Student's Performance Based on E-Learning Platform Behaviour using K-means Clustering

Abdella Mame Abdo[1], Nur Maisara Ahmad Rasid[2], Nur Amira Husna Mohd Badli[3], Siti Nur Aisha Sulaiman[4], Sharyar Wani[5], Zuraini Zainol[6]

[12345]Department of Computer Science, Kulliyyah of Information & Communication Technology, International Islamic University Malaysia, Kuala Lumpur, Malaysia
[6]Department of Computer Science, Faculty of Defence Science and Technology, National Defence University of Malaysia, Kuala Lumpur, Malaysia
abdellamame@gmail.com maisarasido3@gmail.com, amirahusna1113@gmail.com, 99sitinuraishah@gmail.com, sharyarwani@iium.edu.my, zurainiz@upnm.edu.my

***Abstract*—** Measuring student engagement on e-learning platforms is critical to student learning because online platforms can make some students feel isolated and disconnected, leading to loss of interest and affecting academic performance. This paper explores the relation between student's responses on e-learning platforms and performance of various skill levels. Using K-Means clustering technique the students are grouped by their total responses and relative total skill levels to determine cluster having various total response rate and total skill levels in order to measure student performance based on their engagement. This will also assist instructors to focus on students/clusters that require more attention in the teaching and learning process. The clustering results indicate students with higher engagement tend to perform better compared to the cluster with moderate and lower engagement patterns.

***Keywords*—** Student Performance, E-Learning, Platform, K-means Clustering, Online Engagement

I. Introduction

The rapid progress of technology in the modern world, has made online learning as one of the main options in many institutions, especially during Covid-19. Online learning, or e-learning, offers students all the advantages of attending a physical class, with the added advantage of having a flexible schedule. Recently students have increasingly turned to online learning as a viable alternative to on-campus study since courses are available in almost every major. This gave the chance for many students to attend their dream university remotely from the comfort of their home [1].

Students must obtain the ability to articulate ideas, develop understanding and engage with others through spoken language. The interactive learning environment that comes with e-learning provides students with the ability to gain skills such as self-direction and critical thinking. Educators have long argued that the semi-autonomous and self-directed nature of the virtual classroom boosts innovative and creative skills in education. In the online environment, the facilitator and student collaborate to create a dynamic learning experience [2], [3].

A student's behaviour in a class is a measure of various factors. In the current pandemic, online education has been majorly the source of teaching and learning. Online learning is quite different compared to face-to-face learning. Students are having difficulty coping with online learning because of the nature and environment of learning. Additionally, they have other duties to fulfil while being home.

During face-to-face learning, instructors can usually assess their students directly and make sure that the students stay focused in class. However, in online learning, instructors can only assess the students virtually. It does not necessitate that the students can really understand or concentrate in the class. It has been observed that both the parties are having a hard time adapting to the new environment. This situation is quite concerning especially to the students as this impacts their academic performance. Compared to physical learning, online learning requires a higher level of self-discipline and motivation to study and perform well in class. During this time, the e-learning platform is the only medium for students and instructors to communicate [4].

Thus, universities have started tracking the progress of their students in gaining mental skills such as critical thinking, creativity, collaboration and other skills that have a direct impact on their grades in the short term and their professional life in the long term [5]. At the same time, institutions measure and monitor the effectiveness of online platforms for the necessary skill building strategy.

This study aims to study the student's behaviour while using e-learning platform by understanding the relationship between a student's response on the online platforms and their performance in class. Using clustering techniques, the study intends to classify groups that require more engagement and attention in the teaching and learning process.

II. Related Work

Student's behaviour can be analysed by using various features such as managing projects, time management, etc. Using K-Means Clustering, the authors in [6] present that most of the data lies in the agree cluster, indicating student agreement to project-based learning models for satisfaction when the project is complete and is consider it an innovative learning model. It helps to practice project management skills as well as challenges traditional learning skills. Therefore, it is considered suitable to be applied in the lecture design and strategy courses. Similarly in [7], researchers explore variables hidden hidden curriculum and character building among others for self-motivation in college students using K-Means Clustering. The two aforementioned factors were found to have a profound effect of about 60.53%. Meanwhile researchers in [8] propose an improved K-Means Clustering technique for identifying student groups with varied behaviour in a big data setting. The results indicate better performance in grouping and computational performance compared to the traditional K-Means clustering.

