ICIC Express Letters
Volume **17**, Number **1**, January **2023**

ICIC International ©2023 ISSN 1881-803X
pp. 49–59

MODIFICATION OF K-MEANS AND K-MODE ALGORITHMS TO ENHANCE THE PERFORMANCE OF CLUSTERING STUDENT LEARNING STYLES IN THE LEARNING MANAGEMENT SYSTEM

Emy Setyaningsih[1,*], Nurul Hidayat[2], Uning Lestari[3] and Anindita Septiarini[4]

[1]Department of Computer Systems Engineering
[3]Department of Informatics
Institut Sains & Teknologi AKPRIND
Kalisahak Street No. 28, Yogyakarta 55222, Indonesia
uning@akprind.ac.id
*Corresponding author: emysetyaningsih@akprind.ac.id

[2]Department of Informatics
Jenderal Soedirman University
Prof. DR. HR Boenyamin Street No. 708, Purwokerto 53122, Indonesia
nurul@unsoed.ac.id

[4]Department of Informatics
Mulawarman University
Kuaro Street, Gn. Kelua, Samarinda 75119, Indonesia
anindita@unmul.ac.id

Received January 2022; accepted April 2022

Abstract. *During the Corona Virus Disease (COVID-19) pandemic, many access to learning used the e-learning system through the Learning Management System (LMS) platform. One of the weaknesses of the learning process through e-learning is that it cannot detect student learning styles based on actual behavior patterns during online learning. Most of the methods used to study automatic detection techniques use classification methods. One of the weaknesses of the classification method is the determination of class labels, so a learning style detection model was developed using the concept of clustering before classification to produce class labels with a high level of validation. This study focuses on increasing the validity of the clustering method by comparing the performance of the modified K-Means and K-Mode algorithms. The proposed modification of the two algorithms is carried out at the initial centroid determination stage. The performance of the two modified algorithms was carried out by measuring the validation values of the Davies-Bouldin Index (DBI) and Silhouette Index (SI) using log file data from 88 students taking computer programming courses. The validation results of the DBI and SI values indicate that the proposed model has better performance when implemented in the K-Mode algorithm than the K-Means algorithm.*
Keywords: Davies-Bouldin index, Felder Silverman learning style model, K-Means, K-Mode, Silhouette index

1. **Introduction.** Information and Communication Technology (ICT), which is increasingly developing, offers great potential to overcome the problems of access to learning due to the COVID-19 pandemic in higher education. Access to knowledge used during the COVID-19 pandemic includes the e-learning process through the Learning Management System (LMS). LMS is a software application or web-based technology used to plan, implement, and assess learning.

The learning process through e-learning has a weakness, namely the inability to personalize learning [1-3]. Learning styles describe attitudes and behaviors that determine

DOI: 10.24507/icicel.17.01.49

students' preferred way of learning. E-learning does not consider the needs and characteristics of each student's learning style due to student socio-cultural differences [4]. Therefore, to improve student learning efficiency, it is necessary to consider learning techniques in the learning process through e-learning.

The learning style models most often used in online learning systems include the Gregorc' Mind Styles Model, Riding Cognitive Style, Myer-Briggs Type Indicator (MBTI), Kolb's Experiential Learning Theory, Honey and Mumford's Model, and Felder-Silverman Learning Style Model (FSLSM) [5]. Currently, the FSLSM learning style model is the most widely used in the education system because it has good internal consistency, reliability, and validity [6-8]. FSLSM distinguishes student learning style models based on student behavior patterns when using e-learning into four different dimensions (Active/Reflective, Sensitive/Intuitive, Visual/Verbal, Sequential/Global) [8, 9].

The learning style model of students can be detected using two approaches, namely, automatically and statically [9]. The static approach is a learning style detection approach carried out using a questionnaire, but this approach is less accurate. The second approach is automated, based on actual behavioral patterns during online learning. The development of learning platforms such as LMS is possible to detect student learning styles automatically [10]. Students who study using e-learning system activities will be recorded in a log file in the LMS. The automatic detection process is much more accurate because student activities are directly recorded without students realizing it and do not require a particular time.

