UTILIZING CLUSTERING ALGORITHMS TO PROVIDE VARK LEARNING STYLE RECOMMENDATIONS

MS Hasibuan
Faculty Computer Science
Institute Informatics and Business Darmajaya
Bandar Lampung, Indonesia
msaid@darmajaya.ac.id

RZ Abdul Aziz
Faculty Computer Science
Institute Informatics and Business Darmajaya
Bandar Lampung, Indonesia
rz_aziz@darmajaya.ac.id

Sigit A
Faculty Computer Science
Institute Informatics and Business Darmajaya
Bandar Lampung, Indonesia
sigit.andriyanto.2021211028@mail.darmajaya.ac.id

***Abstract*— The term "learning style" refers to the approach preferred by individuals to acquiring knowledge. Fleming and Mills introduced an interesting method called VARK (Visual, Audio, Read/Write, Kinesthetic) to observe learning styles. VARK has been widely used in numerous studies to understand how students prefer to receive information. This research aims to identify clusters of student learning styles in order to enhance the quality of education. E-learning enables students to learn in a manner similar to traditional classrooms, even without the presence of teachers or educators. The study collected data from 138 students attending 16 meetings on campus. Based on the findings and discussions, the following conclusions can be drawn: the Split Validation accuracy test indicates that the K-Means algorithm has a Davies-Bouldin Index (DBI) value of 0.475, the X-Means algorithm has a DBI value of 0.478, and the K-Medoids algorithm has a DBI value of 0.699. The K-Means algorithm yields the best result, with a DBI value of 0.475, which is considered good as it is the smallest value obtained. A smaller DBI value or a value closer to zero indicates a more accurate cluster. Based on the visual results, students predominantly exhibit a preference for the VISUAL and AUDIO learning styles, as recommended by the K-Means algorithm.**

Keywords—Learning Style, Clustering, K-Means

I. INTRODUCTION

E-learning is defined as any form of electronic support of the learning process, which is oriented to the character of the user and aims to build a system that refers to knowledge contraction references, experience, and user or student training [1]. An example of E-learning must have instructional content and provide a learning experience through electronic technology[2] [3][4].

Learning styles are defined as different types of student signals. These differences are sometimes referred to as "lifestyles" and sometimes interpreted as personality types, so that learning styles are referred to as part of an individual's intellectual style [5] [6][7]. Also, learning styles often mean differences in academic ability, skills and interests[8][9] [10]. Learning style is also interpreted as an approach or method that is preferred by individuals to learn and acquire knowledge [11]. VARK (Visual, Aural, Read/Write, Kinesthetic) is an interesting approach to observing learning styles proposed by Fleming and Mills [12]. It is used extensively in many current studies [13][14][15][16][17]. The concept of this approach is to consider the type or style of student interest in receiving the information provided.

The VARK questionnaire was developed by Lincoln University of New Zealand in 1998. The questionnaire is based on three principles, namely: (a) Everyone can learn academic problems; otherwise, everyone has their own learning style, (b) student motivation increases when students' learning styles are adopted from learning content, and (c) the concept of education is learned through the use of different senses and different perceptions.

According to this perspective, individuals acquire knowledge about their surroundings through four sensory modes: visual, auditory, reading/writing, and kinesthetic. Fleming described these four sensory models in the VARK framework. Students who are visual (V) prefer receiving information in the form of images, which aids their learning through the use of maps, graphs, charts, and other visual materials. Aural (A) students learn more effectively through auditory means. Reading/writing (R) students assimilate new information by reading and writing. Kinesthetic (K) students rely on physical experiences to learn, engaging multiple sensory modalities such as touch, hearing, smell, taste, and sight.

