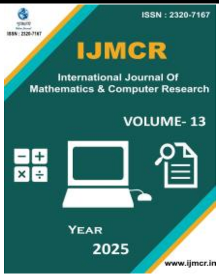
International Journal of Mathematics and Computer Research

ISSN: 2320-7167

Volume 13 Issue 03 March 2025, Page no. – 5000-5013

Index Copernicus ICV: 57.55, Impact Factor: 8.615

DOI: 10.47191/ijmcr/v13i3.13



Dataset Generation and Cluster Creation for Adaptive E-Learning System Using the Rectilinear Technique of K-Means Clustering, Demonstrated Using Java

Ibuomo R. Tebepah[1], Efiyeseimokumo S. Ikeremo[2]

[1]Department of Computer Science, Nigerian Defence Academy, Kaduna State.

[2]Department of Mathematical and Computer Science, University of Africa, Bayelsa State.

ARTICLE INFO	ABSTRACT
Published Online: **29 March 2025** **Corresponding Author:** **Ibuomo R. Tebepah**	Data is an essential element in research which can be challenging to obtain especially when such data is termed classified. Consequently, researchers depend on dataset or direct collection from respondents. The volume of data collected through this means is grossly limited and laborious putting into consideration, the resources involved in the collection and accuracy rate. In order to ease this, researches that requires demonstration, can rely on internally synthetic generated data. This work looks at how data can be generated using Java multi-dimensional array, and the classification of generated data into cluster using the k-means rectilinear technique, that is used to classify adaptiveness of learners in an eLearning environment. With the combination of simple and complex codes, the work adequately and accurately generated 125 elements, created 5 clusters based on the fusion of known and adopted learning pedagogies, which can be used to determine how learners learn different subject matters.

1. BACKGROUND

Learning is an integral part of human existence, and it is believed that humans learn every day. However, the concept of learning today, has gone beyond the conventional methods people were used to. Learning comes with information. In the past, information was usually sought after; now information locates us. The future of education (teaching and learning) is intrinsically linked with emerging technologies and computing. This is obvious in the accelerated pace at which technology has been embedded in content development, content presentation, content distribution, learner assessment techniques, and mode of learning. A large number of studies attest to the fact that learning is no longer appealing without the influence of technology in today's world. It could be regarded as 'old fashion'. It is, therefore a fact that learning has become more interesting, vibrant, and accessible due to the adaptive e-learning system.

Not only has educational institutions standardized the way in which learners' content is developed or presented, they have also incorporated a digital culture into the curriculum, which requires a certain level of computer literacy even for primary school pupils. This is because, schools have implemented the utilization of learning management system such as Moodle, and learning platforms such as Google classroom, Edmodo, Power school etc. There are also massive open source online courses such as courser.com and ocw.mit.edu that provide free degree level programmes.

The advent of artificial intelligence and its related components has further altered education in a radical way with the booming domain that has produced new teaching and learning solutions that are now undergoing testing in different contexts (United Nations Educational, Scientific and Cultural Organization [UNESCO], 2019). Artificial intelligence takes advantage of datafication (objectively quantified data used for tracking, modeling, and prediction for individualized learning) to mobilize human-computer agent interactions.

Using the internet or intranet to facilitate learning has become a trend in modern places of education and corporate organizations. Generally, E-Learning means Electronic Learning. It means the presentation of learning content in an electronic format hard or soft copy. The hard copy aspect of E-Learning is gradually fading away as more people spend time on the internet. For the purposes of this work, E-Learning is limited to online virtual learning using internet, intranet or standalone systems that have access to the learning software.

E-Learning is interactive learning in which the learning content is online. It is presented as applications which can be installed on a computer, mobile device, or online-based (Cloud). This makes it interesting, because it can be accessed from anywhere and at any time.

2. TRENDS IN ADAPTIVE ELEARNING

Adaptive learning has leveraged on Natural Processing Language (NPL) capability found in programming languages such as Python and Java, used in developing systems. For instance, Wambsganss et al. (2020), incorporated NPL adaptive learning to enable students write better argumentation structure texts using the Argumentation Tool in the system. The tool was also used as a feedback mechanism to ascertain how beneficial it is.

The research of Dunn and Kennedy (2019) focused on the usage of Technology Enhanced Learning (TEL); a sharp move away from learning style advocacies. According to the research, although there is significant usage and engagement with the system based on learning style, yet what determines usage is extrinsic motivation and not intrinsic. The extrinsic predictor is that of the use of social media, not recorded lectures, course content blog or virtual environments. This is agreeable if we consider the time students spend on social media content. The issue then is how do developers incorporate or integrate social media and a learning curriculum without the rigor of interoperability issues.

Dziuban et al. (2017) in their work noted that adaptive learning systems have potentials for accommodating student differences in an increasingly diverse population. The work demonstrated and addressed demography variability by customizing course content according to differences in students' skill sets. According to them, adaptive systems have demonstrated the capability to understand where students are and take these students where they need to be, while making assessment part of the learning process.

The fact that designers of adaptive systems are taking into cognizance, not only the technology but the indicators of learning—cognitive, affective, behavioral, and learning style has further strengthened the use of these systems. The research confirmed that students showed a higher level of retention, competence and adaptability when they were introduced into this system due to the varied number of exercises they had to do, the feedback mechanism, and the various learning activities incorporated in the system.

The reviewed features of adaptive systems in the work of Ennouamani and Mahani (2017) provided developers a premise to build model goals, preferences, and knowledge of each individual user, in order to adapt the learning to their needs and characteristics. Their research also gave the similarities and differences in features of the adaptive E-Learning systems. It further revealed that scientific research in the field of E-Learning is oriented towards learning platforms where learner's expectations, motivations, learning styles, habits and needs were increasingly taken into consideration. It identified the main source of adaptive learning been related to the objective of each system, as well as the result. The approaches to adaptive learning as identified in the paper are macro aptitude, and micro adaptive. The suggestions and findings were not implemented in the paper.

Ahmed (2018) study revealed the automatic approach as a better approach to identifying learning styles in online learning, because it is based on the student's behavioral pattern while learning. Their main objective was to identify the learning style based on pattern of behavior for 20 students who studied interactive multimedia. The findings indicate that predicted interactive learning styles are different, because they depend on the actual behavioral pattern of students.

Explaining further Richter and Latchem (2018) reviewed 3,674 articles that bordered on computers and education. It showed the rapid growth from an era of computer based instruction to standalone multimedia learning. It moved to networked computers as tools for collaborative learning and finally to online learning in this digital age. The work demonstrated the influence of technology on education and its impact on the way students learn. It also showed how researchers are finding new mechanisms of adapting the rapid changes in technology to education.

Xie et al. (2019) reviewed the trends and developments of technology enhanced adaptive learning in the last ten years (2007-2017), and the outcome showed that despite the many adaptive systems, majority are not mobile friendly. This is not good, as learners have moved from desktops and laptops to smart devices. It is therefore necessary to keep the trend, if technology has to be used for learning. It is no use developing an adaptive system that is not used by the targeted users.

