Research Article

Sage Science Review of Educational Technology (SSRET)
3(1), pp. 1–13

2020

JEL: G20: General
G29: Other

Clustering Students Based on Virtual Learning Engagement, Digital Skills, and E-learning Infrastructure: Applications of K-means, DBSCAN, Hierarchical, and Affinity Propagation Clustering

Sandeep Trivedi

IEEE Member, Graduated from Technocrats Institute of Technology ,
contact email ID : sandeep.trived.ieee@gmail.com

Nikhil Patel

IEEE Member
Graduated from the University of Dubuque, USA,
contact email ID: Patelnikhilr88@gmail.com

Abstract

Clustering algorithms allow schools and teachers to provide students with more tailored education services. Educators are able to determine the relevant educational material to send, choose the ideal education channels for the target students, uncover new and valuable insights, and launch new instruction techniques by acquiring a better knowledge of learner characteristics. This research segmented the data for 200 students of 6 different high schools using K-means, DBSCAN, Hierarchical, and Affinity Propagation clustering algorithms. The students' segments were determined based on their Virtual Learning Engagement, Digital Skills, and E-learning Infrastructure, which are presumably the most important characteristics to use when establishing the segments of the students attending virtual classes. This study highlights and recommend that different machine learning approaches should be implemented in order to develop segmented and personalized instruction strategies and class policies in order to enhance the effectiveness of the online classes and the level of student success.

Keywords: *Clustering, Digital skill, E-learning, Students, Virtual learning engagement*

Article type:	research article
Keywords:	*Clustering, Digital skill, E-learning, Students, Virtual learning engagement.*
JEL codes:	G20: General; G29: Other

Received: 16 October 2019 Revised: 28 December 2019 Accepted: 20 Jan 2020

Suggested citation:

Trivedi, S., Patel, N. (2020) 'Clustering Students Based on Virtual Learning Engagement, Digital Skills, and E-learning Infrastructure: Applications of K-means, DBSCAN, Hierarchical, and Affinity Propagation Clustering, *SSRET*, 1(1), pp. 1–13.

1. Introduction

Students acquire knowledge in distinct ways and at different rates. This is the notion around which the modern type of instruction known as "personalized learning" is built[1]. Every student has what's called a "learning plan" developed for them that takes into account the way they study, what they already know, as well as their abilities and interests. It is the contrast of the "one size fits all" strategy that is followed by the majority of current educational institutions [2].

The students and their instructors collaborate to determine both the short-term and the long-term objectives for their learning [3], [4]. Students are better able to take responsibility for their own learning with this "personalized learning" approach [3].

The goal of personalized learning would be to facilitate the realization of each student's full academic potential by tailoring their educational experiences to best meet their individual requirements. Its ultimate objective is to assure the academic success of each and every student. And it accomplishes that objective by developing one-of-a-kind teaching strategies that are geared at both challenging and motivating each particular learner.

Learning in the classroom has often been approached by schools using a model that is intended to be universally applicable. It is typical practice for a single educator to educate to the average in situations where they are responsible for the instruction of a large number of students. This means that they tailor their teachings to the typical pupil in the class. This might cause frustration for slower learners as they fall more and further behind. Also, quicker students run the risk of becoming bored and finally disinterested in their studies.

However, the emergence of new technologies has at last made it possible to overcome the limitations posed by the fact that only one instructor is responsible for supervising such a large number of pupils. Students won't be able to take responsibility of their own education unless they have the drive to do so, and personalized learning is designed to provide them that [5], [6].

Teachers in classrooms that have implemented personalized learning invest substantially less time standing at the front of the room presenting lectures to pupils. This frees up the teacher to spend more time working individually with students [7]. The students, on the other hand, are given the flexibility to pursue their own specific lesson plans and to study in a manner that is best suited to them personally.

However, this does not imply that classrooms are full with students who are studying on their own in complete quiet; rather, personalized learning is intended to be very interactive. Students who have the same interests or approaches to learning often collaborate on group work, which provides them with the opportunity to develop important skills in communication, teamwork, and leadership.

The role of educators is to give pupils with direction and monitor their development as learners. The time that they would have typically spent giving a lecture to the class has been released, which enables them to concentrate on assisting the individual development of each student rather than having to lecture.

2. Personalized Learning

Students and teachers alike can reap the benefits of adopting a personalized learning strategy in the classroom. This strategy allows students to continue pursuing their preferences at their own pace while receiving individualized learning experiences, and it frees up teachers to concentrate on providing guidance, support, and longer-term planning. The following is a list of the primary advantages that may result from implementing personalized learning within the educational setting:

1. Personalized learning makes it possible for students to excel in their academic subjects. When students are given the opportunity to study at their own speed and in a manner that is most suited to their preferred mode of learning, they get the confidence necessary to become proficient in every topic that is covered in the curriculum [8]. In a standard school setting, teachers are required to provide material at a certain speed, which may cause some pupils to fall behind and may bore others who are able to grasp concepts more quickly. Through the use of differentiated instruction, each student goes through the classes at a rate that is appropriate for them, enabling them to effectively absorb the content and eventually become knowledgeable in the topic [9].

2. Personalized learning provides Students with a sense of responsibility [3]. Students have more responsibility over their studies and are better able to plan for their own careers when they participate in personalized learning, which aids to keep them more interested in the classroom. It inspires children to discover and pursue their interests while at school and to learn in a variety of various ways via its encouragement. Students are not only able to acquire information from their teachers and via self-directed study, but they are also encouraged to collaborate closely with their classmates in order to share their abilities, expertise, and mutual interests [10]. Students are given the ability to take charge of their own education when they participate actively in the process of determining what and how they study in the classroom via personalized learning.