Most methods used to detect learning styles automatically use the data-driven approach (data-driven). This method builds a classification model using sample data and imitates the Index Learning Style (ILS) instrument. This model uses a classification model for data mining [6, 11-17]. However, the data mining method using the classification concept has not provided a high level of accuracy. According to [1], increasing the accuracy of the automatic detection model can be done by determining class labels with a high level of validation using the concept of clustering before classification. The clustering algorithms that are often used are K-Means and K-Mode. The advantages of the two algorithms are simple, easy to adapt, and the speed of the process to produce data groupings [18]. The K-Mode algorithm, a development of the K-Means algorithm, has advantages in grouping data with many dimensions, and can effectively deal with categorical data. The weakness of both methods is the determination of the initial centroid, which is done randomly [19]. The resulting grouping becomes less optimal if the randomly selected data set for initialization is not good [18]. There are two ways to improve the weaknesses of the two algorithms, namely 1) using better initialization, 2) repeating the clustering algorithm several times with different initial centroids [18].

Based on the above background, so that the accuracy of the learning style detection model is more optimal, an appropriate clustering algorithm is needed to obtain a class label with a high level of validation. Therefore, the main contribution of this paper is to improve the performance of the K-Means and K-Mode algorithms by proposing an initial centroid determination model to improve class label validation. The proposed model measures the performance of cluster grouping results using the Davies-Bouldin Index (DBI) and Silhouette Index (SI) matrices.

This paper is structured as follows: Section 2 describes the proposed model, followed by Section 3 containing the results of the analysis and discussion of the proposed model, and finally, Section 4, the conclusion of this paper.

2. **Proposed Method.** This study uses datasets derived from computer programming courses held by the Institut Sains & Teknologi AKPRIND, Yogyakarta, with a total of 88 participants. The research process includes three main steps: pre-processing, clustering of learning style models using the K-Means algorithm, and a modified K-Mode algorithm.

The performance of the two algorithms is measured based on their validation values using DBI and SI.

2.1. **Pre-processing.** This step aims to obtain features that correlate with the type of FSLSM learning style. This stage analyzes the log file data based on the four dimensions of the FSLSM model. The observation process is carried out to classify data from 36 activities from the data log file. Logfile data will automatically be generated when students use the LMS system. The system on the LMS will record all student activities while using the LMS, such as accessing course profiles, forums, quizzes/practices/assignments/exams, sending assignments/practices/quizzes, and frequency of accessing subject matter.

The results of pre-processing of 36 activities from the data log file produce ten features consisting of features for the processing dimension of 3, the perception dimension of 2, the input dimension of 2, and the understanding dimension of 3, as shown in Table 1.

This study uses clustering $K = 16$ according to the grouping of the FSLSM learning style model as shown in Table 2.

Table 1. Pre-processing learning style detection features

FSLSM dimension	Feature	Description of student behavior
Processing	Forum_Online	F1-Actively present in the forum $\geq 50\%$
		F2-Actively present in the forum $< 50\%$
		F3-Never attended and active in the forum
	Chat	C1-Very frequently used
		C2-Sometimes
		C3-Never
	Participation/Attandance	P1-Present $\geq 75\%$
		P2-75% $>$ Present $\geq 50\%$
		P3-Present $< 50\%$
Perception	Assesment_Result	A1-Average assessment value ≥ 70
		A2-70 $>$ Average assessment value ≥ 50
		A3-Average assessment value < 50
	Assesment_submition	S1-Doing all tasks/practice/quiz
		S2-Doing partial tasks/practice/quiz
		S3-Never do assignments/practice/quiz
Input	Text input	T1-Text-based learning object $\geq 75\%$
		T2-75% $>$ Text-based learning object $\geq 25\%$
		T3-Not using text-based learning objects
	Video input	V1-Video-based learning object $\geq 75\%$
		V2-75% $>$ Video-based learning object $\geq 25\%$
		V3-Not using video-based learning objects
Understanding	Activity LMS	L1-Accessing $\geq 75\%$
		L2-75% $>$ Accessing $\geq 50\%$
		L3-Accessing $< 50\%$
	Exam	E1-Exam score ≥ 70
		E2-70 $>$ Exam score ≥ 50
		E3-Exam score < 50
	Summary	R1-Summary read $\geq 70\%$
		R2-70% $>$ Summary read $\geq 25\%$
		R3-Summary not read

Table 2. Combination of FSLSM learning styles

Cluster	Learning style	Cluster	Learning style	Cluster	Learning style	Cluster	Learning style
1	A, S, Vi, Seq	5	A, I, Vi, Seq	9	R, S, Vi, Seq	13	R, I, Vi, Seq
2	A, S, Vi, G	6	A, I, Vi, G	10	R, S, Vi, G	14	R, I, Vi, G
3	A, S, Ve, Seq	7	A, I, Ve, Seq	11	R, S, Ve, Seq	15	R, I, Ve, Seq
4	A, S, Ve, G	8	A, I, Vi, G	12	R, S, Ve, G	16	R, I, Vi, G