Reference [9] argues the need to evaluate examine and assess students results to measure effectiveness of current educational systems in centres of learning. They apply clustering and other statistical tests on the result of postgraduate students in the computer majors. They successfully identify the clusters of good performance and report presence of more females in these clusters compared to male students.

Motivation and engagement are crucial factors for e-learning platforms. In [10], the authors focus on twelve engagement variables divided into two categories of interaction and effort to tackle issues related to motivation and. Results using k-means algorithm reveal number of logins and assignment submit duration highly representative of the aforementioned issues.

The era of big data has brought forward different aspects of data to be dealt for efficient interpretations. As such clustering algorithms have gained momentum in this direction. Different tuned clustering algorithms have been presented in recent times to cover multiple dimensionalities of data in recent time. A fusion k-means algorithm has been proposed specifically in [11] to study student behaviour, which is close to the scope of the current work.

The famous VAK model emphasizes on visual, kinaesthetic and auditory as three major styles of learning. Most participants require more than one of these for effective learning. Therefore, it becomes crucial to identify the style of learning for individuals in a learning setting so that appropriate materials can be provided. K-means, decision trees and SVM have been used in [12] for prediction of learning style combination and suggest field of studies accordingly. SVM successfully classifies the combinations at an accuracy of about 92%.

III. Experimental Setup

A. Dataset

The dataset used in this study is "E-Learning Student Reactions: Students Reactions and Posts from E-Learning " publicly available and obtained from Kaggle[13]. The dataset was compiled after four months of an algorithm introductory class at a Brazilian University. It consists 16 variables with 71 instances, all collected as quantitative data. Each row is a measurement of a unique student, type of reactions from other students and skill grading by the instructor.

B. Tools & Techniques

The research mostly uses the data science stack libraries such as sicikit-learn, matplotlib, seaborn, NumPy and pandas among others for analysis.

K-Means clustering has been used where the data is clustered into groups based on the total number of reactions by the students and their total grade of skills. Clustering analysis can be performed on the basis of features where we try to determine the subgroups of samples based on similar characteristics. The K-Means algorithm is one of the well-known unsupervised machine learning and clustering algorithms. Clustering is considered as unsupervised since there is no label to compare with the output. In particular, K-Means works by specifying the number of clusters, k, the centroid is initialized by selecting a random k data point for each cluster and keeps iterating until there is no change to the centroids [14], [15].

During the implementation of the K-Means algorithm, Elbow method was used to determine the optimal k. The Elbow method works by calculating the sum of squared error (SSE) from each of the k. Based on the Elbow method, the number of clusters used are 3. The SSE is calculated using the square of the distance between the data point and cluster centre [16].

C. Implementation

Post pre-processing such as treatment of missing values, data inconsistencies, etc, the entire dataset was used for training. Using the Elbow method, the clustering algorithm was initialized with three clusters. The clusters were based on students, their total reactions and the total grade of their skills. The initial clustering wasn't deemed satisfactory. Therefore, MinMaxScaler was used for normalizing the feature scale for better clustering. The scaling helped to obtain a better clustering visualization. The results have been discussed in the forthcoming sections.

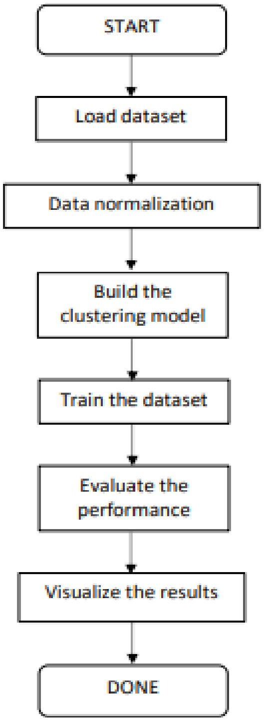

Fig. 1 Modeling Overview

IV. Results & Discussion

This section presents the data findings highlighting the relationship between the three essential features – student behaviour, peer reaction and skill grades.