3. Personalized learning is Beneficial to the Development of Soft Skills [11]. Learning that is personalized imparts essential abilities beyond the scope of conventional topics such as mathematics, English, and history. In addition to this, it places an emphasis on the cultivation and refinement of soft skills such as empathy, creativity, communication, and teamwork. As children prepare for a future that may not be predictable, these abilities are essential. Despite the fact that we do not understand how technology will affect the employment market in the following decades, this prediction was made. And in a

world where automation is increasing, the soft talents that students possess would still be in demand. Students are better prepared for the uncertain labor market of the future because to personalized learning, which provides them with transferrable skills.

4. The feedback given by students is Improved with personalized learning. The integration of various forms of technology into educational settings is a common component of personalized learning. Through the use of personalized tests and quizzes designed specifically for each student, technology also enables instructors to provide students with feedback in a more timely and frequent manner. Students, parents, and teachers can all receive a clearer picture of how effectively a kid is assimilating new information, and they can fix any information gap in real time [12].

5. Personalized learning motivates more collaboration, which often results in better collaboration [13]. Learning that is personalized includes elements of collaboration. Students and teachers collaborate to determine learning plans and objectives in a way that satisfies both the standards of the educational course and the needs of the individual student. In addition, students are strongly encouraged to collaborate closely with their classmates on group projects and to share their knowledge in a casual setting with one another. This helps students learn alongside one another and fosters a sense of community [14].

6. Personalized learning gives teachers the opportunity to optimize the utilization of their time [15]. Because it makes use of technology, personalized learning may save teachers time that would otherwise be spent on administrative activities, allowing them to devote more of their attention to assisting students with their own learning and resolving problems as they occur. The amount of time spent marking papers may be reduced by using online learning applications. Less time is spent speaking in the classroom when using personalized learning, which frees up the instructor to spend more time working one-on-one with each student. Parents are often given the opportunity to get insight into their children's academic performance via the use of technology in the classroom. This may include the ability to learn about forthcoming tasks or examine current grades. In this manner, parents are able to remain current, and teachers are spared the strain of additional work [16].

7. As a result of personalized learning, one's academic performance will improve [17]. Increasing the number of students who are successful in their studies is perhaps the advantage of introducing customized learning within the classroom that stands out the most. Students are able to study at their own speed, and they are encouraged to explore their interests thanks to personalized learning, which also develops closer ties between students and their teachers. When schools make the transition from conventional learning to personalized learning, the results of classroom performance frequently show a discernible upward trend [18]. Students who started considerably below national standards for standardized test scores were able to perform above national standards within two years, according to new research conducted by the Rand Corporation [19], [20]. This finding was based on the fact that schools moved to personalized learning programs.

3. Methodology

The process of categorizing students into clusters that are similar with regard to observable qualities is the objective of the machine learning technique known as cluster analysis [21].

In this study, we used the method of cluster analysis. The process of classifying things (such as objects, animals, and people, among other things) into clusters that are similar across a variety of observable characteristics is referred to as cluster analysis [22]. As soon as such homogeneous groups have been formed, the researcher will be able to focus his or her attention on a limited number of classes rather than the large number of original items [22]. Instead of dealing with data for a very large number of different things, the researcher focuses on data for a very small number of different groups that all have the same things in common [23]. As a reflection of this, cluster analysis is frequently considered to be an exploratory method of data analysis, the purpose of which is to generate ideas rather than to validate existing ones. In the field of education research [24], cluster analysis can be applied for purposes more than merely data exploration. In education, clustering methods are used in order to uncover and define students' subgroups, which are the central focus of teaching strategy. This is accomplished via the use of "clustering" Cluster analysis has for a very long time been the method that is most often used and recommended. Clustering is a generic phrase that may be used to refer to a variety of approaches that seek to discover clusters, as well as groupings that have internal cohesion and groups that are isolated from one another externally [25].

The first task is to establish k, which stands for the total number of clusters that will be produced. After that, the centers of the clusters are selected randomly from a collection of k different places. After calculating the proximity between each centroid and each observation, the observations are then assigned to the centroids that is spatially situated the closest to them [26]. The data are then grouped together at the same time into k distinct clusters. After that, we go on to the next step, which is computing unique centers for every group. After that, the distances between every data point and the new centroids are determined, and the data are reallocated to the centroids that is located the closest to them based on those distances. After then, new centers are determined for every cluster, and this process is repeated until the clusters remain unchanged.

The within-cluster sum of squares is a measure that is used to measure the performance of the algorithm. It is also sometimes referred to as the inertia [27]. We will refer to this distance as di, and we will define it as the separation measure between the middle point and the ith observation. Then [28]:

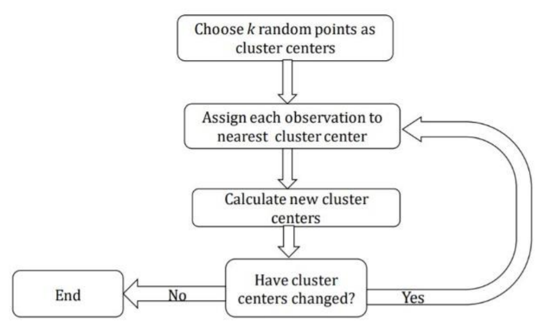

Figure 1 The k-means algorithm

$$Inertia = WCSS = \sum_{i=1}^{n} d_i^2$$

The number of observations is represented by the letter n, which stands for "n." The reduction of inertia should be the objective of the k-means method no matter what value of k is being used. There is a possibility that the initial centroids that are selected will have an effect on the result of only one iteration of the algorithm. As a direct result, the procedure has to be carried out several times using a variety of beginning cluster centers. The result with the fewest amounts of inertia is going to be the most favorable throughout all of the runs [29], [30].