In the light of the above, Hermanwan et al. (2018) looked at implementing an adaptive learning system on mobile platforms rather than on the conventional work station or desktop. The reason for the shift is the rise in percentage of smart device users who also happen to be in the active learning age bracket. The implementation was divided into: adaptive content, adaptive assessment, and adaptive sequence. It also recommended user, content, skill or difficulty level, and performance as the common factors that should be used in building or developing an adaptive mobile learning system. It is worthy of note that implementing all these in a single adaptive learning system will be difficult because it requires systematic thinking skills and sophisticated algorithms.

Alian and Al-Akhnas (2010) work is a web-based adaptive E-Learning environment called AdaLearn. The work AdaLearn saves learners' responses into profiles which are used for future guidance. The system responds to learners differently, in adapting the presentation of learning content to meet the varying needs and learning preferences of individual learners. The learner can select their modular content and customize their learning environment, to enable them get flexible solutions that dynamically adapt content to fit individual learning needs. The algorithm used is based on navigation;

i.e. data collected based on the learner's usage of the system. The work is based on post analysis and using user navigation as the only input for analysis will create errors and not well defined profiles. This is due to the fact that; first-time users of the system are bound to make a lot of errors. In addition, a learner may merely navigate through the interface, just to have a feel of the system. So, basing content presentation and its delivery on user navigation only is not representative enough. Other parameters should be included to make it more acceptable and robust in its analysis.

3. ADOPTED LEARNING PEDAGOGY REVIEW

The selection and adoption of learning pedagogies was based on; usage, popularity, recommendation by specialist, and availability of previous works.

4 MAT: The 4 Mode Application Techniques (4 MAT) learning style model was developed by Bernice McCarthy, and it is modelled along four continuum or fields which can be sub grouped into two categories. According to McCarthy (1990), learners capitalize on learning styles that they are comfortable with. These learning styles involve brain dominance in a way that the learner has processing preferences. In the work, the research stated the four fields as: the imagination field, the analytic field, the common-sense field, and the dynamic field. It infers that every learner perceives or processes information that comes to them in one of the stated ways. 'Perceiving' is the way learners take in new information, and 'processing' is what people do with the new information as stated by Seker and Ovez (2018), which results into change or corresponding actions.

In Irfan et al. (2016), the 4 MAT system of instruction is a strategy that deals with individual differences. It works to develop or modify the procedures of teaching so that they specifically address individual differences. This supports the work of Aktas and Bilgin (2014) in using the model on 7th grade students to improve their performance of a science subject matter. It showed that 4 MAT increased students' motivation levels, participation and self-confidence in the science subject. The model has also been used in inter-disciplinary structures and academic assessments for engineering students. It was discovered to be effective, and it produced workable results.

Despite its successes, the model limits learning styles to just four possible ways. This is not justifiable as a learner can perceive and process information in several other ways different from those enumerated by the 4 MAT learning style model. Also, the performance variation was only based on a singular group and on a specific subject matter test. More learning groups and subject matter should be simultaneously used to ascertain the validity of the recorded improved performance on the learners.

Gregorc Mind Style Model: This provided an organized way to consider how the mind works. According to him, learners learn with ease when the learning environment is in sync with the learning style but becomes challenging when it is not. The Gregorc Style Delineator (GSD) is designed to assist learners recognize and identify ways or mediums through which the learner receives and expresses information in an effective, efficient and economical way. This means that learners will most likely use what they have conceived as the best way to learn for them. The model pre-supposes that there is usually a premise for learning which may be formal or informal and with this experience of knowledge, learning occurs. It supports the 'none tabula rasa' concept of educational philosophers. The work of Alduals (2018) explains that in the model, none of the perceptual abilities is absolute, rather there is a dominating one that a person would use more comfortably than the other.

However, it is difficult to say if this generalization is applicable to all the four modalities. In other words, if there is a dominating perceptual ability and a dominating orderly ability, would that formula be applicable to a dominating learning modality? If it is, then the learning model contradicts itself. Also, categorizing the model into four distinct combinations of Concrete-Sequential (CS), Abstract-Sequential (AS), Abstract-Random (AR), and Concrete-Random (CR) in a metrics of 10 – 40, with a maximum of 100 points for all four is challenging (Hawk and Shah, 2007). This is because, it is difficult to completely categorize all learners' preferences and attributes into just four categories. It showed some form of effectiveness when it was used to measure learning styles of the participants in a mental model assessment. It gave evidence that gender and learning styles can be used to associate mental models in order to provide a group base (Lau and Yuen, 2010). However, the application is still limited to the four categories, which are not comprehensive enough considering the frequency in the learners' preferences changes depending on content and presentation.

Kolb Learning Style: The Kolb Learning Style sets out four distinct learning continuum across two dimensions: concrete vs abstract, and active vs reflective (Kolb and Kolb, 2019). This involves experiences based on feelings, watching others and developing opinions, creating theories to explain observations and using the developed theories to solve problems and make decisions. The Kolb model presents a less formal lecture/classroom orientation to the learners. It is a subtle approach to instructional methodology of Svinicki and Dixon (2010) that enhanced learning. It takes learning to be an experiential cycle in which a learner experiences information and tries to reflect or rationalize the information to make meaning to him or her. This reflection could be abstract or reflective observation, after which a theoretical perception will be formed, that will lead to active experience or concrete experience.

This model has formed the premise for other learning style models such as 4 MAT and Honey & Mumford. So, it has been widely used and applied in various fields of studies.

Despite its achievements, the fact remains that the subject matter influences learners' choice of style and often times, it

requires a combination of more styles to achieve learning. Since learners' preferences are not stagnated, it means that the learning styles of learners are also not stagnant. It evolves over a period of time, with maturity, exposure (experience), and the frequency of brain work performed by the learner.
Honey and Mumford Learning Style: This model is a modification of the Kolb style but with few differences. Sangvigit (2012) identified four stages that includes identification of problem, experience, gathering of information, and thinking through to identify new ideas. According to the model, learners usually go through these four stages, during the learning process. Distinctively stating the stages as the way learning occurs is not accurate, as other researches have proved otherwise. The way learning occurs varies from learner to learner, and it is the reason teachers and psychologists advocate for learner preference considerations when determining the learning to style to use for a learner.

Felder and Silverman Learning Style Model: Felder and Silverman learning style has four paired option dimensions that expresses different aspect of learning with a linguistic variable (Birol, 2016). Each learner, according to Felder and Silverman is characterized by a specific preference for each dimension. These dimensions are:
Birol (2016) work states the active or reflective way of processing information. Active learners learn best by working actively with the learning material, and they tend to work better in teams. On the other hand, reflective learners are loners. They prefer to work alone or in very small groups of two or three people. They learn by critical thinking and analysis.
Intuitive or sensing learners learn by understanding facts and concrete materials that deal with theories and underlying principles.
Visual or verbal – This group prefers to learn from what they see, such as diagrams, pictures, images, and videos. The verbal learners on the other hand, learn from texts, spoken words and audio content presentation.
Sequential or global – Sabine et al. (2007) in their work revealed that sequential learners learn in small incremental steps and therefore have a linear learning process. They are very logical in their approach to solving problems. In contrast, global learners use the holistic thinking process and learn in large random leaps. They eventually put all the pieces together and solve the problem but may not necessarily be able to explain how they did it.
In reviewing the learning style models, while some highlights just two pair possibilities, only Felder and Silverman highlights four pairs which reasonably covers most aspects of the human personality.
However, learning has taken a different route in our world today, as everything is done on the go now; including learning. So, while the original learning style models which are learner character-oriented will always form the premise for most research works, content materials also have great influence in determining learners' learning style. It has been observed that a learner's style of learning depends on the subject matter and individual's attributes. Determining an adaptive learning style for learners has become even more difficult, and it has become clear that a singular learning style may not be able to capture the totality of a learner's preference for all subjects. This is the reason, generalization, discretion, and standardization is necessary when designing an adaptive multiple learning system for learners.