Some individuals may have more than one modality learning style, they can be unimodal or multimodal[18] [19]. In addition, according to Felder and Silverman, learning styles (part of the conventional style) describe students' preferences about how some subjects are presented, how to work with that subject matter, and how to internalize (obtain, process, and store) information[20][21]. According to Willingham, students may have diverse preferences when it comes to learning. Thus, knowing a student's learning styles can help in finding the most suitable way to improve the learning process. There are several studies showing that learning styles enable adaptive e-learning systems to enhance student learning[22][23][24].

This research is to determine the clustering of student learning styles, in order to improve the quality of education. E-learning helps students to learn as they would in class, even though there are no lecturers or educators to help and accompany them. This study took a sample of students from 16 (sixteen) meetings with a total of 138 (one hundred and thirty-eight) students.

II. METHOD

This study carried out the following stages to complete its research and suit the objectives.

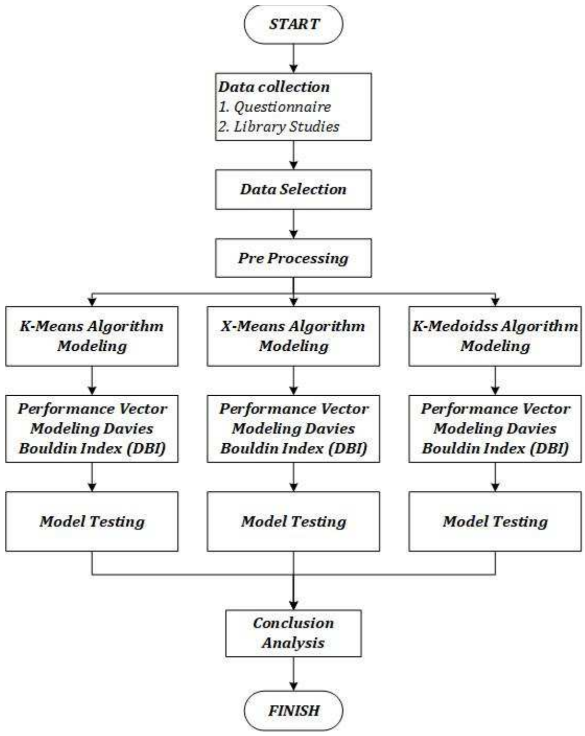

Figure 1. Methodology Research

The picture of the research stages above can be explained as follows: the raw data is reduced first and entered into Microsoft Excel to become training data. A total of 276 data were collected. Using the pre-processing method, the data is processed to eliminate duplicates and missing data, which are already available in RapidMiner Studio software. Furthermore, the testing data will be tested with the training data that has been obtained. Then the data is tested using the K-Means, X-Means and K-Medoids algorithms . The data that has been tested is then compared to which algorithm performance results are the best in determining the VARK learning style cluster:

Data Preparation

The data used is secondary data, i.e., the results of activities in e-learning and the results of the questionnaire, which were obtained from campus. In these data it can be seen that the VARK learning style cluster is formed, which consists of 8 predictor attributes and 1 target attribute. The attributes that become parameters are shown in Table 1.

Table 1. Data Attribute

No.	Attribute
1	NIM
2	Name
3	Class
4	Program Study
5	*Visual*
6	*Audio*
7	*Reading*
8	*Kinesthetic*

Data Selection

The data selection, or data selection stage, is the attribute selection stage of the data to be analyzed, because not all data contained in the raw data will be used. Thus, we obtain some attributes that will be used. The results of the data selection table can be seen in Table 2.

Table 2. Data Selection

Attribute	Indicator	Usage Details
NIM	√	Used
Name	X	Not Used
Class	X	Not Used
Program Study	√	Used
Visual	√	Used
Audio	√	Used
Reading	√	Used
Kinesthetic	√	Used
Nilai	√	Used

The table above describes the attributes that will be used in this study and those that won't. The "V" indicator indicates that the variable is used, while the "X" indicator indicates that the attribute is eliminated or not used at the stage of determining the criteria. Some of these attributes are eliminated because they do not affect the results of the assessment process.