2.2. **The proposed initial centroid determination model.** The initial centroid determination algorithm proposed in this paper is to develop the proposed model [1]. The model proposed to modify the K-Means and K-Mode algorithms in this study is a process to determine the dataset used as the initial centroid. The initial centroid determination algorithm proposed for K-Means and K-Mode in this study is shown in Table 3, where $N = 88$, $f = 10$, and $K = 16$ are variables that state the number of data sets, the number of features, and the number of clusters used to test the proposed method in this study.

2.3. **Proposed modification of the K-Means algorithm.** The modified K-Means algorithm process is as follows.

1) Initialization process: Initialization processes the number of groupings (K) in the study used $K = 16$ according to the FSLSM learning style model group as shown in Table 3. The value of the Function Objective (FO) is determined with a value large enough so that the iteration process is not carried out only once so that the results of the clustering can be optimal. The FO value in this study is filled with 1000.
2) Determination of initial centroid: The initial centroid is determined by selecting K data sets that have similarities with 16 combinations of FSLSM learning style models. The rules are used to determine the initial centroid using the algorithm shown in Table 3.
3) Calculating the distance of each data set to the centroid: Calculating the length of each information assigned to the centroid is done using the Euclidean distance formula as shown in Equation (1).

$$d_{ik} = \sqrt{\sum_{j=1}^{m} (c_{ij} - c_{kj})^2} \tag{1}$$

where d_{ik} is the data set distance at i, c_{ij} is the data set at i, and c_{kj} is the data centroid at j.
4) Calculating the new centroid value: Calculating the new centroid value is done by finding the average value of the data set member of the cluster, using Equation (2).

$$c_{kj} = \frac{\sum_{i=1}^{p} x_{ij}}{p} \tag{2}$$

where $x_{ij} \in$ cluster at k and p is the number of cluster members at k.
5) Calculating the new centroid: Calculating the new centroid aims to group the data into clusters with the shortest distance using the new centroid generated in step 4). The next step is to calculate the value of the Function Objective (FO), which aims to get the value of the distance to the new centroid, closest distance, and cluster from each dataset. The calculation of the FO value is obtained from the most relative distance from the new centroid to each data originating from the cluster results from the previous iteration.

Table 3. Proposed modification of the K-Means and K-Mode algorithms at the initial centroid determination process

```
Input: Data = array[1...N; 1...f + 1] {Preprocessed data}
Output: CA = array[1...K; 1...f] {Initial Centroid}
Proses:
C ← [ ]
for i = 1 to N
    {Determining the dimensions of learning styles from N data sets}
    P1 ← data(i, 2) + data(i, 3) + data(i, 4)
    if (P1/3) ≥ 2 then k1 = 1
    else k1 = 0
    endif
    P2 ← data(i, 5) + data(i, 6)
    if (P2/2) <= 2 then k2 = 1
    else k2 = 0
    endif
    if data(i, 8) <= data(i, 7) then k3 = 1
    else k3 = 0
    endif
    U ← data(i, 9) + data(i, 10) + data(i, 11)
    if (U/3) <= 2 then k4 = 1
    else k4 = 0
    endif;
    {Define cluster label according to learning style}
    Class = [k1  k2  k3  k4]
    if Class = [1111] then C = [C 1]
    elseif Class = [1110] then C = [C 2]
    elseif Class = [1101] then C = [C 3]
    elseif Class = [1100] then C = [C 4]
    elseif Class = [1011] then C = [C 5]
    elseif Class = [1010] then C = [C 6]
    elseif Class = [1001] then C = [C 7]
    elseif Class = [1000] then C = [C 8]
    elseif Class = [0111] then C = [C 9]
    elseif Class = [0110] then C = [C 10]
    elseif Class = [0101] then C = [C 11]
    elseif Class = [0100] then C = [C 12]
    elseif Class = [0011] then C = [C 13]
    elseif Class = [0010] then C = [C 14]
    elseif Class = [0001] then C = [C 15]
    else C = [C 16]
    endif;
endfor
{Determine the dataset to be the initial centroid (CA)}
for j = 1 : K
    CA(j) ← find(C = j)
endfor
```

6) Determine the convergence condition of the iteration process: Iteration in this algorithm is declared convergent if the Delta value is less than the threshold value (T) and there is no cluster displacement. If it has not converged, then repeat steps 3) to 6).

