who has lower grades compared to cluster 0 and has moderate and high alcohol consumption. The plot is formed in a downward slope pattern; in other words, alcohol consumption has a negative effects on students' GPA. Our findings are closely aligned with those of [38] and [39], who come to the same conclusion about the effects of alcohol consumption on students' academic performance.

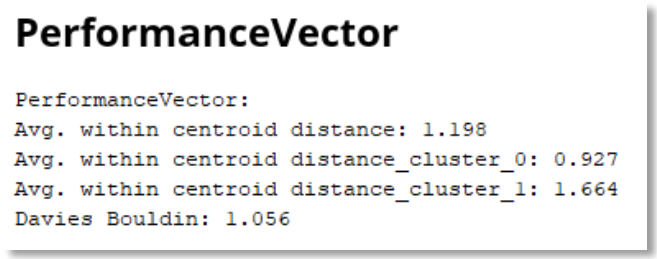

Figure 6. Cluster Performance Accuracy on Alcohol Consumption and Academic Performance.

Figure 6 shows the performance accuracy of the cluster model. The operator used to measure the K-means cluster accuracy is % Performance. It is used to evaluate the performance of centroid-based clustering techniques [40]. One of the assessment parameters of % Performance is Davies Bouldin Index. The result of Davies Bouldin Index produces a value of 1.056, based on the results of these performance, it can be summed up as a good cluster algorithm.

4.2 Cluster Performance on Sleep Quality and Students' Academic Performance

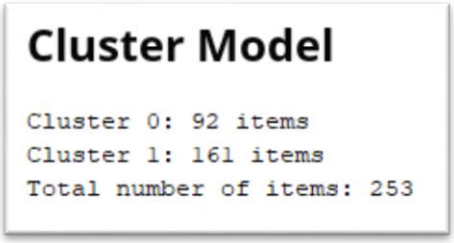

Figure 7. Clustering Results of Sleep Quality and Academic Performance.

In figure 7, the cluster results of sleep quality and academic performance was presented. The total participants were clustered into two groups: Cluster 0 and Cluster 1. Cluster 0 contains 92 students, while Cluster 1 contains 161 students possessing the highest number of students contained.

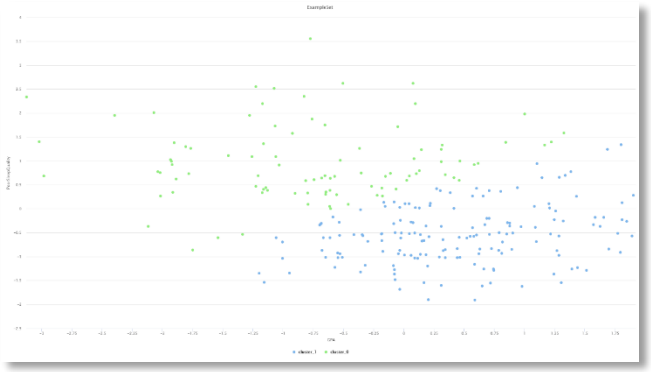

Figure 8. Results of K-means Algorithm on Sleep Quality and Academic Performance.

Figure 8 describes the result of K-means clustering on sleep quality and academic performance. Cluster 1 specifies students who has high GPA and rich sleep quality. On the other hand, cluster 0 specifies students who has somewhat low GPA compared to cluster 1 and has poorer sleep quality. Hence, the figure shown above clearly shows sleep quality has an effect on students' academic performance. The implication of these findings were also discussed in the study of [41] where sleep quality is associated to students' academic performance.

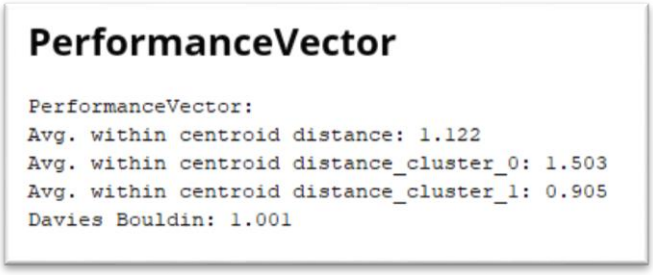

Figure 9. Cluster Performance Accuracy on Sleep Quality and Academic Performance.