Data Cleaning

The pre-processing stage is the stage of cleaning missing value data, namely inconsistent or empty data, and also selecting variables for the data that one wants to use in the data mining process. Clustering the learning styles of campus students will be a Visual, Audio, Reading, Kinesthetic cluster or a combination of VARK.

Data Integration and Transformation

To improve the accuracy and efficiency of the algorithm, the data in this study are grouped. After obtaining the data, the next step is to group the attributes because they affect the cluster results. This data clustering is done with the help of the Ms.Excel tool. Student data that has been classified is taken 7 attributes.

Modelling

a. Clustering Model

The resulting data source is a dataset with a total of 134 data records containing student data. Then the data will be used as the initial basis for the data mining process, so that a significance output and an analysis of the results are obtained which state the most dominant learning style with the VARK approach. The data set is divided into two parts, namely data testing which is data used to measure the extent to which the cluster has succeeded in clustering correctly. The training data (training set) is the data used by the clustering algorithm (K-Means, X-Means, K-Medoids) to form a cluster model[25][26]. This model is a knowledge representation that will be used to classify new data classes that have never existed.

This data set table is pure data that has not been processed in such a way as training data. If we look carefully at this data set table, we will find several unused attributes and duplicate records. Not only is it divided into training data which is used as the basis for modeling the K-Means, X-Means, K-Medoids algorithms, but from the large number of existing datasets, it is necessary to divide the test data (testing data) which is used to test the extent to which the clustering algorithm succeeds in grouping correctly.

b. Algorithm Model

At this stage, the selection of appropriate modeling techniques is carried out. In this study, using the K-Means, X-Means and K-Medoids models or algorithms in RapidMiner, which is a solution for analyzing data mining, grouping. RapidMiner is a standalone software for data analysis and as a data mining engine integrated into its own product. After going through the preparatory stage of data processing, this study used a dataset of 138 detection records and 138 non-detection records.

Implementation and Method Testing

Implementation of clustering testing in this study using RapidMiner software. RapidMiner is software programming that works in data processing. RapidMiner is a programming language that has a wide range of capabilities using data mining principles and algorithms. Besides that, RapidMiner can extract patterns from very large final data by combining statistical methods, artificial intelligence and databases. RapidMiner is devoted to the use of data mining. The models provided are quite complete, including K-Means, X-Means and other K-Medoids models. The algorithm validation test used is the Davies-Bouldin Index (DBI), while the accuracy of the algorithm is tested using Split Validation.

III. Results and Discussion

The cluster accuracy test is used to evaluate the results of the cluster analysis quantitatively, so that the optimal group is produced. The accuracy of forming the number of clusters can be validated by using the cluster accuracy method to find out the best input in cluster formation. The cluster accuracy test method used is the Split Validation method, as shown in Figure 2.

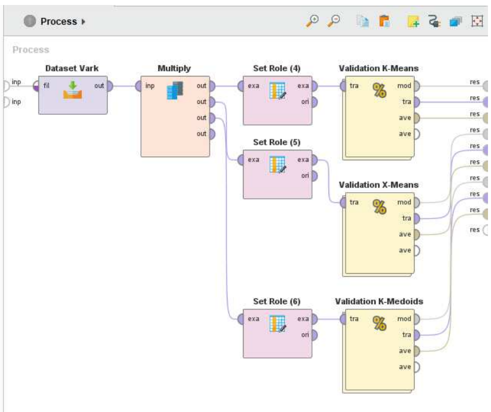

Figure 2. Test Cluster Validity Using Split Validation

1. Validation of the K-Means Data Testing Method and Data Training Using Split Validation

Modeling the validation of the accuracy of the K-Means Algorithm Cluster, using Split Validation, is shown in Figure 3.