2.4. **Proposed modification of the K-Mode algorithm.** The modified K-Mode algorithm process is as follows.

1) Initial centroid determination: Initial centroid determination in the modified K-Mode algorithm uses the same algorithm as the initial centroid determination process in the K-Means algorithm, as shown in Table 3.
2) Calculate the distance between each object and the cluster mode: Assign objects to the cluster whose center has the closest space to the thing using Equation (3).

$$d(X,Y) = \sum_{j=1}^{r} \epsilon(X_j, Y_j) \tag{3}$$

where $d(X,Y)$ is the data distance X to Y, X_j is the feature value from X at j, Y_j is the feature value from Y at j, r is the number of features, and $\epsilon(X_j, Y_j)$ is the matched value using Equation (4).

$$\epsilon(X_j, Y_j) = \begin{cases} 0 & \text{if } X_j = Y_j \\ 1 & \text{if } X_j \neq Y_j \end{cases} \tag{4}$$

3) Updating the centroid center: Updating the centroid center of each cluster is determined from the category value that frequently occurs in each cluster.
4) Determine the convergence condition of the iteration process: Iteration in this algorithm is declared convergent if it meets the requirements, namely (a) the data in the cluster does not move or (b) the position of the centroid center does not change.

2.5. **Testing the proposed modification of the K-Means and K-Mode algorithms.** The validation test of the modified K-Means algorithm and the modified K-Mode algorithm used two measures, namely DBI and SI.

1) DBI: DBI is used to measure the validation of the entire cluster in the data set. The DBI value is obtained from Equation (5) [1,20].

$$DBI = \frac{1}{C} \sum_{i=1}^{C} \max(r_{i,j}), \text{ provided that } i \neq j \tag{5}$$

where C is the number of clusters and $r_{i,j}$ is formulated by Equation (6).

$$r_{i,j} = \frac{W_i + W_j}{B_{i,j}} \tag{6}$$

Sum of square Within-cluster (W) as a cohesion metric in cluster at i is formulated by Equation (7).

$$W_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \tag{7}$$

where m_i is the number of data in the cluster at i, while c_i is the cluster centroid at i. While the metric for the separation between the two clusters used the formula sum of square Between-clusters (B) by measuring the distance between the centroids c_i and c_j using Equation (8).

$$B_{i,j} = d(c_i, c_j) \tag{8}$$

2) SI: SI is used to measure the validation of data, single cluster, or whole-cluster [20,21]. The SI value of a cluster is obtained by calculating the SI average of all data that joins the cluster, as in Equation (9)

$$SI_j = \frac{1}{m_j} \sum_{i=1}^{m_j} SI_i^j \tag{9}$$

The value of $SI_i^j = \frac{b_i^j - a_i^j}{\max\left\{a_i^j, b_i^j\right\}}$ where a_i is the average distance of the data at i to all other data in one cluster. At the same time, b_i is obtained by calculating the average length of the data at i to all data from other clusters not in the same cluster with the data i; then, the smallest is taken. The global SI value is obtained by calculating the average SI value of all clusters as in Equation (10)

$$SI = \frac{1}{C} \sum_{j=1}^{C} SI_j \tag{10}$$

where C is the number of clusters.

3. **Result and Discussion.** This paper's learning style clustering model is implemented using the Matlab R2013a application.

3.1. **Results of clustering learning styles using modified K-Means and K-Mode algorithms.** The modified K-Means algorithm clustering process in the test data set will converge after five iterations. In comparison, the modified K-Mode algorithm clustering process results on the test data set will converge after seven iterations.

The clustering results using the modified K-Means and K-Mode algorithm can be seen in Figure 1.