4. K-MEANS CLUSTERING ALGORITHM

The clustering method is widely used for segmentation and identification of data points and the cluster the data belongs to.
Many researchers had worked with the applications of the K-Means Clustering and Fuzzy-C Means Theory on educational areas such as in student assessment. Both algorithms have been used for an assessment system to translate numeric grades into letters.
K-Means Clustering (KMC) and Fuzzy-C Means (FCM) has been regarded appropriate for application in automatic navigation, variable computation and complex logic analysis (Chen et al., 2017). This is mainly because of its capacity to process large quantities of incomplete and inaccurate input signals and data (Huang, 2017).
However, despite the similarities between the two algorithms, the computation process and clustering process varies, which also affects the overall performance of each algorithm. For instance, Velmurugan (2014) and Wihanto and Surani (2020) made a comparative analysis of FCM and KMC and discovered that performance of KMC drops when the number of clusters is less than four, and FCM takes more processing time. This is similar to the fact that FCM presupposes that a data point partially belongs to more than one cluster, thus the iteration operation increases the processing time. The paper did not clearly state which is preferred of the two, as the performance of each is relatively stable as shown in the research.
In their work, Zeynel and Yildiz (2018) further reiterated the issue of performance and highlighting the high accuracy rate of both algorithms; with KMC leading or recommended for separated cluster structures spreading regular patterns in the data set.
The K-Means Clustering (KMC) algorithm computes or determines the centroid (center point) in a dataset by iterating through the dataset until the data points belonging to a particular centroid and the centroid remains significantly unchanged. The data set is usually unlabeled and unsupervised. This makes the algorithm applicable in a scenario where the output is undefined.
The learning process is an internal cognitive event says Machado et al. (2016), and the use of computing tools will stimulate the changes that occur during the teaching and learning process. Therefore, modeling or mapping computer aided learning system to learners' needs is of high priority

and value. It harmonizes the content, learning style and content presentation that is brought to the learner.

K-Means Clustering has been used for several experiments in relation to health, education and others. In the research of Ambaselker and Bagwan (2016), K-Means Clustering was used as the base for selecting top ranked mining rules for genes. It identified and associated the weighted condensed support and weighted condensed confidence with genes. Though it had predetermined numbers of clusters, it was able to deduce the appropriate clusters for the different data points in the gene dataset.

K-Means Clustering can also be combined with other algorithms for an improved result. The strength of the algorithm can be used to compliment the weaknesses of other algorithms for an improved result. Dharshinni et al. (2019) used K-Means Clustering with Apriori algorithm, and it was able to reduce the iteration time greatly thereby improving the result. K-Means Clustering algorithm uses a faster means of iteration and at the same time creates association between related data points.

The K-Means Clustering algorithm has different implementation techniques as shown in the work of (Xiuchange, 2014) in which an improved K-Means Clustering algorithm was used to show behavioural patterns in a website that is not sensitive to time. It was also used to show the trajectory usage of electricity during the seasons, and the peak usage in each season.

In using the K-Means Clustering algorithm, the fact that it is expandable and of high efficiency (Liu et al., 2014), makes it easy to improve upon. It can be modified and combined easily with other algorithms to improve the end result. The work introduced the concept of the smallest rule covering set in K-Means Clustering, targeted at audit monitoring and discovery, as well as extraction of processes.

Although the K-Means Clustering can be implemented or applied with modification, the initial clustering process or standard is still adhered to. The general standards of K-Means Clustering are:

i. There is a given set of ‘n’ data points in d-dimensional space
ii. An integer ‘k’ determines a set of ‘k’ points in d-dimension called centroid
iii. Each ‘n’ data point has to minimize the mean square distance from each data point to its nearest centroid
iv. Data point iteration and new centroid identified
v. Iteration is altered when the centroid of a data point remains constant.

5. METHODOLOGY

This section highlights various segments of the work; identification of learning pedagogies, dataset generation, creating of clusters based on learning pedagogies, and stating the techniques and usage.

5.1 Procedure

5.1.1 Identification of learning pedagogies

In section 3 (adoption of learning pedagogy), a review of 5 popular learning styles were made. Table 1 is a summary of findings

Table 1 Adopted Learning Style Summary

s/no	Theorist / Model	Model Features
1	4MAT	• Imagination • Common sense • Analysis • Dynamic
2	Gregorc Mind Style	• Concrete-Sequential (CS) • Abstract-Sequential (AS) • Abstract-Random (AR) • Concrete-Random (CR)
3	Kolb	• Concrete vs abstract • Active vs reflective
4	Honey & Mumford	• Identification • Experience • Gathering of information • Thinking through
5	Felder & Silverman	• Active or reflective • Intuitive or sensing • Visual or verbal • Sequential or global

Using observation, extraction, usage, and comparison techniques on each model, common elements were identified, which were classified into the following categorization as;

i. Activity
ii. Experience
iii. Theory (Abstract conceptualization)
iv. Demonstration (Concrete)
v. Visuals

This implies that a learner's learning styles will be a combination of two or more of these based on their preferences and the subject matter that is been presented.

5.1.2 DataSet Generation

DataSet was generated using Java programming language. The code contains 5 sets of 2 dimensional arrays containing 5 elements in each arrays. The 5 paired (2 dimensional) arrays represent the 5 learning categorization, the 5 elements represent the possible expected responses, and 5, 5 representing the arrays rows and columns.

```
public static void main(String [] args){
    int [][] da = {{5, 4, 3, 2, 1}, { 1, 2, 3, 4, 5 }};
    int [][] dt = {{5, 4, 3, 2, 1}, { 1, 2, 3, 4, 5 }};
    int [][] de = {{5, 4, 3, 2, 1}, { 1, 2, 3, 4, 5 }};
    int [][] dv = {{5, 4, 3, 2, 1}, { 1, 2, 3, 4, 5 }};
    int [][] dd = {{5, 4, 3, 2, 1}, { 1, 2, 3, 4, 5 }};
}
```

For each set of array, it will generate possible 25 values originating from the combination within it. This will create 25 datapoints. In total, there will be 125 datapoints that makes up the generated dataset. Each datapoint is access using the array name and index.