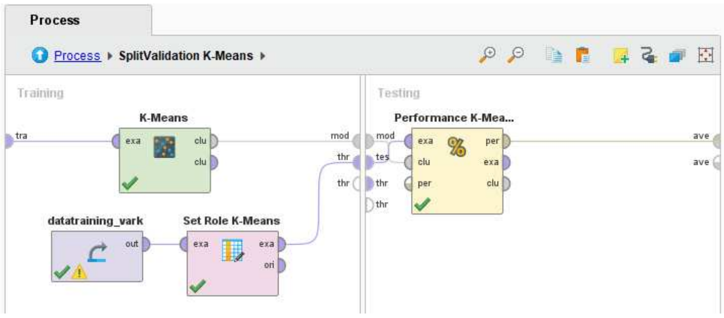

Figure 3. K-Means Algorithm Cluster Validation Model Using Split Validation

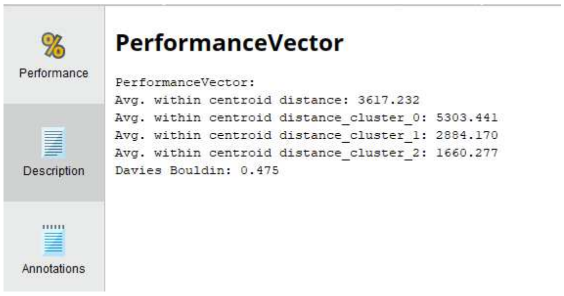

Figure 4. Performance Test Results for the Davies Bouldin Index Value on the K-Means Algorithm

The Performance Vector results above show the resulting Davies Bouldin Index value of 0.475. The resulting DBI value is higher than the DBI value in the previous experiment, namely 0.391.

2. Validation of the X-Means Data Testing Method and Data Training Using Split Validation.

Modeling the accuracy of validating the X-Means Algorithm, Cluster using Split Validation, is shown in Figure 5.

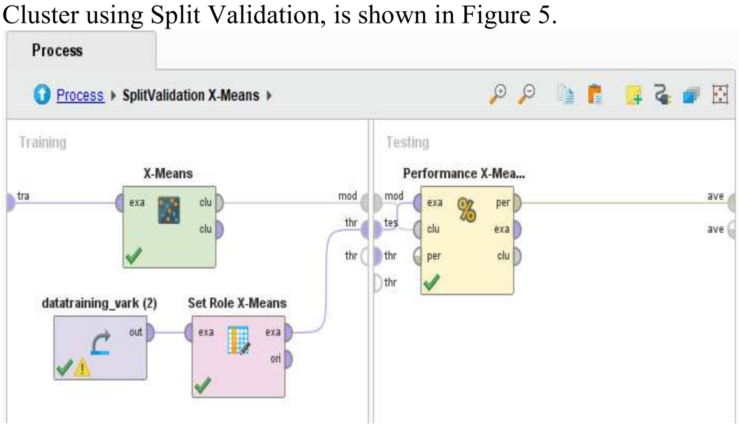

Figure 5. X-Means Algorithm Cluster Validation Model Using Split Validation

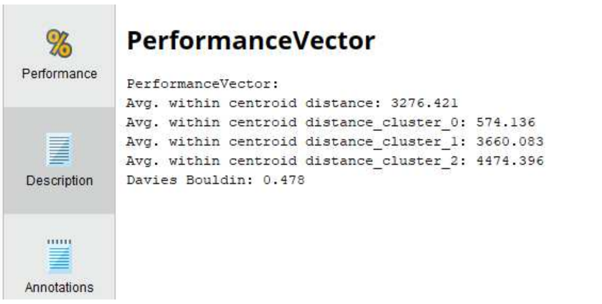

Figure 6. Performance Test Results for the Davies Bouldin Index Value on the X-Means Algorithm

The Performance Vector results above show the resulting Davies Bouldin Index value of 0.47. The resulting DBI value is higher than the DBI value in the previous experiment, namely 0.391.

3.Validation of the K-Medoids Data Testing Method and Data Training Using Split Validation

Modeling the validation of the accuracy of the K-Medoids Algorithm Cluster, using Split Validation, is shown in Figure 7.