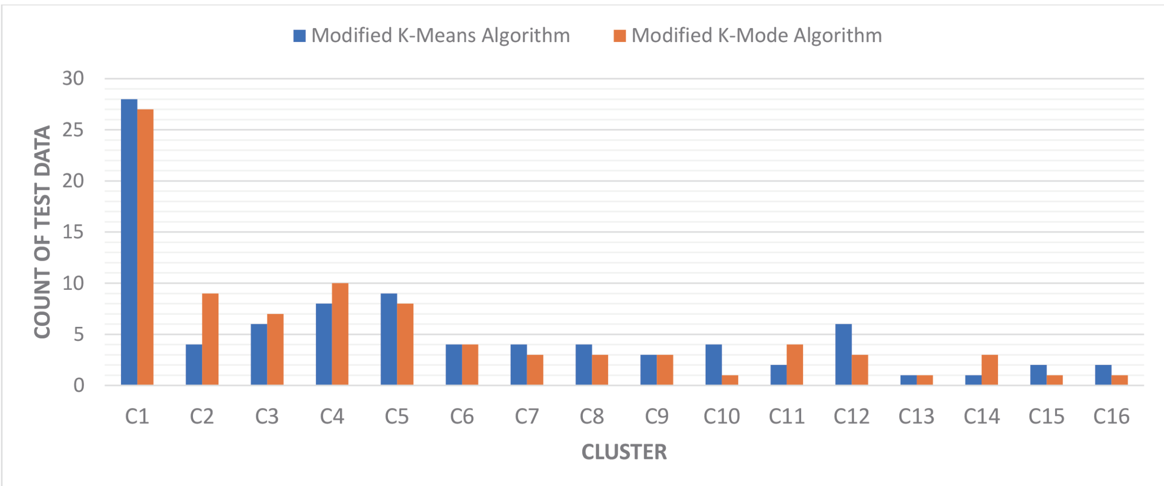

Figure 1. The results of clustering using modified K-Means and K-Mode algorithms

Based on Figure 1, the clustering of student learning styles using the two algorithms, especially in the learning style group in cluster 1, namely Active, Sensing, Visual, Sequential (A, S, Vi, Seq). This cluster shows that computer programming students prefer to learn by doing activities such as discussions and tests, choosing to discover by seeing or hearing directly. Students will also remember what they saw, such as pictures, videos, or animations, and have learning styles sorted by topic.

3.2. **Validation of the proposed model modification of the K-Means and K-Mode algorithm.** The DBI values for the modified K-Means algorithm and the original K-Means algorithm for each test data set of 10 trials are shown in Figure 2. The DBI values for the modified K-Mode algorithm and the original K-Mode algorithm for each test data set of 10 trials are shown in Figure 3.

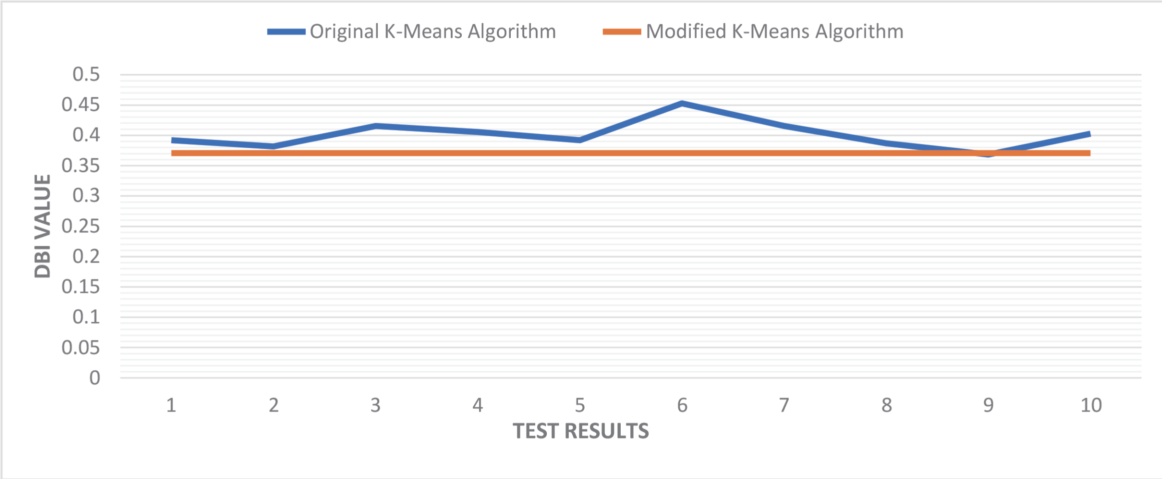

Figure 2. Comparison of the results of the modified K-Means algorithm validation test with the original K-Means using the DBI value