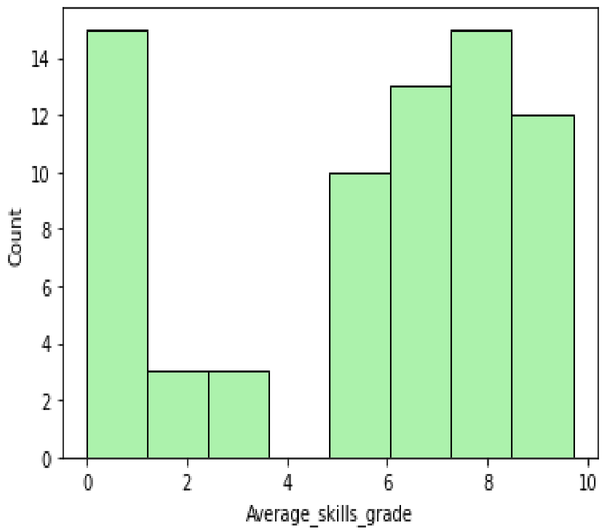

Fig. 2 Average grade for skills

The histogram in Fig 2 reveals average skills grade range from 0 to 9. Most of the students in the class have an average skills grade of 0 to 1 and 8.

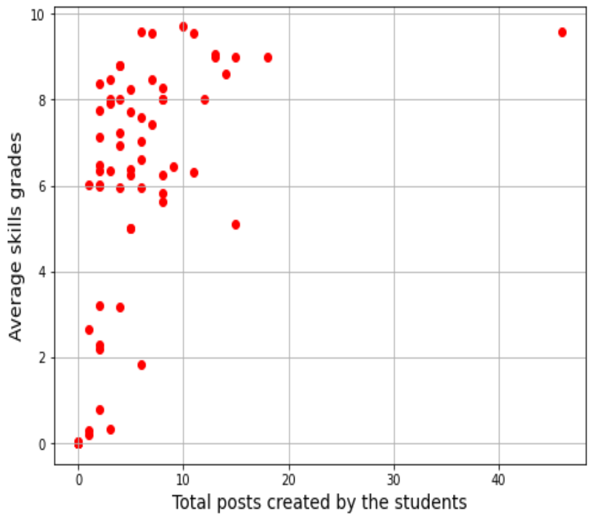

Fig. 3 Students posts against average skills grade

The scatterplot in Fig 3 presents student' final grade with the number of total posts created by students in the online class. The scatter plot is moderate, positive and linear. The uphill pattern proves that the total posts created by the students convey good performance of students. Therefore, higher the posts created by the students, the higher their average skills grade in class.

The heatmap in Fig 4 demonstrates the relationship between the reactions and the type of skills. Critical thinking and problem-solving skills, creativity and innovation skills, and, constant and self-learning skills are associated with the reaction of collaborative post. For collaboration and self-direction skills, and, social and cultural responsibility, the students' reaction is correlated to the amazing post reaction. In a nutshell, students that