Figure 9 presents the cluster performance accuracy of the model. Here, the Davies Bouldin Index display a score of 1.001. Based on the results, this model is considered to be a satisfactory algorithm.

5. SUMMARY

The aim of this study was to categorize students based on sleep quality, alcohol consumption, and academic performance. K-means clustering algorithm was used to determine the group of students based on the attributes assigned. Student's sleep pattern data obtained from R package called "Lock5Data" were used in this study. In this study, the tool used for experiment was RapidMiner. The dataset obtained were pre-processed using Z-transformation to minimize data redundancy and ensures data integrity. In RapidMiner, the "Clustering" operator was used to cluster the students into to two groups: Cluster 0 and Cluster 1. Next, Davies Bouldin Index was utilized to assess cluster performance. Better clustering is indicated by a model with a lower value. In this study, there were two clustering operations performed. One is to cluster students based on sleep quality and academic performance. Second is to categorize students based on alcohol consumption and academic performance. In clustering the students based on sleep quality and academic performance, Davies Bouldin Index displays a score of 1.001 while to the other displays a score of 1.056. Overall, both clustering operations generated a good clustering accuracy performance. Hence, sleep quality and alcohol consumption has an effect on students' academic performance.

6. CONCLUSION AND RECOMMENDATION

The researchers have been successful in using K-means to categorize students based on sleep quality, alcohol consumption, and academic performance. Based on the results, minority of the students are considered to be heavy drinkers and these students had poor academic performance. In other words, an increase of alcohol consumption negatively affects the students' academic performance. Another, majority of the students tend to have better sleep quality and academic performance while minority of the students have poor sleep quality and also has poor academic performance. Therefore, sleep quality has a difference between students with high academic performance and those with low academic performance. Overall, the method provides sufficient results which acknowledged the research questions. 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 [42]. All in all, these recommendations can be used for future research which can draw conclusion with a valuable findings.

7. REFERENCES

[1] 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

[2] El Ansari, W., Salam, A., & Suominen, S. (2020). Is Alcohol Consumption Associated with Poor Perceived Academic Performance? Survey of Undergraduates in Finland. *International journal of environmental research and public health*, *17*(4), 1369. https://doi.org/10.3390/ijerph17041369

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

[4] Jalali, R., Khazaei, H., Paveh, B. K., Hayrani, Z., & Menati, L. (2020). The Effect of Sleep Quality on Students' Academic Achievement. *Advances in medical education and practice*, *11*, 497–502. https://doi.org/10.2147/AMEP.S261525

[5] Onyper, S. V., Thacher, P. V., Gilbert, J. W., & Gradess, S. G. (2012). Class start times, sleep, and academic performance in college: a path analysis. Chronobiology International, 29(3), 318-335.

[6] Lovibond PF, Lovibond SH. (1995). The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav. Res.Ther. 33:335–343.

[7] Anunaya, S. (2021, August 10). Data Preprocessing in Data Mining -A Hands On Guide. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2021/08/data-preprocessing-in-data-mining-a-hands-on-guide/

[8] Lanio, K. (2022, April 11). Data Mining Tools. RapidMiner. https://rapidminer.com/glossary/data-mining-tools/

[9] Better Evaluation. (2014, June 12). RapidMiner. https://www.betterevaluation.org/en/resources/tool/rapidminer

[10] Techopedia. (2016, November 30). K-Means Clustering. Techopedia.Com. https://www.techopedia.com/definition/32057/k-means-clustering

[11] Lu, X. S., & Zhou, M. (2016, April). Analyzing the evolution of rare events via social media data and k-means clustering algorithm. In 2016 IEEE 13th International Conference on Networking, Sensing, and Control (ICNSC) (pp. 1-6). IEEE.