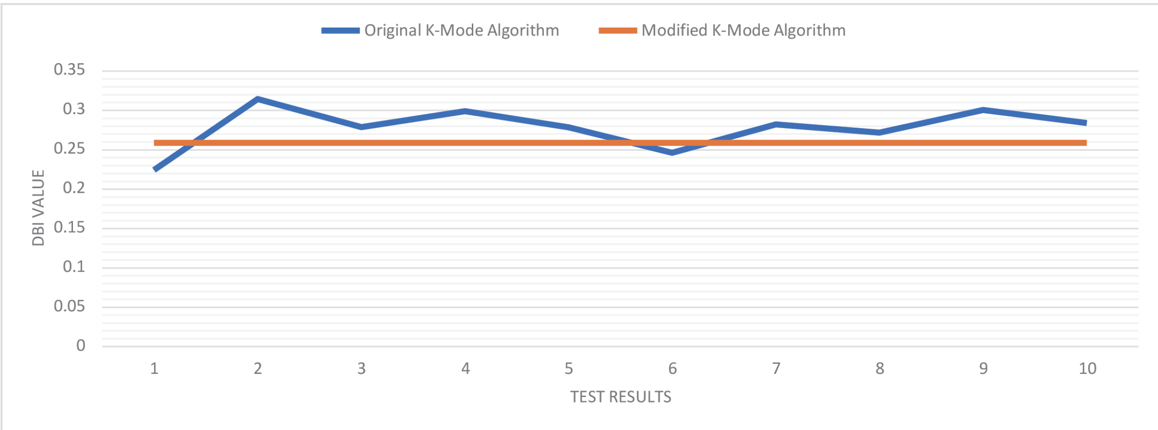

Figure 3. Comparison of the validation testing results of the modified K-Mode algorithm with the original K-Mode using the DBI value

Figures 2 and 3 show the DBI values that are always different in each experiment for the original K-Means and K-Mode algorithms. DBI values that are always different indicate that each time the clustering process is carried out, it will produce cluster labels that are not the same. The difference in DBI value is due to the initial centroid value generated from the randomization process in each clustering process. The results of testing the proposed modification of the K-Means and K-Mode algorithms for each test by measuring the DBI value in each algorithm can be seen in Figures 2 and 3. The proposed development of the two algorithms results in no cluster displacement in the dataset for each clustering process so that the cluster data formed is more optimal. Value validation for each cluster can be done using the SI. A comparison of the results of SI calculations using the K-Means and K-Mode algorithms that have been modified for each cluster can be seen in Figure 4.

The SI values shown in Figure 4 show that 75% of the data set is in the appropriate cluster. In addition, Figure 4 also shows that two clusters have an SI value close to one, which means the data set has been grouped correctly in that cluster. Furthermore, the average value of the validation of the entire cluster for the two proposed clustering algorithms is shown in Figure 5. Figure 5 shows that the proposed modification of the two clustering algorithms is good in grouping test data sets, as evidenced by the DBI validation value, which is close to 0 [20], and the SI value is close to 1 [21, 22]. Based on the validation values of the DBI and SI values shown in Figure 5, the modified K-Mode

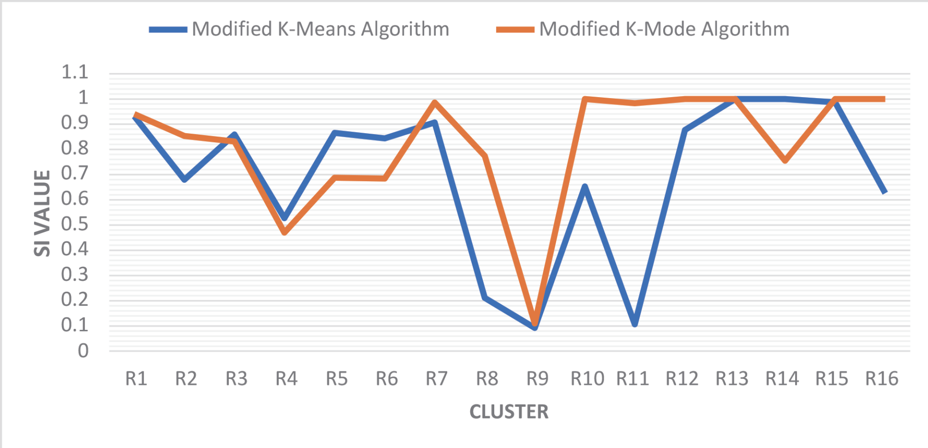

FIGURE 4. Comparison of SI values for each cluster between modified K-Means and K-Mode algorithms