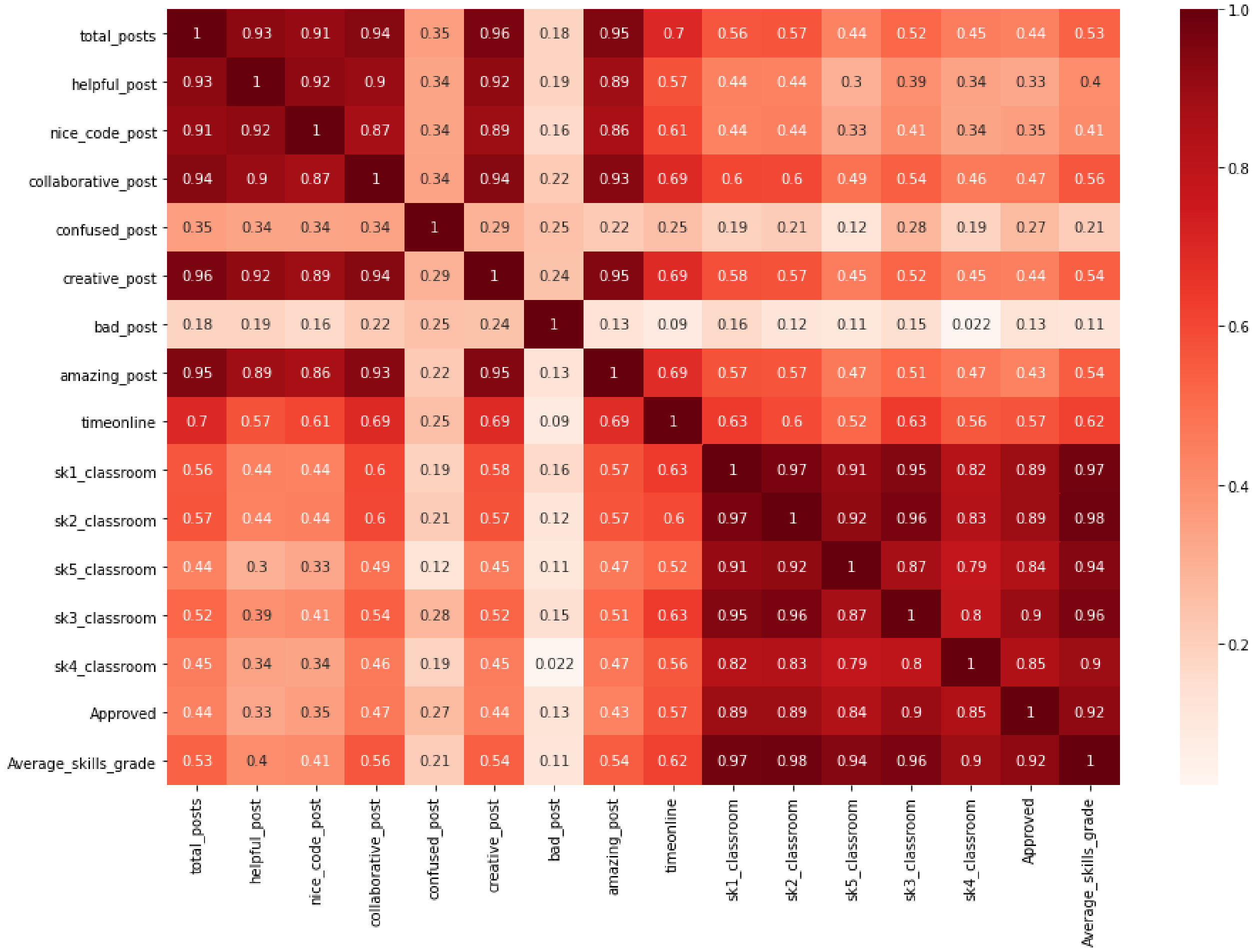

Fig. 4 Reactions and skills relationship

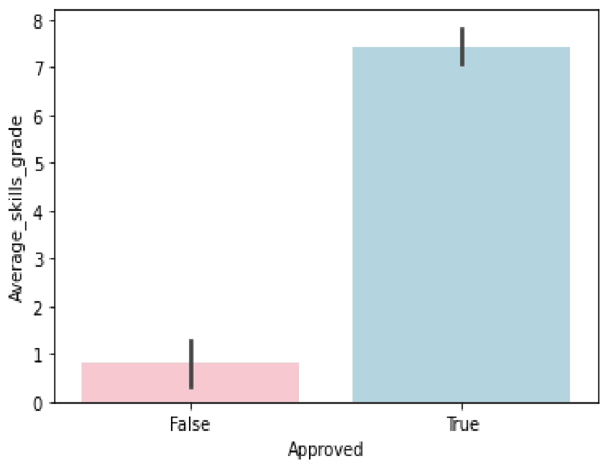

Fig. 5: Student average and instructor approval

Fig 5 visualizes the student average grade in class against the student's approval in class. It can be observed that most of the students that are approved in the class have a high average grade. It seems that there is a tendency of better approval in class if one scores high marks.

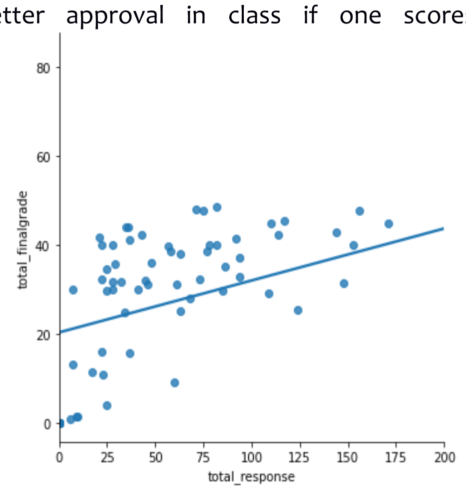

Fig. 6 Total response against total grade of skills

Fig 6 and Fig 7 present total response and time spent online with the total final grade for the students. There is a rising trend between values of the variables. Based on this trend, it can be concluded that total response and time spent online contribute to better student performance. The students who are active in class tend score higher grades.