The unsupervised syntactic dataset comprises of 125 datapoints, which means 125 possible answer combination, giving Table 2

Table 2 125 generated datapoints

A_{1}(5,5)	D_{1}(5,5)	E_{1}(5,5)	V_{1}(5,5)	T_{1}(5,5)
A_{2}(5,4)	D_{2}(5,4)	E_{2}(5,4)	V_{2}(5,4)	T_{2}(5,4)
A_{3}(5,3)	D_{3}(5,3)	E_{3}(5,3)	V_{3}(5,3)	T_{3}(5,3)
A_{4}(5,2)	D_{4}(5,2)	E_{4}(5,2)	V_{4}(5,2)	T_{4}(5,2)
A_{5}(5,1)	D_{5}(5,1)	E_{5}(5,1)	V_{5}(5,1)	T_{5}(5,1)
A_{6}(4,5)	D_{6}(4,5)	E_{6}(4,5)	V_{6}(4,5)	T_{6}(4,5)
A_{7}(4,4)	D_{7}(4,4)	E_{7}(4,4)	V_{7}(4,4)	T_{7}(4,4)
A_{8}(4,3)	D_{8}(4,3)	E_{8}(4,3)	V_{8}(4,3)	T_{8}(4,3)
A_{9}(4,2)	D_{9}(4,2)	E_{9}(4,2)	V_{9}(4,2)	T_{9}(4,2)
A_{10}(4,1)	D_{10}(4,1)	E_{10}(4,1)	V_{10}(4,1)	T_{10}(4,1)
A_{11}(3,5)	D_{11}(3,5)	E_{11}(3,5)	V_{11}(3,5)	T_{11}(3,5)
A_{12}(3,4)	D_{12}(3,4)	E_{12}(3,4)	V_{12}(3,4)	T_{12}(3,4)
A_{13}(3,3)	D_{13}(3,3)	E_{13}(3,3)	V_{13}(3,3)	T_{13}(3,3)
A_{14}(3,2)	D_{14}(3,2)	E_{14}(3,2)	V_{14}(3,2)	T_{14}(3,2)
A_{15}(3,1)	D_{15}(3,1)	E_{15}(3,1)	V_{15}(3,1)	T_{15}(3,1)
A_{16}(2,5)	D_{16}(2,5)	E_{16}(2,5)	V_{16}(2,5)	T_{16}(2,5)
A_{17}(2,4)	D_{17}(2,4)	E_{17}(2,4)	V_{17}(2,4)	T_{17}(2,4)
A_{18}(2,3)	D_{18}(2,3)	E_{18}(2,3)	V_{18}(2,3)	T_{18}(2,3)
A_{19}(2,2)	D_{19}(2,2)	E_{19}(2,2)	V_{19}(2,2)	T_{19}(2,2)
A_{20}(2,1)	D_{20}(2,1)	E_{20}(2,1)	V_{20}(2,1)	T_{20}(2,1)
A_{21}(1,5)	D_{21}(1,5)	E_{21}(1,5)	V_{21}(1,5)	T_{21}(1,5)
A_{22}(1,4)	D_{22}(1,4)	E_{22}(1,4)	V_{22}(1,4)	T_{22}(1,4)
A_{23}(1,3)	D_{23}(1,3)	E_{23}(1,3)	V_{23}(1,3)	T_{23}(1,3)
A_{24}(1,2)	D_{24}(1,2)	E_{24}(1,2)	V_{24}(1,2)	T_{24}(1,2)
A_{25}(1,1)	D_{25}(1,1)	E_{25}(1,1)	V_{25}(1,1)	T_{25}(1,1)

5.1.3 Cluster Creation using the rectilinear technique

With the dataset in place, the K-means clustering (KMC) algorithm was used to calculate the clustering each datapoint belongs to. This dataset is used to generate distinct centroid for each category, as a rule in K-means clustering, a datapoint can only belong to a single centroid.

Generally, since the dataset is grouped based on five the learning styles attributes, 5 clusters and their centroids are identified to start the clustering processing. The Rectilinear Distance Observation formula is used to iterate the dataset to obtain an "unchangeable" datapoints that belongs to a particular centroid. For datapoints to belong to same cluster, it means that the datapoints have similar attributes. Even though the K stands for the number of clusters, the number is usually defined based on how the dataset is organized. However, K (number of clusters) must be less than n (number of objects, obervations or datapoints) ($k < n$).

$$d_{u,v} = | u_1 - v_1 | + | u_2 - v_2 | + \ldots\ldots\ldots. + | uq - vq |$$

with v and u as vectors.

source: Msigwa (2022) Data science & machine learning

This takes the sum of the absolute differences between the observations in each vector.

The initial mean is a randomly selected point that serves as the initial centroid. The number of mean selected will be determined by the number of expected clusters, which is five (5) in this case.

To determine the first cluster, the randomly selected mean values are used to calculate the new mean point to in other to determine the new cluster each datapoint belongs. The following mean point were selected from the default clusters.

C1 – 5,5 (selected from Activity)
C2 – 1,2 (selected from Demonstration)
C3 – 4,3 (selected from Experience)
C4 – 2,3 (selected from Visual)
C5 – 3,5 (selected from Theoretical)

Each datapoint or observation is place in a cluster depending on the observation distance to the mean. The closer an observation is to a cluster, the most likely that the datapoint will fall into that cluster. Using these mean and the point to calculate the x and y, new cluster is identified. The new cluster a datapoint belongs is determined by its closest or nearest in distance in relation to the identified centroid.

Datapoint mean (C1)

x_1 , y_1 **x_2 , y_2**

5,5 **5,5**

$P(a,b) = | x_2-x_1 | + | y_2-y_1 |$

0+0 = 0

This formula is used to calculate each datapoint to arrive at a new or maintain the datapoint new or previous cluster.

Table 2: first cluster analysis after iteration

POINTS	C1 5,5	C2 1,2	C3 4,3	C4 2,3	C5 3,5	NEW CLUSTER
A_1(5,5)	0	7	3	5	2	**1**
A_2(5,4)	1	6	2	4	3	**1**
A_3(5,3)	2	5	1	3	4	**3**
A_4(5,2)	3	4	2	4	5	**3**
A_5(5,1)	4	5	3	5	6	**3**
A_6(4,5)	1	6	2	4	1	**1**
A_7(4,4)	2	5	1	3	2	**3**
A_8(4,3)	3	4	0	2	4	**3**
A_9(4,2)	3	3	1	3	4	**3**
A_{10}(4,1)	5	4	2	4	5	**3**
A_{11}(3,5)	2	5	3	3	0	**5**
A_{12}(3,4)	3	4	2	2	1	**5**
A_{13}(3,3)	4	3	1	1	2	**3**
A_{14}(3,2)	5	2	3	2	3	**2**
A_{15}(3,1)	6	3	3	3	4	**2**
A_{16}(2,5)	3	4	3	2	1	**5**
A_{17}(2,4)	4	3	3	1	2	**4**
A_{18}(2,3)	5	2	2	0	3	**4**
A_{19}(2,2)	6	1	3	1	4	**2**
A_{20}(2,1)	7	2	4	2	5	**2**
A_{21}(1,5)	4	5	5	3	2	**5**
A_{22}(1,4)	5	2	4	2	4	**2**
A_{23}(1,3)	6	1	5	1	4	**2**
A_{24}(1,2)	5	0	4	2	5	**2**
A_{25}(1,1)	8	1	5	3	6	**2**
D_1(5,5)	0	7	3	5	2	**1**
D_2(5,4)	1	6	2	4	3	**1**
D_3(5,3)	2	5	1	3	4	**3**
D_4(5,2)	3	4	2	4	5	**3**
D_5(5,1)	4	5	3	5	6	**3**
D_6(4,5)	1	6	2	4	1	**1**
D_7(4,4)	2	5	1	3	2	**1**
D_8(4,3)	3	4	0	2	4	**3**
D_9(4,2)	3	3	1	3	4	**3**
D_{10}(4,1)	5	4	2	4	5	**3**