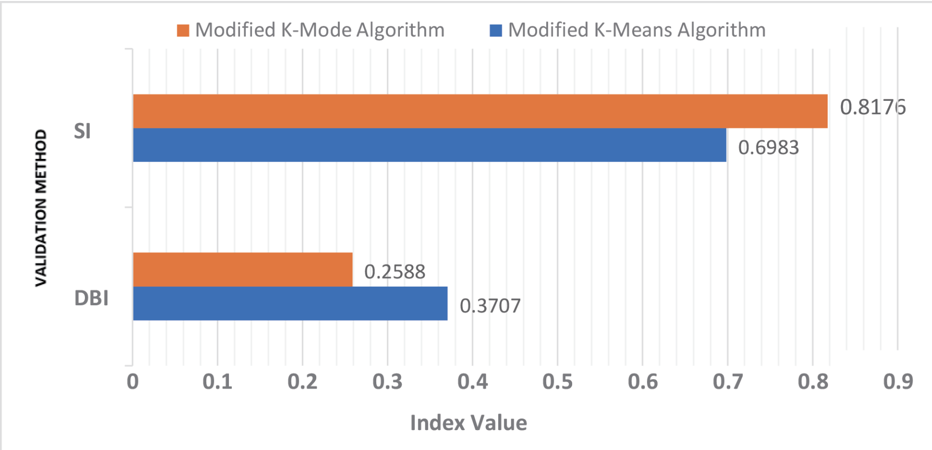

FIGURE 5. Comparison of DBI and SI values between the modified K-Means and K-Mode algorithm

algorithm can improve the performance of the modified K-Means algorithm. The DBI value of the modified K-Mode algorithm is 0.2588, which is closer to the value 0 than the modified K-Means algorithm. The SI value of the modified K-Mode algorithm is 0.8176, which is also more relative to the value of 1 than the modified K-Means algorithm.

4. **Conclusion.** This paper proposed a modification of the K-Means and K-Mode algorithms to enhance the performance clustering of the learning style method in the LMS. The proposed modification of the two algorithms is carried out at the initial centroid determination stage. The performance of the two modified algorithms is carried out by measuring the DBI and SI validation values using $N = 88$, $f = 10$, and $K = 16$ with the variable N, f, and K denoting the number of data sets, the number of features, and the number of clusters. Based on the DBI and SI values generated from the test, the proposed modification of the two algorithms is good enough for clustering data sets. The comparison of DBI and SI values for the two algorithms shows that the modified K-Mode algorithm performs better than the modified K-Means algorithm. The DBI value of the modified K-Mode algorithm is close to 0, which is 0.2588 compared to the modified K-Means algorithm. The SI value of the modified K-Mode algorithm is also better, namely 0.8176, which is close to the value of 1 compared to the modified K-Means algorithm. The test results using the modified two algorithms show that the grouping of students' learning styles in computer programming courses is dominated by active, sensing, visual,

sequential learning styles. This learning style means that students learn by participating in discussions and tests. Students choose to discover by seeing or hearing in person. Students will remember what they saw, such as pictures, videos, or animations, and have learning styles sorted by topic. As part of future work, the proposed model can be applied to detecting learning styles by combining classification methods using data mining methods. The aim is to determine the combination of the modified K-Mode algorithm with the appropriate classification method to optimize the automatic detection of student learning styles.

REFERENCES

[1] N. Hidayat, R. Wardoyo, A. SN and H. D. Surjono, Enhanced performance of the automatic learning style detection model using a combination of modified K-Means algorithm and Naïve Bayesian, *International Journal of Advanced Computer Science and Applications*, vol.11, no.3, pp.638-648, 2020.

[2] M. P. P. Liyanage, K. S. L. Gunawardena and M. Hirakawa, Using learning styles to enhance learning management systems, *International Journal on Advances in ICT for Emerging Regions (ICTer)*, vol.7, no.2, DOI: 10.4038/icter.v7i2.7153, 2014.

[3] L. K. P. Suryapranata, G. P. Kusuma, Y. Heryadi and W. Budiharto, Adaptive gamification framework for improving motivation and achievement of students in learning process, *ICIC Express Letters*, vol.15, no.4, pp.327-334, 2021.

[4] S. Graf, T.-C. Liu and Kinshuk, Analysis of learners' navigational behaviour and their learning styles in an online course, *Journal of Computer Assisted Learning*, vol.26, no.2, pp.116-131, 2010.