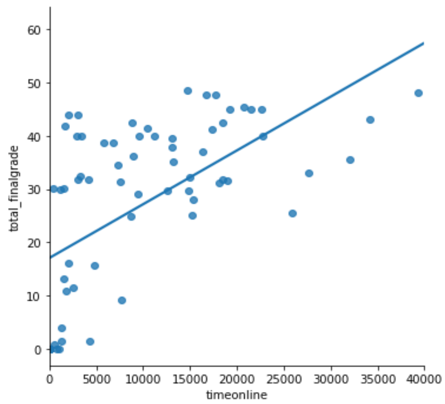

Fig. 7: Time spent online against total grade of skills

Fig 8 presents the total response by the students against the score of their skills. From the scatterplot it can be inferred that despite low total response by some students, they still managed to score high for certain skills compared to others that gave a huge amount of total response, but some of their skills grade is lower than those that provided less total response.

Fig 9 displays the total number of reactions by the student against the total grade of their skills. It is generally observed that higher the total response, higher the total grade of all skills. However, some students are able to score well despite low number of total responses.

The Elbow method is visualized in the plot presented in Fig 10 that we applied to determine the optimal k. The most elbow-shaped form from the plot above falls at the scale of 3, therefore we have 3 clusters.

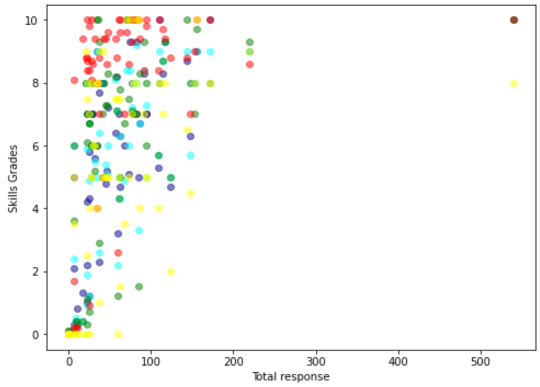

Fig. 8 Grade of skill based on total response

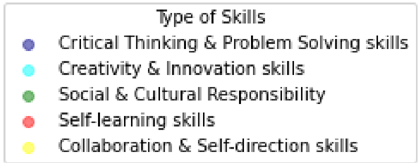

Fig. 9 Colour Code for Fig 8

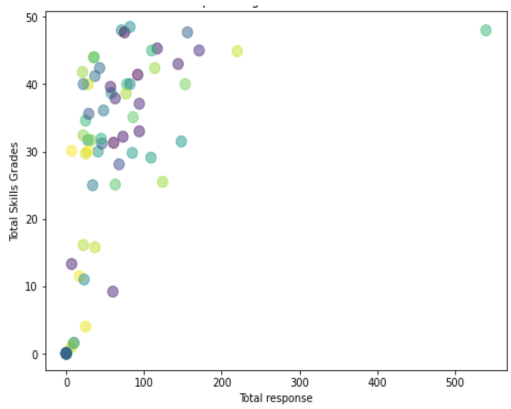

Fig. 9 Total response versus total skills grade10

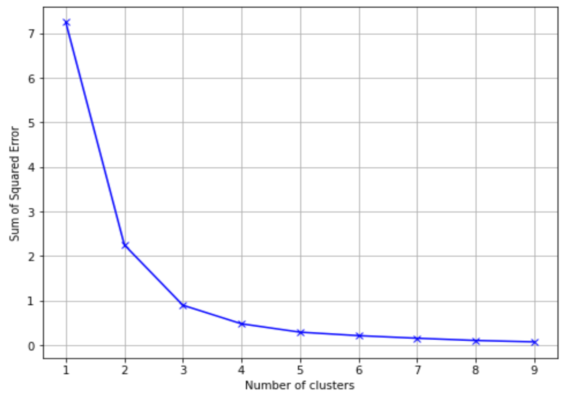

Fig. 10 Elbow method for optimal k

The scatter plot in Fig 11 visualizes the effect of K-Means algorithm on the total response and total skills grade. Based on the Elbow method, there should have been 3 clusters. However, without optimizing and normalizing the clustering model, the result indicates that most of the students in cluster 0 and cluster 2. There is a single data point for cluster 1, considerably isolated from the other two clusters. Therefore, post normalization using the MinMaxScaler, the three clusters can be observed distinctly in Fig 12.