$D_{11}(3,5)$	2	5	3	3	0	**5**
$D_{12}(3,4)$	3	4	2	2	1	**5**
$D_{13}(3,3)$	4	3	1	1	2	**3**
$D_{14}(3,2)$	5	2	3	2	3	**2**
$D_{15}(3,1)$	6	3	3	3	4	**2**
$D_{16}(2,5)$	3	4	3	2	1	**5**
$D_{17}(2,4)$	4	3	3	1	2	**4**
$D_{18}(2,3)$	5	2	2	0	3	**4**
$D_{19}(2,2)$	6	1	3	1	4	**2**
$D_{20}(2,1)$	7	2	4	2	5	**2**
$D_{21}(1,5)$	4	5	5	3	2	**5**
$D_{22}(1,4)$	5	2	4	2	4	**2**
$D_{23}(1,3)$	6	1	5	1	4	**2**
$D_{24}(1,2)$	5	0	4	2	5	**2**
$D_{25}(1,1)$	8	1	5	3	6	**2**
$E_{1}(5,5)$	0	7	3	5	2	**1**
$E_{2}(5,4)$	1	6	2	4	3	**1**
$E_{3}(5,3)$	2	5	1	3	4	**3**
$E_{4}(5,2)$	3	4	2	4	5	**3**
$E_{5}(5,1)$	4	5	3	5	6	**3**
$E_{6}(4,5)$	1	6	2	4	1	**1**
$E_{7}(4,4)$	2	5	1	3	2	**3**
$E_{8}(4,3)$	3	4	0	2	4	**3**
$E_{9}(4,2)$	3	3	1	3	4	**3**
$E_{10}(4,1)$	5	4	2	4	5	**4**
$E_{11}(3,5)$	2	5	3	3	0	**5**
$E_{12}(3,4)$	3	4	2	2	1	**5**
$E_{13}(3,3)$	4	3	1	1	2	**3**
$E_{14}(3,2)$	5	2	3	2	3	**2**
$E_{15}(3,1)$	6	3	3	3	4	**2**
$E_{16}(2,5)$	3	4	3	2	1	**5**
$E_{17}(2,4)$	4	3	3	1	2	**4**
$E_{18}(2,3)$	5	2	2	0	3	**4**
$E_{19}(2,2)$	6	1	3	1	4	**2**
$E_{20}(2,1)$	7	2	4	2	5	**2**
$E_{21}(1,5)$	4	5	5	3	2	**5**
$E_{22}(1,4)$	5	2	4	2	4	**4**
$E_{23}(1,3)$	6	1	5	1	4	**4**
$E_{24}(1,2)$	5	0	4	2	5	**2**
$E_{25}(1,1)$	8	1	5	3	6	**2**
$V_{1}(5,5)$	0	7	3	5	2	**1**
$V_{2}(5,4)$	1	6	2	4	3	**1**
$V_{3}(5,3)$	2	5	1	3	4	**3**
$V_{4}(5,2)$	3	4	2	4	5	**3**
$V_{5}(5,1)$	4	5	3	5	6	**3**
$V_{6}(4,5)$	1	6	2	4	1	**1**
$V_{7}(4,4)$	2	5	1	3	2	**3**
$V_{8}(4,3)$	3	4	0	2	4	**3**
$V_{9}(4,2)$	3	3	1	3	4	**3**
$V_{10}(4,1)$	5	4	2	4	5	**3**
$V_{11}(3,5)$	2	5	3	3	0	**5**
$V_{12}(3,4)$	3	4	2	2	1	**5**
$V_{13}(3,3)$	4	3	1	1	2	**3**
$V_{14}(3,2)$	5	2	3	2	3	**3**

$V_{15}(3,1)$	6	3	3	3	4	**2**
$V_{16}(2,5)$	3	4	3	2	1	**5**
$V_{17}(2,4)$	4	3	3	1	2	**4**
$V_{18}(2,3)$	5	2	2	0	3	**4**
$V_{19}(2,2)$	6	1	3	1	4	**2**
$V_{20}(2,1)$	7	2	4	2	5	**2**
$V_{21}(1,5)$	4	5	5	3	2	**5**
$V_{22}(1,4)$	5	2	4	2	4	**4**
$V_{23}(1,3)$	6	1	5	1	4	**1**
$V_{24}(1,2)$	5	0	4	2	5	**2**
$V_{25}(1,1)$	8	1	5	3	6	**2**
$T_{1}(5,5)$	0	7	3	5	2	**1**
$T_{2}(5,4)$	1	6	2	4	3	**1**
$T_{3}(5,3)$	2	5	1	3	4	**4**
$T_{4}(5,2)$	3	4	2	4	5	**4**
$T_{5}(5,1)$	4	5	3	5	6	**4**
$T_{6}(4,5)$	1	6	2	4	1	**5**
$T_{7}(4,4)$	2	5	1	3	2	**3**
$T_{8}(4,3)$	3	4	0	2	4	**3**
$T_{9}(4,2)$	3	3	1	3	4	**3**
$T_{10}(4,1)$	5	4	2	4	5	**3**
$T_{11}(3,5)$	2	5	3	3	0	**5**
$T_{12}(3,4)$	3	4	2	2	1	**5**
$T_{13}(3,3)$	4	3	1	1	2	**3**
$T_{14}(3,2)$	5	2	3	2	3	**3**
$T_{15}(3,1)$	6	3	3	3	4	**2**
$T_{16}(2,5)$	3	4	3	2	1	**5**
$T_{17}(2,4)$	4	3	3	1	2	**4**
$T_{18}(2,3)$	5	2	2	0	3	**4**
$T_{19}(2,2)$	6	1	3	1	4	**2**
$T_{20}(2,1)$	7	2	4	2	5	**2**
$T_{21}(1,5)$	4	5	5	3	2	**5**
$T_{22}(1,4)$	5	2	4	2	4	**4**
$T_{23}(1,3)$	6	1	5	1	4	**1**
$T_{24}(1,2)$	5	0	4	2	5	**2**
$T_{25}(1,1)$	8	1	5	3	6	**1**

In other to be relatively accurate in determining the centroid of a given data, the calculation is iterated or repeated until, the centroid in which a point belongs remains unchanged. The iteration process brings together datapoints with similar attributes in this case, datapoints with high value representing the most desirable learning preferences. That is datpoints remains within same cluster like it was in the previous cluster setting, and to achieve this, a new centroid or mean is determined, which will be used to calculate points distance to its new mean and the cluster in which it belongs to.