The Density-based Spatial clustering of applications with noise (DBSCAN) has been hailed as one of the most efficient and often mentioned algorithms for identifying clusters of random form and size in big datasets that have been contaminated by noise [31], [32]. The DBSCAN method has the benefit of not requiring datasets to be predetermined in terms of the quantity of clusters [33].

Another unsupervised learning approach is hierarchical clustering, which is used to group together unlabeled data points with comparable features. Hierarchical Clustering comes in two versions. 1. Divisive, and 2. Hierarchical Agglomerative Clustering Agglomerative [34]. Hierarchical Clustering is generally recognized as a bottom-up technique, whereby each data or observation is considered as its cluster. Clusters are consolidated until all data is included in a single massive cluster.

Hierarchical clustering begins with k = N clusters and progresses to k = N-1 clusters by combining the two nearest days into one cluster. The procedure of combining two clusters to produce k-1 clusters is continued until the required number of segments K is reached. To determine which clusters to merge, we utilize the Euclidean distance [35]. The centroid then represents the final cluster allocations. Hierarchical clustering is generally deterministic, which implies that it can be replicated. It is, however, results in local solutions [36], [37].

In a broad sense, the inertia / WCSS value will decrease as k increases. In the case where k is equal to the number of observations and the inertia is zero, there will be one cluster for each observation. This will occur when there is no inertia. The elbow method is a standard method that is used to determine the number of clusters that are present [38]. A method that is less open to interpretation is the silhouette technique, which may be used to count the size of clusters. Frey and Dueck introduced Affinity Propagation in 2007 [39], and it has grown in popularity owing to its simplicity, wide application, and performance. Affinity Propagation finds exemplars among data sets and generates clusters of datasets around them. It works by treating all data points as prospective exemplars at the same time and communicating between them until a good collection of exemplars and clusters develops [40]. Affinity Propagation does not require a number of clusters to be specified and offers benefits in terms of accuracy and

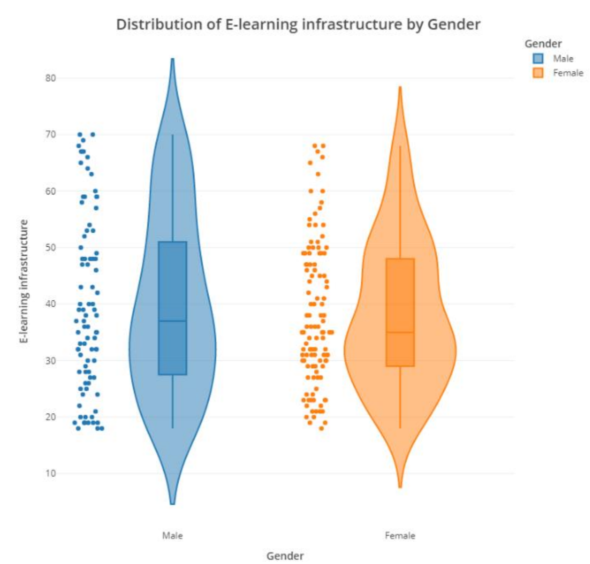

Figure 2. Distributions E-learning infrastructure by gender

efficiency, but it is not ideal for large-scale clustering [41]–[43]. The dataset included information from 200 students who were taking online classes.

4. Results

The median E-learning infrastructure score of the male students in the dataset is 37, while the median E-learning infrastructure

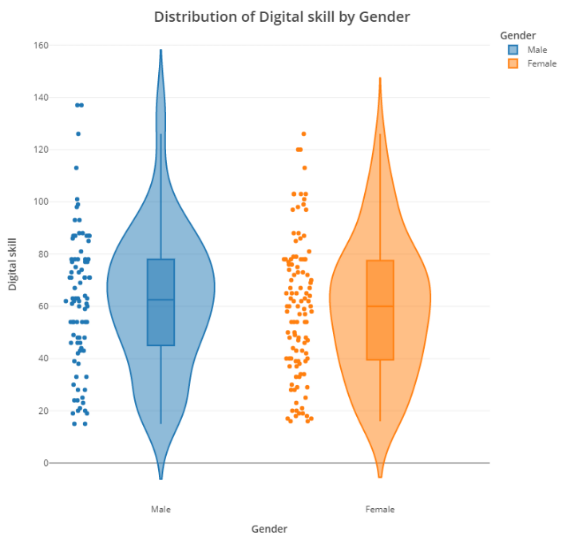

Figure 3. Distributions of digital skills by gender

score for female students in the dataset is 35. The median Digital competence score of male students in the dataset is 62.5, whereas the median Digital skill score for the female