The clusters are labelled as low response, medium response, and high response where the low response corresponds to cluster 1 (green), medium response corresponds to cluster 2 (red) and the high response represents cluster 3 (orange).

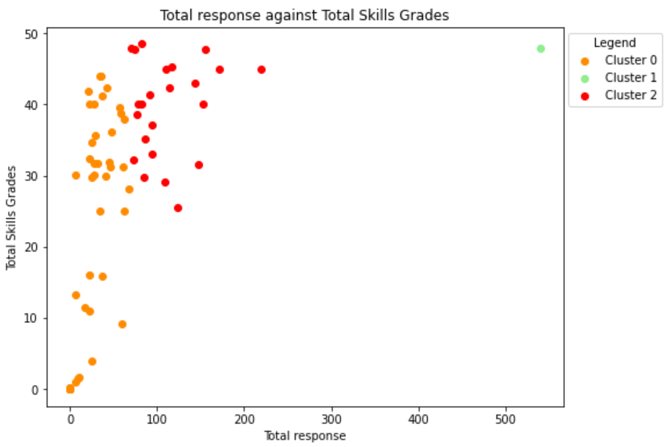

Fig. 11 K-Means clustering

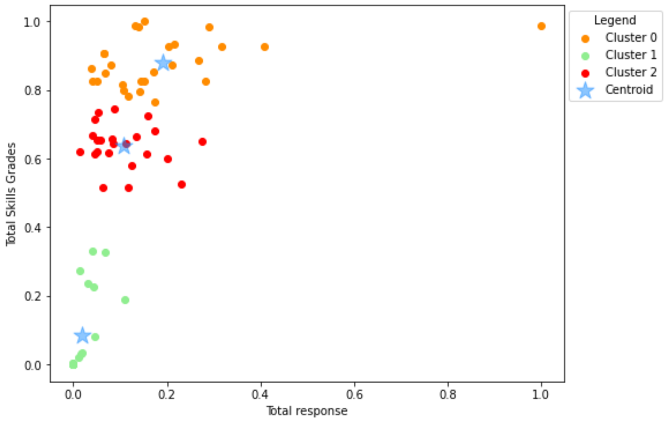

Fig. 12 Normalized K-Means clustering

From these 3 clusters, cluster 0 consists of students that often provide responses on the e-learning platform and managed to score higher total grade for skills. Cluster 1 consist of students that rarely react and are observed to score lower total grade for skills while cluster 2 consists of students presenting a middle ground in giving reactions and managed to score medium total score for their skills. The centroid for each cluster has also been presented in Fig 12. The three clusters are distinct and therefore the k-value of 3 successfully charts and predicts performance based on the underlying features in the data.

V. Conclusion

The paper investigated the relationship of student engagement on e-learning platforms and their performance. A well engaged student indicates higher interest to the topic of discussion. Engagement goes past the passive learning where the student tends to contribute their analysis and information to the domain of study. This enables better comprehension as the topic of discussion is not merely sought from the lecture material and also tends to boost chances of learning even more advanced material especially in depth. Such measurements and student clustering is essential and constantly required in both online and offline mode of teaching and learning. This will help the instructors to focus on groups with lower engagement levels determining the cause and adjust the class and materials accordingly.

Such studies need to be replicated with dataset with larger instances in order to further generalize the results. The data should include more features such as attendance, moral behaviour and others for deeper insights. Furthermore, varied clustering algorithms should be used to visualize an overall picture of the problem behaviour.

References

[1] B.-V. Cioruța, M. Lauran, M. Coman, A. L. Pop, and A. Lauran, “About the Benefits of Adopting E-Learning in the Current Romanian Educational System,” *Asian J. Educ. Soc. Stud.*, vol. 15, no. 3, pp. 1–13, Mar. 2021, doi: 10.9734/ajess/2021/v15i330379.

[2] M. C. Félix Antonio, “Teaching Strategies Used To Promote EFL Autonomous Learning In Distance Education Undergraduate Students: An Initial Approach In The Framework Of The Colombian Research Context,” Oct. 2020. https://core.ac.uk/download/pdf/344725158.pdf (accessed Jul. 07, 2021).