In other to determine the new mean (centroid) of a cluster for a datapoint, the following steps are performed.

i. Count the number of points that belongs to a cluster
ii. Find the sum of x and y separately
iii. Find the average of x and y separately.

The result will form the new mean for the cluster.

C1 contains 13 points;

X = 60/13 = 4.6

Y=61/13 = 4.7

New mean for C1 = 4.6,4.7

This is done for the clusters and the new mean for each cluster is then used to calculate the new mean for each datapoint. The number of datapoint iteration is not definite neither is it infinite. However, ones datapoints cluster remains unchanged or relatively remain same after iteration, it is assumed that some level of clustering accuracy has been attained.

Using the java code snippet 1;

DataKM.java, importing the java.util.* packages.

The DataCluster.java class contains void methods such as getCentroid(), setCentroid(), setPoints();getPoints(), etc to retrieve data from the array points. See part of the code

```
import java.util.*;
public class DataCluster {
  public List points;
        public DataPoint centroid;
        public int id;
```

```
            public DataCluster(int id) {
                        this.id = id;
                        this.points = new ArrayList();
                        this.centroid = null;
            }
            public List getPoints() {
                        return points;
            }
            public void addPoint(DataPoint point) {
                        points.add(point);
            }
```

Code snippet 2

The DataKM.java class implements the clustering of the datapoints. Each datapoint is explicitly place into a cluster after its coordinates to the mean has been determined.

```
private List getCentroids() {
            List centroids = new ArrayList(NUM_CLUSTERS);
            for(DataCluster cluster : clusters) {
                        DataPoint aux = cluster.getCentroid();
                        DataPoint point = new DataPoint(aux.getX(),aux.getY());
                        centroids.add(point);
            }
            return centroids;
}

private void assignCluster() {
    double max = Double.MAX_VALUE;
    double min = max;
    int cluster = 0;
    double distance = 0.0;

    for(Point point : points) {
        min = max;
      for(int i = 0; i &lt; NUM_CLUSTERS; i++) {
        Cluster c = clusters.get(i);
         distance = Point.distance(point, c.getCentroid());
         if(distance &lt; min){
            min = distance;
            cluster = i;
         }
      }
```

6. DISCUSSION OF RESULTS

The final iterated table 3 is below showing each value after 4 iterative sessions

Table 3: final iterated results

POINTS	C1 4.6,4.7	C2 1.7,2	C3 4.2,2.4	C4 2,3.7	C5 2.2,5	NEW CLUSTER
A_{1}(5,5)	0.7	6.3	3.4	4.3	2.8	**1**
A_{2}(5,4)	0.9	5.3	2.4	3.3	3.8	**1**
A_{3}(5,3)	1.0	4.3	1.4	3.7	2.8	**3**
A_{4}(5,2)	2.9	3.3	1.2	4.7	4.8	**3**
A_{5}(5,1)	3.9	4.3	2.2	5.7	6.8	**3**
A_{6}(4,5)	1.1	5.3	2.8	3.7	1.8	**1**
A_{7}(4,4)	1.3	4.3	1.8	2.3	2.8	**1**
A_{8}(4,3)	2.3	3.3	0.8	2.7	3.8	**3**
A_{9}(4,2)	3.3	2.3	0.6	3.7	4.8	**3**
A_{10}(4,1)	4.3	3.3	1.6	4.7	5.8	**3**
A_{11}(3,5)	2.1	4.3	3.8	2.3	0.8	**5**
A_{12}(3,4)	2.3	3.3	2.8	1.3	1.8	**4**
A_{13}(3,3)	3.3	2.3	1.8	1.7	2.8	**3**
A_{14}(3,2)	4.3	1.3	1.6	2.7	3.8	**2**
A_{15}(3,1)	5.3	2.3	2.6	3.7	4.8	**2**
A_{16}(2,5)	3.1	3.3	4.8	1.3	0.2	**5**
A_{17}(2,4)	3.3	2.3	3.8	0.3	1.2	**4**
A_{18}(2,3)	4.3	1.3	2.8	0.7	2.2	**4**
A_{19}(2,2)	5.3	0.3	2.6	1.7	3.2	**2**
A_{20}(2,1)	6.3	1.3	3.6	2.7	4.2	**2**
A_{21}(1,5)	4.1	3.7	5.8	2.3	1.2	**5**
A_{22}(1,4)	4.3	2.7	4.8	1.3	2.2	**4**
A_{23}(1,3)	5.3	1.7	3.8	1.7	3.2	**2**
A_{24}(1,2)	6.3	0.7	3.6	2.7	4.2	**2**
A_{25}(1,1)	7.3	1.7	4.6	3.7	5.2	**2**