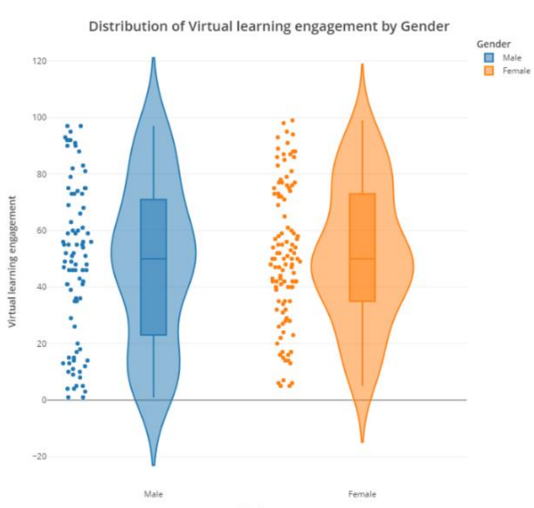

Figure 4. Distributions virtual engagement by gender

students is 60. In terms of Virtual learning engagements, both male students and female students have the same median value of 50. The *maximum numbers* nevertheless reveal the greatest Virtual learning engagement for male students to be 97, whilst that of female students to be 99.

We perform an assessment of the total number of virtual learning interactions that male students and female students had with regard to each category of digital skills. It can be shown that, across the board for digital skill groups, the total number of virtual learning engagement that female students

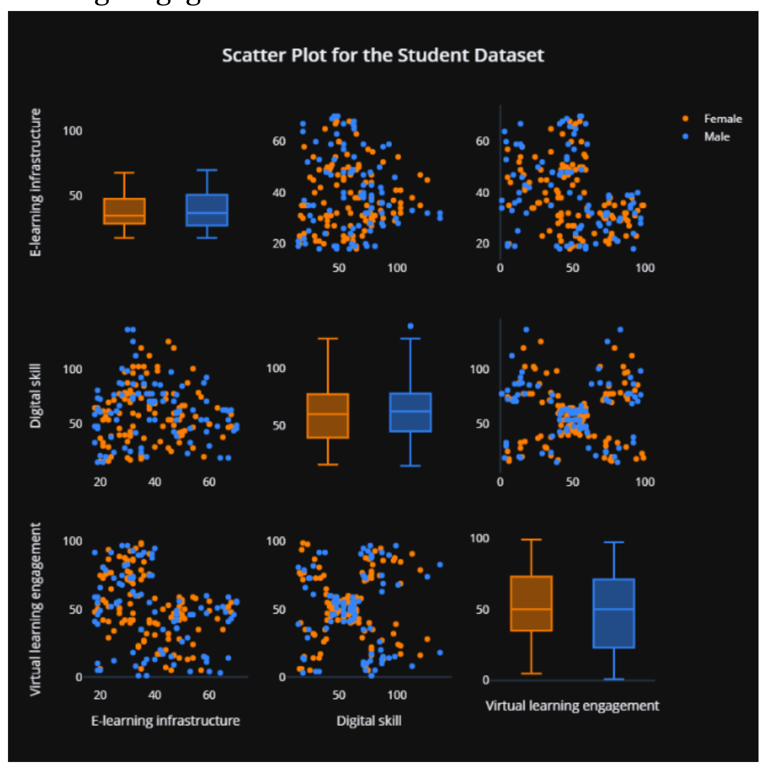

Figure 5. Scatter plot for student dataset

have is greater than that of male students in the same group. The most typical range for digital skills is between 70 and 79, with the total number of virtual learning engagements for female students coming in at value 1076 and the total number for male students coming in at value 823.

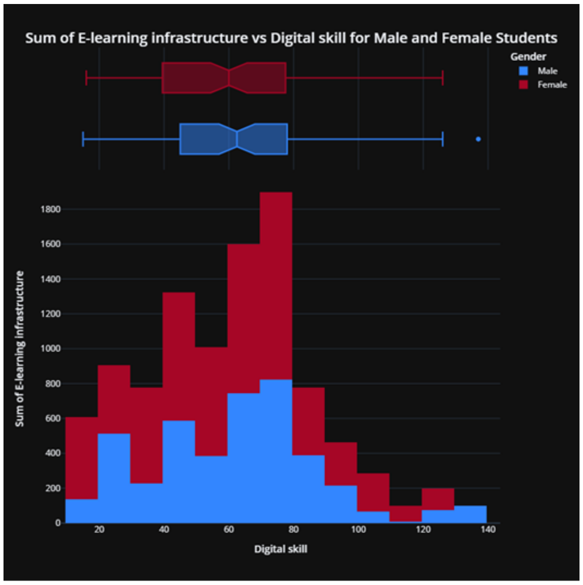

Figure 6. E-learning infrastructure vs Digital skill for Male and Female Students

The dataset also contains an anomaly in the form of a male student with an E-learning infrastructure value of 30 who has a Digital skill score of 137.

We also examined the distribution of the various data points for the various characteristics by using the scatterplotmatrix that corresponds to the dataset. Scatter plots, however, make it difficult to see and identify the link between the aspects of the dataset. This makes it difficult to draw conclusions about the data.

There is a marginally positive association between some of the traits and others. For instance, because the connection between Virtual learning engagement and Digital skill is just 0.01, we may believe that these two factors are unrelated to one another. Aside from this, the dataset also contains weak negative correlations. For example, Virtual learning engagement and E-learning infrastructure have a correlation of -0.330, which suggests that students with low E-learning infrastructure have low Virtual learning engagement.

Result from K-Means Clusters: Cluster 0 denotes the group of students that have a poor level of digital proficiency (less than 40) and thus have a low level of involvement in their virtual learning (less than 40). Students that have a high level of digital proficiency (more than 50) and a high level of involvement in virtual learning (more than 70) belong in Cluster 1.