[12] Das, A. (2022, January 15). Deciding number of Clusters using Gap Statistics, Davies-Bouldin Index, & Calinski-Harabasz Index for K-Means and Hierarchical Clustering using Python. Medium. https://medium.com/mlearning-ai/deciding-number-of-clusters-using-gap-statistics-davies-bouldin-index-calinski-harabasz-index-2ce9acfb6118

[13] Virmani, D., Taneja, S., & Malhotra, G. (2015). Normalization based K means Clustering Algorithm. arXiv preprint arXiv:1503.00900.

[14] Wolfson, A. R., & Carskadon, M. A. (1998). Sleep schedules and daytime functioning in adolescents. Child development, 69(4), 875-887.

[15] Carskadon, M. A., Seifer, R., & Acebo, C. (1991). Reliability of six scales in a sleep questionnaire for adolescents. Sleep Research, 20, 421.

[16] Wechsler D. (1997). Wechsler Adult Intelligence Scale—Third Edition. San Antonio: The Psychological Corporation.

[17] Casagrande M, Violani C, Curcio G, Bertini M. (1997). Assessing vigilance through a brief pencil and paper letter cancellation task (LCT): effects of one night of sleep deprivation and of the time of day. Ergonomics 40:613–630.

[18] 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

[19] 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.

[20] 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

[21] 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

[22] 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

[23] 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

[24] Houlis, A. (2020, September 30). Reasons People Drink. Retrieved from Alcohol Rehab Help: https://alcoholrehabhelp.org/blog/why-people-drink/

[25] Tembo C, Burns S, Kalembo F (2017) The association between levels of alcohol consumption and mental health problems and academic performance among young university students. PLoS ONE 12(6): e0178142. https://doi.org/10.1371/journal.pone.0178142

[26] El Ansari, W., Salam, A., & Sakari, S. (2020). Is Alcohol Consumption Associated with Poor Perceived Academic Performance? Survey of Undergraduates in Finland. International Journal of Environmental Research and Public Health.

[27] Makongho, A.-M. (2018, March 18). The Effect of Alcohol Use on Academic Performance of College Students. Retrieved from https://minds.wisconsin.edu/: https://minds.wisconsin.edu/bitstream/handle/1793/78353/MakonghoAnne-Marie.pdf?sequence=3&isAllowed=y

[28] Kaushik, S. (2016, November 3). An Introduction to Clustering and different methods of clustering. Retrieved from

AnalyticsVidhya: https://www.analyticsvidhya.com/blog/2016/11/an-introduction-to-clustering-and-different-methods-of-clustering/

[29] Dimitriadis, Stavros & Messaritaki, Eirini & Jones, Derek. (2021). The Impact of Graph Construction Scheme and Community Detection Algorithm on the Reliability of Community and Hub Identification in Structural Brain Networks. 10.1101/2020.05.07.082271.

[30] Milecia, M. (2020, September 21). 8 Clustering Algorithms in Machine Learning that All Data Scientists Should Know. Retrieved from freecodecamp.org: https://www.freecodecamp.org/news/8-clustering-algorithms-in-machine-learning-that-all-data-scientists-should-know/

[31] Sharma, P. (2019, August 19). The Most Comprehensive Guide to K-Means Clustering You'll Ever Need. Retrieved from AnalyticsVidhya: https://www.analyticsvidhya.com/blog/2019/08/comprehensive-guide-k-means-clustering/#h2_9

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

[33] 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.

[34] Baker, R. (2010). Data Mining for Education. International Encyclopedia of Education (3rd edition), 1.

[35] 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

[36] 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

[37] 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

[38] Idoko, J. O., Muyiwa, A. S., & Agoha, B. E. (2015). The effect of alcohol consumption on the academic performance of undergraduate students. Psychology and Behavioral Sciences, 4(4), 147-153.

[39] Singleton, R. A., & Wolfson, A. R. (2009). Alcohol consumption, sleep, and academic performance among college students. Journal of studies on alcohol and drugs, 70(3), 355-363.

[40] Windarto, A. P., & Wanto, A. (2018, September). Data mining tools| rapidminer: K-means method on clustering of rice crops by province as efforts to stabilize food crops in Indonesia. In IOP Conference Series: Materials Science and Engineering (Vol. 420, No. 1, p. 012089). IOP Publishing.

[41] 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.

[42] 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.