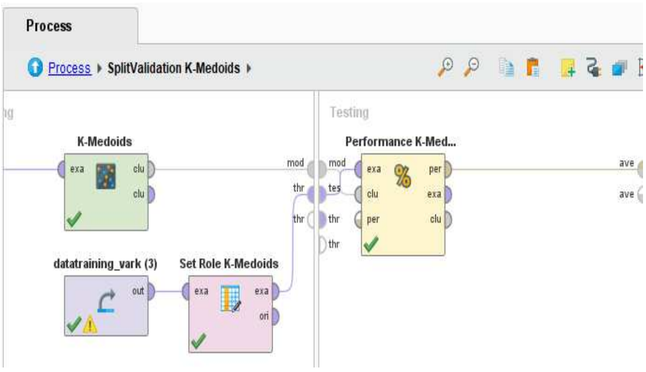

Figure 7. K-Medoids Algorithm Cluster Validation Model Using Split Validation

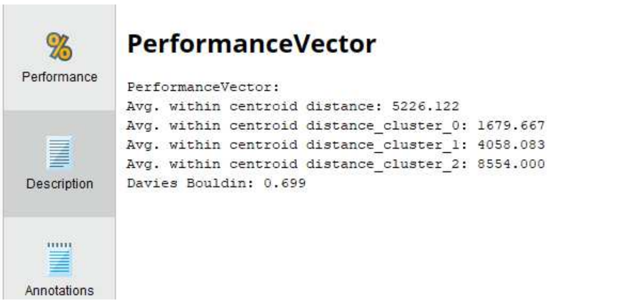

Figure 8. Performance Test Results for the Davies Bouldin Index Value on the K-Medoids Algorithm

The Performance Vector results above show the resulting Davies Bouldin Index value of 0.699. The resulting DBI value is higher than the DBI value in the previous experiment, namely 0.674.

The research that has been done shows the results of an information pattern in using the data mining process to cluster the 4th (fourth) semester students. This research produces an information pattern that is in accordance with the objectives of data mining, namely data training and data testing patterns to cluster the learning styles of students using the VARK approach of each attribute using training data and data testing to obtain new information, are dominant in using Visual, Audio, Reading, Kinesthetic learning styles or a combination of VARK. The clustering process uses the K-Means, X-Means and K-Medoids algorithms.

After the clustering process has been completed, the process of finding the level of accuracy of the three algorithms is then carried out. The process of clustering and searching for accuracy values uses the RapidMiner software. The results of testing the accuracy using the Split Validation Model that has been carried out are the K-Means algorithm, the X-Means algorithm and the K-Medoids algorithm. The test is carried out by looking at the DBI value of each algorithm.

Table 3. Davies-Bouldin Index value after Split Validation Accuracy Test

No	Algorithm	Davis-Bouldin Index
1	K-Means	0.475
2	X-Means	0.478
3	K-Medoids	0.699

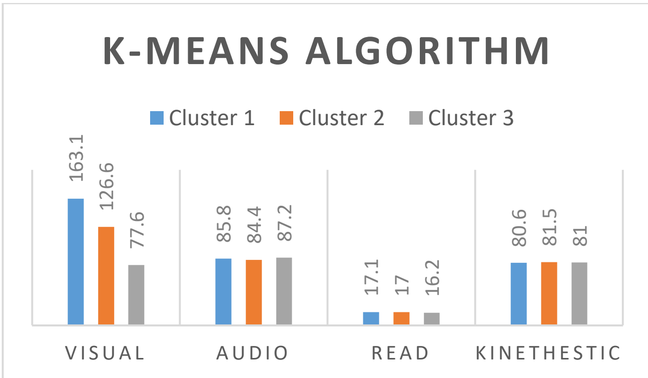

Figure 9. Comparison of Accuracy Values

The Split Validation accuracy test value produces a DBI of the K-Means algorithm is 0.475. The X-Means algorithm is 0.478, while the DBI value of the K-Medoids algorithm is 0.699. The best result based on the Davies-Bouldin Index value is found in the K-Means algorithm with a DBI value of 0.475. This value is said to be good, because this result is the smallest value obtained, and because the smaller the DBI value or the closer to zero, the more accurate the resulting cluster. Meanwhile, from the results of the visual data displayed graphically in Figure 9 the visual, audio, reading and kinesthetic learning styles recommended by the K-Means algorithm, students are more dominant using the VISUAL and AUDIO learning styles.