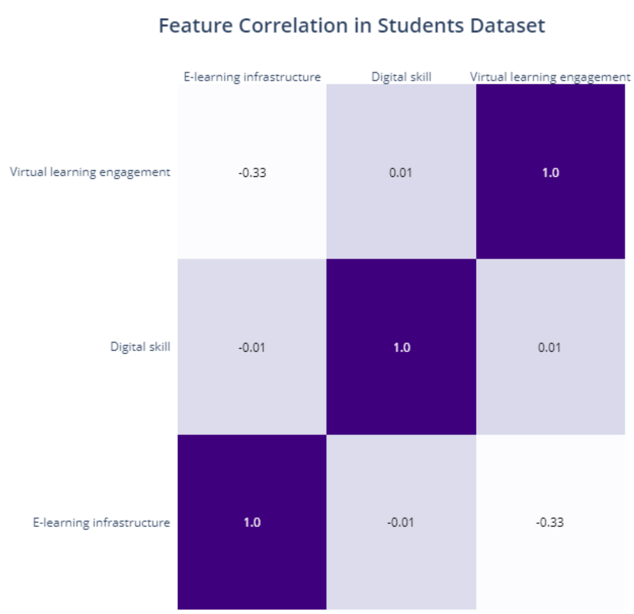

Figure 7. Feature correlations

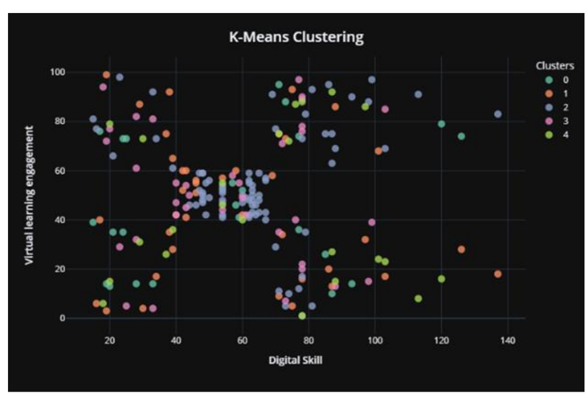

Figure 8. K-means clusters among students

Students who fall into the category of having an average E-learning infrastructure Digital skill (between 40 and 70) and an average E-learning infrastructure Virtual learning engagement may be identified with the

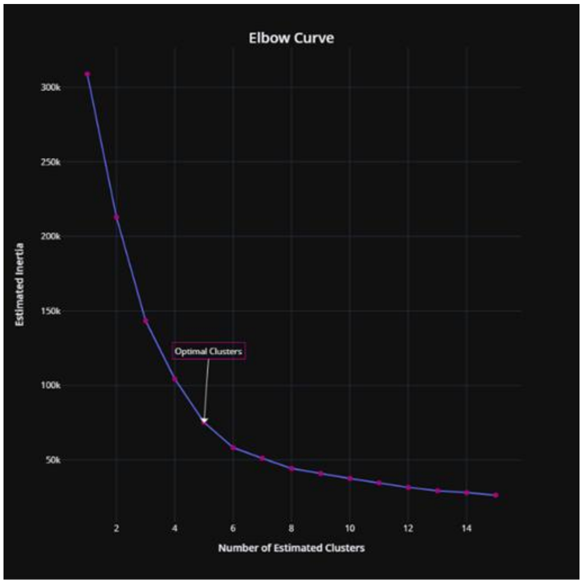

Figure 9. Elbow curve

assistance of Cluster 2. (40 - 60). Cluster 4 is for the group of students who, although having poor Digital skills, have a high Virtual learning engagement. These students are engaged in their learning in a virtual environment. Cluster 5 is the last group, and it reflects the category of students who, although having high levels of digital proficiency, have a low level of involvement in virtual learning.

Result from DBSCAN Clusters: Outliers in the dataset are denoted by white circles, and these are observations that do not fall into any of the three categories. The class of students that fall into Cluster 0 are those who have a Digital skill of less than 60 and a Virtual learning engagement of more than 40. Students that fall into Cluster 1 have an exceptionally low level of digital proficiency (below 40) and a low level of involvement in virtual learning (below 20). Cluster 2 represents the class of students who have a high level of involvement in virtual learning (>60) and a high level of digital ability (>70). Students that fall into Cluster 3 have a high level of digital proficiency (above 70) but a low level of engagement with virtual learning (below 40).

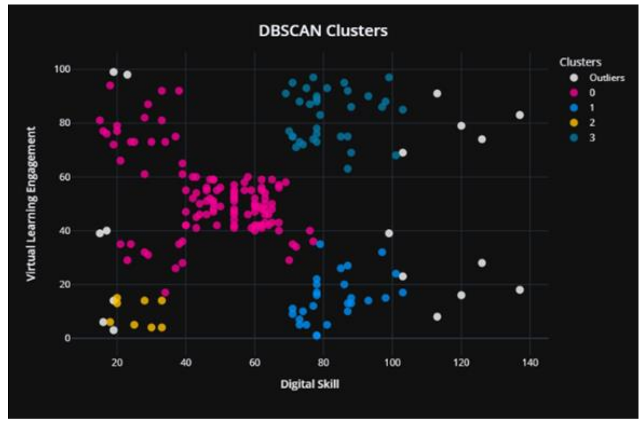

Figure 9.1 DBSCAN clusters

Result from Hierarchal Clusters: Students who fall into the Cluster 0 category have an average E-learning infrastructure Digital skill (between 40 and 60) and an average E-learning infrastructure Virtual learning engagement (40-60).