D_1(5,5)	0.7	6.3	3.4	4.3	2.8	**1**
D_2(5,4)	0.9	5.3	2.4	3.3	3.8	**1**
D_3(5,3)	1.0	4.3	1.4	3.7	2.8	**1**
D_4(5,2)	2.9	3.3	1.2	4.7	4.8	**3**
D_5(5,1)	3.9	4.3	2.2	5.7	6.8	**3**
D_6(4,5)	1.1	5.3	2.8	3.7	1.8	**5**
D_7(4,4)	1.3	4.3	1.8	2.3	2.8	**1**
D_8(4,3)	2.3	3.3	0.8	2.7	3.8	**3**
D_9(4,2)	3.3	2.3	0.6	3.7	4.8	**3**
D_{10}(4,1)	4.3	3.3	1.6	4.7	5.8	**3**
D_{11}(3,5)	2.1	4.3	3.8	2.3	0.8	**5**
D_{12}(3,4)	2.3	3.3	2.8	1.3	1.8	**5**
D_{13}(3,3)	3.3	2.3	1.8	1.7	2.8	**4**
D_{14}(3,2)	4.3	1.3	1.6	2.7	3.8	**3**
D_{15}(3,1)	5.3	2.3	2.6	3.7	4.8	**3**
D_{16}(2,5)	3.1	3.3	4.8	1.3	0.2	**5**
D_{17}(2,4)	3.3	2.3	3.8	0.3	1.2	**4**
D_{18}(2,3)	4.3	1.3	2.8	0.7	2.2	**4**
D_{19}(2,2)	5.3	0.3	2.6	1.7	3.2	**2**
D_{20}(2,1)	6.3	1.3	3.6	2.7	4.2	**2**
D_{21}(1,5)	4.1	3.7	5.8	2.3	1.2	**5**
D_{22}(1,4)	4.3	2.7	4.8	1.3	2.2	**4**
D_{23}(1,3)	5.3	1.7	3.8	1.7	3.2	**2**
D_{24}(1,2)	6.3	0.7	3.6	2.7	4.2	**2**
D_{25}(1,1)	7.3	1.7	4.6	3.7	5.2	**2**
E_1(5,5)	0.7	6.3	3.4	4.3	2.8	**1**
E_2(5,4)	0.9	5.3	2.4	3.3	3.8	**1**
E_3(5,3)	1.0	4.3	1.4	3.7	2.8	**1**
E_4(5,2)	2.9	3.3	1.2	4.7	4.8	**3**
E_5(5,1)	3.9	4.3	2.2	5.7	6.8	**3**
E_6(4,5)	1.1	5.3	2.8	3.7	1.8	**1**
E_7(4,4)	1.3	4.3	1.8	2.3	2.8	**1**
E_8(4,3)	2.3	3.3	0.8	2.7	3.8	**3**
E_9(4,2)	3.3	2.3	0.6	3.7	4.8	**3**
E_{10}(4,1)	4.3	3.3	1.6	4.7	5.8	**3**
E_{11}(3,5)	2.1	4.3	3.8	2.3	0.8	**5**
E_{12}(3,4)	2.3	3.3	2.8	1.3	1.8	**4**
E_{13}(3,3)	3.3	2.3	1.8	1.7	2.8	**4**
E_{14}(3,2)	4.3	1.3	1.6	2.7	3.8	**2**
E_{15}(3,1)	5.3	2.3	2.6	3.7	4.8	**2**
E_{16}(2,5)	3.1	3.3	4.8	1.3	0.2	**5**
E_{17}(2,4)	3.3	2.3	3.8	0.3	1.2	**4**
E_{18}(2,3)	4.3	1.3	2.8	0.7	2.2	**4**
E_{19}(2,2)	5.3	0.3	2.6	1.7	3.2	**2**
E_{20}(2,1)	6.3	1.3	3.6	2.7	4.2	**2**
E_{21}(1,5)	4.1	3.7	5.8	2.3	1.2	**5**
E_{22}(1,4)	4.3	2.7	4.8	1.3	2.2	**4**
E_{23}(1,3)	5.3	1.7	3.8	1.7	3.2	**2**
E_{24}(1,2)	6.3	0.7	3.6	2.7	4.2	**2**
E_{25}(1,1)	7.3	1.7	4.6	3.7	5.2	**2**
V_1(5,5)	0.7	6.3	3.4	4.3	2.8	**1**
V_2(5,4)	0.9	5.3	2.4	3.3	3.8	**1**
V_3(5,3)	1.0	4.3	1.4	3.7	2.8	**1**
V_4(5,2)	2.9	3.3	1.2	4.7	4.8	**3**

$V_5(5,1)$	3.9	4.3	2.2	5.7	6.8	**3**
$V_6(4,5)$	1.1	5.3	2.8	3.7	1.8	**1**
$V_7(4,4)$	1.3	4.3	1.8	2.3	2.8	**1**
$V_8(4,3)$	2.3	3.3	0.8	2.7	3.8	**3**
$V_9(4,2)$	3.3	2.3	0.6	3.7	4.8	**3**
$V_{10}(4,1)$	4.3	3.3	1.6	4.7	5.8	**3**
$V_{11}(3,5)$	2.1	4.3	3.8	2.3	0.8	**5**
$V_{12}(3,4)$	2.3	3.3	2.8	1.3	1.8	**4**
$V_{13}(3,3)$	3.3	2.3	1.8	1.7	2.8	**4**
$V_{14}(3,2)$	4.3	1.3	1.6	2.7	3.8	**2**
$V_{15}(3,1)$	5.3	2.3	2.6	3.7	4.8	**2**
$V_{16}(2,5)$	3.1	3.3	4.8	1.3	0.2	**5**
$V_{17}(2,4)$	3.3	2.3	3.8	0.3	1.2	**4**
$V_{18}(2,3)$	4.3	1.3	2.8	0.7	2.2	**4**
$V_{19}(2,2)$	5.3	0.3	2.6	1.7	3.2	**2**
$V_{20}(2,1)$	6.3	1.3	3.6	2.7	4.2	**2**
$V_{21}(1,5)$	4.1	3.7	5.8	2.3	1.2	**5**
$V_{22}(1,4)$	4.3	2.7	4.8	1.3	2.2	**4**
$V_{23}(1,3)$	5.3	1.7	3.8	1.7	3.2	**2**
$V_{24}(1,2)$	6.3	0.7	3.6	2.7	4.2	**2**
$V_{25}(1,1)$	7.3	1.7	4.6	3.7	5.2	**2**
$T_1(5,5)$	0.7	6.3	3.4	4.3	2.8	**1**
$T_2(5,4)$	0.9	5.3	2.4	3.3	3.8	**1**
$T_3(5,3)$	1.0	4.3	1.4	3.7	2.8	**1**
$T_4(5,2)$	2.9	3.3	1.2	4.7	4.8	**3**
$T_5(5,1)$	3.9	4.3	2.2	5.7	6.8	**3**
$T_6(4,5)$	1.1	5.3	2.8	3.7	1.8	**1**
$T_7(4,4)$	1.3	4.3	1.8	2.3	2.8	**1**
$T_8(4,3)$	2.3	3.3	0.8	2.7	3.8	**3**
$T_9(4,2)$	3.3	2.3	0.6	3.7	4.8	**3**
$T_{10}(4,1)$	4.3	3.3	1.6	4.7	5.8	**3**
$T_{11}(3,5)$	2.1	4.3	3.8	2.3	0.8	**5**
$T_{12}(3,4)$	2.3	3.3	2.8	1.3	1.8	**4**
$T_{13}(3,3)$	3.3	2.3	1.8	1.7	2.8	**4**
$T_{14}(3,2)$	4.3	1.3	1.6	2.7	3.8	**2**
$T_{15}(3,1)$	5.3	2.3	2.6	3.7	4.8	**2**
$T_{16}(2,5)$	3.1	3.3	4.8	1.3	0.2	**5**
$T_{17}(2,4)$	3.3	2.3	3.8	0.3	1.2	**4**
$T_{18}(2,3)$	4.3	1.3	2.8	0.7	2.2	**4**
$T_{19}(2,2)$	5.3	0.3	2.6	1.7	3.2	**2**
$T_{20}(2,1)$	6.3	1.3	3.6	2.7	4.2	**2**
$T_{21}(1,5)$	4.1	3.7	5.8	2.3	1.2	**5**
$T_{22}(1,4)$	4.3	2.7	4.8	1.3	2.2	**4**
$T_{23}(1,3)$	5.3	1.7	3.8	1.7	3.2	**2**
$T_{24}(1,2)$	6.3	0.7	3.6	2.7	4.2	**2**
$T_{25}(1,1)$	7.3	1.7	4.6	3.7	5.2	**2**

From table 3, the following centroids were derived;

C1 = 4.5,4.5
C2 = 1.8,1.6
C3 = 4.4,2.1
C4 = 2.0,3.5
C5 = 2.1,4.8

The result can further be tested using other forms of algorithm. However, for this research, it was adequate enough and further verification was not needed because it successfully, created the clusters. Using the rectilinear took into consideration only absolute values and ignored non-absolute values, which further reduced the complexity of the clustering process. In case where there were duplicity of values within same cluster, the first-in, first-out principle was applied, therefore, ignoring the second entry.