IV. Conclusion

Based on the results and the discussions that have been carried out, it can be concluded that the visual, audio, reading and kinesthetic learning styles recommended by the K-Means algorithm use more dominant visual and audio learning styles; the visual, audio, reading and kinesthetic learning styles recommended by the X-Means algorithm use a more dominant visual and audio learning style; the visual, audio, reading and kinesthetic learning styles recommended by the K-Medoids algorithm use more dominant visual and audio learning styles and the test value for the accuracy of Split Validation produces a DBI of the K-Means algorithm of 0.475. The X-Means algorithm is 0.478, while the DBI value of the K-Medoids algorithm is 0.699. The best result, based on the Davies-Bouldin Index value, is found in the K-Means algorithm with a DBI value of 0.475. This value is said to be good, because this result is the smallest value obtained, and because the smaller the DBI value or the closer to zero, the more accurate the resulting cluster.

ACKNOWLEDGMENTS

The Ministry of Research, Technology, and Higher Education of the Republic of Indonesia funded this study under the Fundamental Research of College Excellence program.

REFERENCES

[1] M. S. Hasibuan and L. Nugroho, "Detecting Learning Style Using Hybrid Model," *2016 IEEE Conf. e-Learning, e-Management e-Services Detect.*, pp. 107–111, 2016.

[2] X. Wei and J. Yan, "Learner profile design for personalized E-learning systems," *Proc. - 2009 Int. Conf. Comput. Intell. Softw. Eng. CiSE 2009*, pp. 1–4, 2009, doi: 10.1109/CISE.2009.5363560.

[3] M. S. Hasibuan, L. E. Nugroho, and P. I. Santosa, "Model E-learning MDP for Learning Style Detection Using Prior Knowledge," *Int. J. Eng. Technol.*, vol. 7, pp. 118–122, 2018.

[4] S. Ennouamani and Z. Mahani, "An overview of adaptive e-learning systems," in *The 8th IEEE International Conference on Intelligent Computing and Information Systems (ICICIS*, 2017, no. February 2018, doi: 10.1109/INTELCIS.2017.8260060.

[5] M. Alshammari, R. Anane, and R. J. Hendley, "The Impact of Learning Style Adaptivity in Teaching Computer Security," *ACM*, no. December, pp. 135–140, 2015, doi: 10.1145/2729094.2742614.

[6] M. İlçin, N. Tomruk, S. S. Yeşilyaprak, D. Karadibak, and S. Savcı, "The relationship between learning styles and academic performance in Turkish physiotherapy students," *Physiotherapy*, vol. 102, pp. e84–e85, 2016, doi: 10.1016/j.physio.2016.10.084.

[7] B. D. Ictenbas and H. Eryilmaz, "Determining learning styles of engineering students to improve the design of a service course," *Procedia - Soc. Behav. Sci.*, vol. 28, pp. 342–346, 2011, doi: 10.1016/j.sbspro.2011.11.065.

[8] M. A. Asiry, "Learning styles of dental students," *Saudi J. Dent. Res.*, vol. 7, no. 1, pp. 13–17, 2016, doi: 10.1016/j.sjdr.2015.02.002.

[9] T. Vasileva-Stojanovska, T. Malinovski, M. Vasileva, D. Jovevski, and V. Trajkovik, "Impact of satisfaction, personality and learning style on educational outcomes in a blended learning environment," *Learn. Individ. Differ.*, vol. 38, pp. 127–135, 2015, doi: 10.1016/j.lindif.2015.01.018.