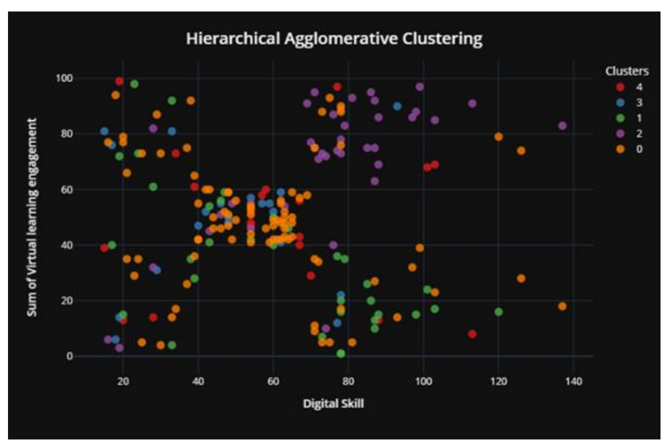

Figure 10. Hierarchal Clusters

Students that fall into Cluster 1 have a strong Digital skill (more than 70) but a low Virtual learning engagement (fewer than 40). Students that fall into Cluster 2 have a high level of digital literacy and participate in more than 60 percent of their available virtual learning opportunities. Students that fall into Cluster 3 have digital skills of less than 40 thousand and a high level of engagement in virtual learning (more than 60). Last but not least, students that fall into Cluster 4 have a digital skill level of less than 40k and a virtual learning engagement level of less than 40.

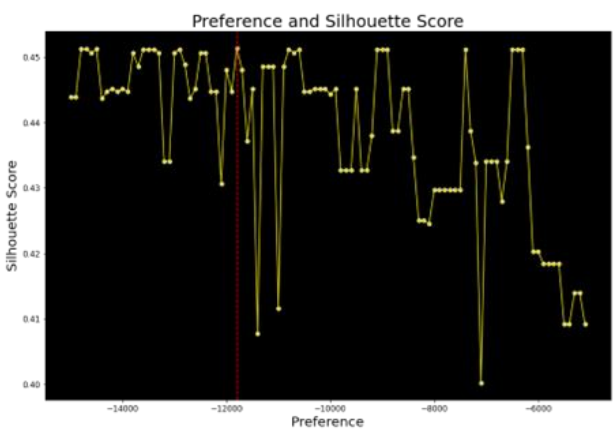

Figure 11. Silhouette score

The Silhouette Score shown in figure 11 is used to determine the value of preference that is picked. Due to the fact that its Silhouette Score is 0.45, which is the greatest possible value, we have decided to go with -11, 800 as our choice.

Result from Affinity Propagation Clusters: Students that fall into the Cluster 0 category have an average level of E-

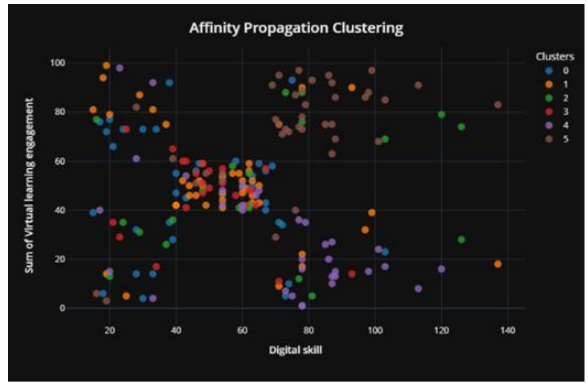

Figure 12. Affinity propagation clusters

learning infrastructure Digital Skill (40-60), as well as an average level of E-learning infrastructure Virtual Learning Engagement (40-60). Cluster 0 and Cluster 1 both reflect a category of students that is virtually identical to one another, and the difference between the two clusters is almost indistinguishable. Students with a digital skill score of less than 40 and a strong level of engagement in virtual learning are grouped together by Custer 2. (60-100). Cluster 3 allows us to identify students that have both a poor level of digital skill (below 40) and a low level of engagement with virtual learning (0-40).

Students in Cluster 4 have a high level of digital skill (>80) and a high level of engagement in virtual learning (>60), making Cluster 4 the polar opposite of Cluster 3. Last but not least, Cluster 5 illustrates students who have a high level of digital skill (>80) but a low level of engagement in virtual learning (40).

The instructors at the schools will be able to develop individual course plans for each of the five clusters that were discovered in this research.