[3] E. Mani and A. Tachie-Menson, “An interdisciplinary approach to Medical Education: the role of Visual Media in teaching and learning of ‘Gross Human Anatomy’ at The University For Development Studies, Tamale, Ghana,” May 2021, Accessed: Jul. 07, 2021. [Online]. Available: http://ir.knust.edu.gh:8080/handle/123456789/13799.

[4] E. Hussein, S. Daoud, H. Alrabaiah, and R. Badawi, “Exploring undergraduate students’ attitudes towards emergency online learning during COVID-19: A case from the UAE,” *Child. Youth Serv. Rev.*, vol. 119, p. 105699, Dec. 2020, doi: 10.1016/j.childyouth.2020.105699.

[5] B. Yusuf, L. M. Walters, and S. N. Sailin, “Restructuring Educational Institutions for Growth in the Fourth Industrial Revolution (4IR): A Systematic Review,” *Int. J. Emerg. Technol. Learn.*, vol. 15, no. 03, pp. 93–109, Feb. 2020, doi: 10.3991/ijet.v15i03.11849.

[6] D. Kuswandi, E. Surahman, Z. Z. A. Thaariq, and M. Muthmainnah, “K-Means Clustering of Student Perceptions on Project-Based Learning Model Application,” in *2018 4th International Conference on Education and Technology, ICET 2018*, Jul. 2018, pp. 9–12, doi: 10.1109/ICEAT.2018.8693932.

[7] I. Gunawan *et al.*, "Hidden Curriculum and Character Building on Self-Motivation based on K-means Clustering," in *2018 4th International Conference on Education and Technology, ICET 2018*, Jul. 2018, pp. 32–35, doi: 10.1109/ICEAT.2018.8693931.

[8] D. Ding, J. Li, H. Wang, and Z. Liang, "Student Behavior Clustering Method Based on Campus Big Data," in *Proceedings - 13th International Conference on Computational Intelligence and Security, CIS 2017*, Feb. 2018, vol. 2018-January, pp. 500–503, doi: 10.1109/CIS.2017.00116.

[9] D. Aggarwal and D. Sharma, "Application of Clustering for Student Result Analysis," *Int. J. Recent Technol. Eng. ISSN*, pp. 2277–3878, 2019.

[10] A. Moubayed, M. Injadat, A. Shami, and H. Lutfiyya, "Student Engagement Level in an e-Learning Environment: Clustering Using K-means," *Am. J. Distance Educ.*, vol. 34, no. 2, pp. 137–156, Apr. 2020, doi: 10.1080/08923647.2020.1696140.

[11] W. Chang *et al.*, "Analysis of university students' behavior based on a fusion K-means clustering algorithm," *Appl. Sci.*, vol. 10, no. 18, p. 6566, Sep. 2020, doi: 10.3390/APP10186566.

[12] A. S. Kuttattu, G. S. Gokul, H. Prasad, J. Murali, and L. S. Nair, "Analysing the learning style of an individual and suggesting field of study using Machine Learning techniques," in *Proceedings of the 4th International Conference on Communication and Electronics Systems, ICCES 2019*, Jul. 2019, pp. 1671–1675, doi: 10.1109/ICCES45898.2019.9002051.

[13] "E-Learning Student Reactions | Kaggle." https://www.kaggle.com/marlonferrari/elearning-student-reactions (accessed Jun. 23, 2021).

[14] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," *Electronics (Switzerland)*, vol. 9, no. 8. MDPI AG, pp. 1–12, Aug. 01, 2020, doi: 10.3390/electronics9081295.

[15] "Meta-Heuristic Algorithms For K-Means Clustering: A Review," *PalArch's J. Archaeol. Egypt / Egyptol.*, vol. 17, no. 7, pp. 12002–12020, 2021, [Online]. Available: https://archives.palarch.nl/index.php/jae/article/view/4630.

[16] R. Nainggolan, R. Perangin-Angin, E. Simarmata, and A. F. Tarigan, "Improved the Performance of the K-Means Cluster Using the Sum of Squared Error (SSE) optimized by using the Elbow Method," in *Journal of Physics: Conference Series*, Dec. 2019, vol. 1361, no. 1, p. 12015, doi: 10.1088/1742-6596/1361/1/012015.