[5] O. Zine, A. Derouich and A. Talbi, A comparative study of the most influential learning styles used in adaptive educational environments, *International Journal of Advanced Computer Science and Applications*, vol.10, no.11, pp.520-528, 2019.

[6] M. P. P. Liyanage, K. L. Gunawardena and M. Hirakawa, Detecting learning styles in learning management systems using data mining, *Journal of Information Processing*, vol.24, no.4, pp.740-749, 2016.

[7] M. P. Pitigala Liyanage, K. S. L. Gunawardena and M. Hirakawa, A framework for adaptive learning management systems using learning styles, *2013 International Conference on Advances in ICT for Emerging Regions (ICTer)*, pp.261-265, DOI: 10.1109/ICTer.2013.6761188, 2013.

[8] T. Sheeba and R. Krishnan, Prediction of student learning style using modified decision tree algorithm in e-learning system, *Proc. of the 2018 International Conference on Data Science and Information Technology (DSIT'18)*, New York, NY, USA, pp.85-90, 2018.

[9] J. Feldman, A. Monteserin and A. Amandi, Automatic detection of learning styles: State of the art, *Artificial Intelligence Review*, vol.44, no.2, pp.157-186, 2015.

[10] E. Gomede, R. Miranda de Barros and L. de Souza Mendes, Use of deep multi-target prediction to identify learning styles, *Applied Sciences*, vol.10, no.5, DOI: 10.3390/app10051756, 2020.

[11] E. S. Amir, M. Sumadyo, D. I. Sensuse, Y. G. Sucahyo and H. B. Santoso, Automatic detection of learning styles in learning management system by using literature-based method and support vector machine, *2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, pp.141-144, 2016.

[12] L. Mahnane and M. Hafidi, Automatic detection of learning styles based on dynamic Bayesian network in adaptive e-learning system, *International Journal of Innovation and Learning*, vol.20, no.3, DOI: 10.1504/IJIL.2016.079067, 2016.

[13] O. El Aissaoui, Y. El Madani El Alami, L. Oughdir and Y. El Allioui, Integrating web usage mining for an automatic learner profile detection: A learning styles-based approach, *2018 International Conference on Intelligent Systems and Computer Vision (ISCV)*, pp.1-6, 2018.

[14] I. Karagiannis and M. Satratzemi, An adaptive mechanism for Moodle based on automatic detection of learning styles, *Education and Information Technologies*, vol.23, no.3, pp.1331-1357, 2018.

[15] O. E. Aissaoui, Y. E. A. El Madani, L. Oughdir and Y. E. Allioui, Combining supervised and unsupervised machine learning algorithms to predict the learners' learning styles, *Procedia Computer Science*, vol.148, pp.87-96, 2019.

[16] A. Kika, L. Leka, S. Maxhelaku and A. Ktona, Using data mining techniques on Moodle data for classification of student's learning styles, *Proc. of the 47th International Academic Conference*, Prague, 2019.

[17] H. H. Aly, A novel approach for harmonic tidal currents constitutions forecasting using hybrid intelligent models based on clustering methodologies, *Renewable Energy*, vol.147, pp.1554-1564, 2020.

[18] P. Fränti and S. Sieranoja, How much can K-Means be improved by using better initialization and repeats?, *Pattern Recognition*, vol.93, no.4, pp.95-112, 2019.
[19] Y. Dai, G. Yuan, Z. Yang and B. Wang, K-Modes clustering algorithm based on weighted overlap distance and its application in intrusion detection, *Scientific Programming*, vol.2021, pp.1-9, 2021.
[20] S. V. Petrovic, A comparison between the silhouette index and the Davies-Bouldin index in labelling IDS clusters, *The 11th Nordic Workshop on Secure IT-Systems*, pp.53-64, 2006.
[21] N. L. R. Amalia, A. A. Supianto, N. Y. Setiawan, V. Zilvan, A. R. Yuliani and A. Ramdan, Student academic mark clustering analysis and usability scoring on dashboard development using K-Means algorithm and system usability scale, *Jurnal Ilmu Komputer dan Informasi*, vol.14, no.2, pp.137-143, 2021.
[22] J. M. Luna-Romera, M. del Mar Martínez-Ballesteros, J. García-Gutiérrez and J. C. Riquelme-Santos, An approach to silhouette and dunn clustering indices applied to big data in spark, in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, O. Luaces, J. A. Gámez, E. Barrenechea et al. (eds.), Cham, Springer, 2016.