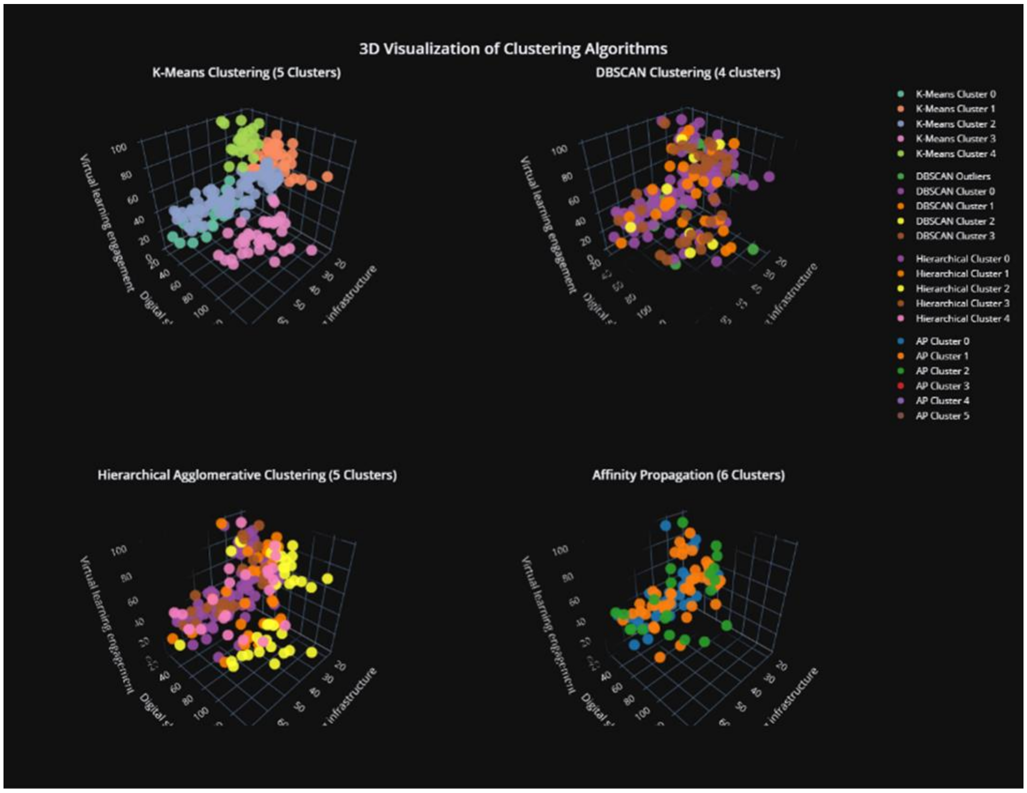

Figure 13. Different clusters in 3D coordinate

5. Conclusion

Colleges and universities in the modern day are confronted with a diverse set of issues, some of which include students who are not interested in their study, increasing rates of dropouts and the inefficiency of the conventional "one size fits all" method to education. However, when Artificial Intelligence and machine learning are applied effectively and responsibly, personalized learning opportunities may be developed, which may in turn assist to alleviate some of these difficulties. The use of intelligent approaches makes it possible to provide the system with features that allow for customization and to adapt it to the specific needs of individual students. It would have been more efficient to develop several versions of training procedures and materials if the interests of each learner had been taken into account. It will be possible to restrict the suggestions to a fixed number of users without losing the ability to personalize them if learners are divided into groups based on the similarities in their characteristics.

The monitoring of how data is used is a significant obstacle that must be overcome before AI technology can be fully implemented. Concerning the ownership of data, and most effective and morally sound approaches to making use of data, choices that are not only challenging but also very crucial will have to be taken across every social level.

Even though algorithms may be useful in directing choices, not all education programs should be managed by computers and algorithms. This is despite the fact that algorithms can be useful in guiding decisions. Instead, the aid that may be offered by AI algorithms need to be used to help promote the development of excellent educational settings.

References

[1] Patrick, Kennedy, and Powell, "Mean What You Say: Defining and Integrating Personalized, Blended and Competency Education," *International Association for K-12 Online*, 2013.

[2] B. Bray and K. McClaskey, "A Step-by-Step Guide to Personalize Learning," *Learning & Leading with Technology*, 2013.

[3] B. Kallick and A. Zmuda, "Students at the center: Personalized learning with habits of mind," 2017.

[4] M. E. Azukas, "Cultivating blended communities of practice to promote personalized learning," *Journal of online learning research*, 2019.

[5] J. Brass and T. L. Lynch, "Personalized Learning: A History of the Present," *JCT Online*, vol. 35, no. 2, Sep. 2020.

[6] Rathee and Rathee, "Concentration, Reactive-Proactive Aggression, Non-verbal Aggression, and Wellness among Video Games Playing Adolescents," *Int. J. High Risk Behav. Addict.*, 2014.

[7] M. B. Horn and H. Staker, *Blended: Using disruptive innovation to improve schools*. London, England: Jossey-Bass, 2017.

[8] C.-M. Chen, "Intelligent web-based learning system with personalized learning path guidance," *Comput. Educ.*, vol. 51, no. 2, pp. 787–814, Sep. 2008.

[9] M. Vandewaetere and G. Clarebout, "Advanced Technologies for Personalized Learning, Instruction, and Performance," in *Handbook of Research on Educational Communications and Technology*, J. M. Spector, M. D. Merrill, J. Elen, and M. J. Bishop, Eds. New York, NY: Springer New York, 2014, pp. 425–437.

[10] D. Nandigam, S. S. Tirumala, and N. Baghaei, "Personalized learning: Current status and potential," in *2014 IEEE Conference on e-Learning, e-Management and e-Services (IC3e)*, 2014, pp. 111–116.

[11] Zmuda, "Personalized learning," *Excellence through equity: Five principles of*, 2015.

[12] A. Pardo *et al.*, "OnTask: Delivering Data-Informed, Personalized Learning Support Actions," *Learning Analytics*, vol. 5, no. 3, pp. 235–249, Dec. 2018.

[13] H. Xie, H.-C. Chu, G.-J. Hwang, and C.-C. Wang, "Trends and development in technology-enhanced adaptive/personalized learning: A systematic review of journal publications from 2007 to 2017," *Comput. Educ.*, vol. 140, p. 103599, Oct. 2019.