7. CONCLUSION AND RECOMMENDATION

The work has successfully shown means of generating unsupervised synthetic dataset for demonstration purposes using Java Programming Language, and how the rectilinear clustering technique of k-means clustering can be used for creating clusters based on attribute (learning pedagogies) categorization. The work has also created the premise for further data clustering technique and manipulation

REFERENCES

1. Ahmed, A. (2018). Teaching and learning vocabulary: insights from learning styles and learning theories. *Journal of Child & Adolescent Behaviour*, *6*(1), 1-4. https://www.researchgate.net/publication/323550447
2. Aktas,I., & Bilgin, I. (2014). The effect of the 4mat learning model on achievement and motivation of 7th grade students on the subject of particulate nature of matter and an examination of students opinions on the model. *Research in Science and Technological Education. 33*(1), 1-21. https://www.academia.edu/10347458
3. Alian, M., & Al-Akharas, M. (2010). Adalearn: an adaptive e-learning environment. *ISWSA '10:Proceedings of the 1st International Conference on Intelligent Semantic Web-Services and Applications*, *21*, 1–7. https://doi.org/10.1145/1874590.1874611
4. Ambaselker, D., & Bagwan, A. (2016). Adaptive k-means clustering for association rule mining from gene expression data. *International Journal of Engineering and Technical Research*, *5*(3), 25-28. https://www.erpublication.org/published_paper/IJETR042111.pdf
5. Chen, C., Wang, C., Wang, Y., & Wang, P. (2017). Fuzzy logic controller design for intelligent robots. *Mathematical Problems in Engineering*, *27*, 1-12. https://www.researchgate.net/publication/320073463
6. Dharshinni, N.P., Azmi, F., Fawaz, I., Husein, A.M., & Siregar, S.D. (2019). Analysis of accuracy k-means and apriori algorithms for patient data clusters. *Journal of Physics Conference Series*, *12*(24). https://www.mdpi.com/2071-1050/12/24/10367
7. Dunn, T., & Kennedy, M (2019). Technology enhanced learning in higher education: motivations engagement and academic achievement. *Computer and Education*, *137*(104).
8. Dziuban, C., Moskal P., Johnson, C., & Evans, D. (2017). Adaptive learning: a tale of two context. *Elsevier*, *14*(1) , 26-55. https://scholarworks.umb.edu/ciee/vol4/iss1/3/
9. Ennouamani, S. (2017). An overview of adaptive e-learning systems. *2017 Eight International Conference on Intelligent Computing and Information Systems (ICICIS) Egypt*. https://www.researchgate.net/publication/322586183
10. Hawk, T.F., & Shah, A.J. (2007). Using learning style instruments to enhance student learning. *Journal of Innovation Education*, *5*(1), 1-19. https://doi.org/10.1111/j.1540-4609.2007.00125.x
11. Hermanwan, H., Wardani, R., Julian, C., Darmawati, A., & Yarmatov, M. (2018). Adaptive mobile learning in the nearby wisdom app. *2018 International Seminar on Intelligent Technology and Its Applications (ISITIA) Indonesia*, 221-225. https://www.researchgate.net/publication/333067941
12. Huang, Q., Yang, D., Jiang, L., Zhang, H., Liu, H., Kotani, K., (2017). An improved k-means algorithm based on association rules. *International Journal of Computer Theory and Engineering* *6*(2), 146-149. http://www.ijcte.org/papers/853-IT143.pdf.
13. Irfan, E., David, M., Itans, C., & Wa, J. (2016). Effect of using 4mat method on academic achievement and attitudes toward engineering economy for undergraduate students. *International Journal of Vocational and Technical Education*, *8*(1), 1-11. https://www.researchgate.net/publication/296057736
14. Kolb, A., & Kolb, D. (2014). Eight important things to know about the experiential learning cycle. *Australian Educational Leader*, *40*(3), 8-14. https://www.researchgate.net/profile/David-Kolb-2/publication/
15. Lau, W., & Yuen, A. (2010). Promoting conceptual change of learning sorting algorithm through the diagnosis of mental models: the effects of gender and learning styles. *Computer & Education*, *54*(1), 275-288. https://www.sciencedirect.com/science/article/abs/pii/
16. Liu, G., Huang, S., & Du, Y. (2014). An improved k-means algorithm based on association rules. *International Journal of Computer Theory and Engineering*, *6*(2), 146-149. http://www.ijcte.org/papers/853-IT143.pdf
17. Machado, M., Moreira, T., Gomes, L., Caldeira, A., & Santos, D. (2016). A fuzzy logic application in virtual education. *Procedia computer science*, *19*, 19-26. https://cyberleninka.org/article/n/676534/viewer
18. Msigwa, O. (2022). Data science machine learning (part 08): k-means clustering in plain mql5. https://www.mql5.com/en/articles/11615
19. Richter, O., & Latchem, C. (2018). Exploring four decades of research in computer and education.

Computer & Education, 122, 136-141. https://www.researchgate.net/publication/

20. Sabine, G. (2007). In-depth analysis of the felder-silverman learning style dimensions. Journal of *Research on Technology in Education*, *40*(1), 79-93.https://doi.org/10.1080/15391523.2007.10782498
21. Sangvigit, P., Mungsing, S., & Theeraroungchaisri, A. (2012). Correlation of honey & mumford learning styles and online machine preference. *International Journal for Computer Technology and Applications*, *3*(3), 1-8. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.442.4639&rep=rep1&type=pdf
22. Svinicki, Manilla D, and Dixon, Nancy M.,(2010). The kolb model modified for classroom activities. *College Teaching*, *35*(4), 141-146. https://www.researchgate.net/publication/254338739
23. Velmurugan, T. (2014). Performance based analysis between k-means and fuzzy c-means clustering algorithms for connection oriented telecommunication data. *Applied Soft Computing*, *19*, 134-146. https://www.sciencedirect.com/science/article/abs/pii/S1568494614000805
24. Wambsganss, T., Niklaus, C., Cetto, M., Sollner, M., Handschuh, S., & Leimeister, J. (2020). AL:an adaptive learning support system for argumentation skills. *CHI Conference on Human Factors in Computing Systems Hawaii*, 1-14, https://www.researchgate.net/publication/338951985
25. Wihanto, W., & Survani, E. (2020). Comparison of clustering algorithms: k-means and fuzzy c-means for segmetation retinal blood vessels. *Journal of Academy of Medical Sciences*, *28*(1), 42-47. https://www.researchgate.net/publication/339440824
26. Xie, W. (2019). Trends and development in technology enhanced adaptive/personalized learning: a systematic review of journal publications from 2007 – 2017. *Computer & Education*, *140*. https://doi.org/10.1016/j.compedu.2019.103599
27. Xiuchange, H. (2014). An improved k-means clustering algorithm. *2011 IEEE 3rd International conference on communication software and networks China*. 6-14. https://10.1109/ICCSN.2011.6014384
28. Zynel, C., & Yildiz, F. (2018). Comparison of k-means and fuzzy c-means algorithms on different cluster structures behavior. *Journal of Agricultural Informatics*, *6*(3), 13-23. https://www.researchgate.net/publication/282861550