[14] S. C. Kong and Y. Song, "An experience of personalized learning hub initiative embedding BYOD for reflective engagement in higher education," *Comput. Educ.*, vol. 88, pp. 227–240, Oct. 2015.

[15] F. Grivokostopoulou, I. Perikos, and I. Hatzilygeroudis, "Using Semantic Web Technologies in a Web Based System for Personalized Learning AI Course," in *2014 IEEE Sixth International Conference on Technology for Education*, 2014, pp. 257–260.

[16] P.-H. Liang and J.-M. Yang, "Virtual Personalized Learning Environment (VPLE) on the Cloud," in *Web Information Systems and Mining*, 2011, pp. 403–411.

[17] M. A. Sereno, "The Impact of a Personalized Learning Framework on Student Achievement," search.proquest.com, 2018.

[18] Tawafak, Romli, Malik, and Shakir, "A systematic review of personalized learning: Comparison between e-Learning and learning by coursework program in Oman," *in Learning*, 2019.

[19] J. F. Pane, E. D. Steiner, M. D. Baird, and L. S. Hamilton, "Continued Progress: Promising Evidence on Personalized Learning," *RAND Corporation*, Nov. 2015.

[20] J. F. Pane, E. D. Steiner, M. D. Baird, L. S. Hamilton, and J. Pane, “Informing Progress: Insights on Personalized Learning Implementation and Effects. Research Report. RR-2042-BMGF,” *RAND Corporation*, Jul. 2017.

[21] R. M. Cormack, B. S. Duran, and P. L. Odell, “Cluster Analysis: A Survey,” *J. R. Stat. Soc. Ser. A*, vol. 139, no. 1, p. 132, 1976.

[22] G. W. Milligan and H. C. Romesburg, “Cluster Analysis for Researchers,” *J. Mark. Res.*, vol. 22, no. 2, p. 224, May 1985.

[23] E. Bynen, *Cluster analysis Survey and evaluation of techniques*. Berlin, Germany: Springer, 1973.

[24] Roelandt and Hertog, “Cluster analysis and cluster-based policy making: the state of the art,” *Boosting innovation: The cluster*, 1999.

[25] B. S. Everitt, S. Landau, M. Leese, and D. Stahl, *Cluster Analysis*, 5th ed. Nashville, TN: John Wiley & Sons, 2011.

[26] Ares, “Cluster analysis: Application in food science and technology,” *Mathematical and Statistical Methods in Food Science*, 2014.

[27] Speece, “Methodological issues in cluster analysis: How clusters become real,” *Learning disabilities: Theoretical and research*, 1990.

[28] W. S. Sarle, L. Kaufman, and P. J. Rousseeuw, “Finding groups in data: An introduction to cluster analysis,” *J. Am. Stat. Assoc.*, vol. 86, no. 415, p. 830, Sep. 1991.

[29] Vermunt and Magidson, “Latent class cluster analysis,” *Applied latent class analysis*, 2002.

[30] V. S. Rathee, H. Sidky, and B. J. Sikora, “Role of associative charging in the entropy–energy balance of polyelectrolyte complexes,” *Journal of the American*, 2018.

[31] Hemanth, “Dimensionality Reduction of Production Data Using PCA and DBSCAN Techniques,” *Intelligent Systems and Computer Technology*, 2020.

[32] V. S. Rathee, A. J. Zervoudakis, H. Sidky, B. J. Sikora, and J. K. Whitmer, “Weak polyelectrolyte complexation driven by associative charging,” *J. Chem. Phys.*, vol. 148, no. 11, p. 114901, Mar. 2018.

[33] Duan, “Density-based clustering and anomaly detection,” *Business intelligence-solution for business*, 2012.

[34] Szymkowiak and Larsen, “Hierarchical clustering for datamining,” *Proceedings of KES-2001*, 2001.

[35] Amar, Labzour, and Bensaid, “Semi-Supervised Hierarchical Clustering Algorithms,” *SCAI*, 1997.

[36] Maimon and Rokach, “Clustering methods,” *Data Min. Knowl. Discov.*, 2005.

[37] V. S. Rathee, H. Sidky, B. J. Sikora, and J. K. Whitmer, “Explicit Ion Effects on the Charge and Conformation of Weak Polyelectrolytes,” *Polymers* , vol. 11, no. 1, Jan. 2019.

[38] Bholowalia and Kumar, “EBK-means: A clustering technique based on elbow method and k-means in WSN,” *Int. J. Comput. Appl.*, 2014.

[39] B. J. Frey and D. Dueck, “Clustering by passing messages between data points,” *Science*, vol. 315, no. 5814, pp. 972–976, Feb. 2007.

[40] A. Rodriguez and A. Laio, “Clustering by fast search and find of density peaks,” *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.

[41] X. Liu, M. Yin, J. Luo, and W. Chen, “An improved affinity propagation clustering algorithm for large-scale data sets,” in *2013 Ninth International Conference on Natural Computation (ICNC)*, 2013, pp. 894–899.

[42] H. Sidky *et al.*, “SSAGES: Software Suite for Advanced General Ensemble Simulations,” *J. Chem. Phys.*, vol. 148, no. 4, p. 044104, Jan. 2018.

[43] K. Wang, J. Zhang, D. Li, X. Zhang, and T. Guo, “Adaptive Affinity Propagation Clustering,” *arXiv [cs.AI]*, 08-May-2008.