[10] T. K. Tee *et al.*, "The Pattern of Learning Styles among Second Year Students in Business Management and Hospitality Programs at One of The Vocational College in Northern Zone," *Procedia - Soc. Behav. Sci.*, vol. 204, pp. 62–72, 2015, doi: 10.1016/j.sbspro.2015.08.110.

[11] P. Sivanandan, T. Letchumanan, M. Ramayah, and N. H. Nasrijal, "Learning Style Preferences : Influence of Cultural Background among Business Students," *Int. J. arts Commer.*, vol. 3, no. 5, pp. 87–98, 2014.

[12] N. D. Fleming, "I'm different; not dumb. Modes of presentation (VARK) in the tertiary classroom," *Res. Dev. High. Educ. Proc. Annu. Conf. High. Educ. Res. Dev. Soc. Australasi*, pp. 308–313, 1995.

[13] A. Amran, A. Desiani, and M. S. Hasibuan, "Detection learning style vark for out of school children (OSC)," *Int. Conf. Electr. Eng. Comput. Sci. Informatics*, vol. 3, 2016, doi: 10.1088/1757-899X/190/1/012031.

[14] V. Marcy, "Adult learning styles: How the VARK Learning Styles Inventory can be used to improve student learning," *Perspect. Physician Assist. Educ.*, vol. 12, no. 2, pp. 117–120, 2001, [Online]. Available: http://www.paeaonline.org/index.php?ht=action/GetDocumentAction/i/25142Smith.

[15] L. I. F. Dutsinma and P. Temdee, "VARK Learning Style Classification Using Decision Tree with Physiological Signals," *Wirel. Pers. Commun.*, pp. 1–29, 2020, doi: 10.1007/s11277-020-07196-3.

[16] S. Wright and A. Stokes, "The application of VARK learning styles in introductory level economics units," *Issues Educ. Res.*, vol. 25, no. 1, pp. 62–79, 2015.

[17] P. I. S. Muhammad Said Hasibuan1, 2 , Lukito Edi Nugroho1, "MODEL DETECTING LEARNING STYLES WITH ARTIFICIAL NEURAL NETWORK," *J. Technol. Sci. Educ.*, vol. 9, no. 1, pp. 86–96, 2019.

[18] M. A. O. Balfaqeeh, "Understanding Engineering Students' Learning Styles," *Int. J. Linguist. Commun.*, vol. 5, no. 1, pp. 44–55, 2017, doi: 10.15640/ijlc.v5n1a5.

[19] P. Kharb, P. P. Samanta, M. Jindal, and V. Singh, "The learning styles and the preferred teaching-learning strategies of first year medical students," *J. Clin. Diagnostic Res.*, vol. 7, no. 6, pp. 1089–1092, 2013, doi: 10.7860/JCDR/2013/5809.3090.

[20] S. Graf, S. R. Viola, T. Leo, and Kinshuk, "In-Depth Analysis of the Felder-Silverman Learning Style Dimensions," *J. Res. Technol. Educ.*, vol. 40, no. 1, pp. 79–93, 2007, doi: 10.1080/15391523.2007.10782498.

[21] C. Ipbuker, "Teaching and Learning Styles in Engineering Education Learning Styles and Teaching Models in Engineering Education Felder-Solomon," no. November 2008, 2014, doi: 10.1109/FIE.2008.4720326.

[22] B. Daoruang, S. Chaichomchuen, and A. Mingkhwan, "An Adaptive Learning System Based on Proportional VARK to Enhance Learning Achievement Concept," no. February, 2018.

[23] T. Larkin and D. Budny, "Learning styles in the classroom: Approaches to enhance student motivation and learning," *ITHET 2005 6th Int. Conf. Inf. Technol. Based High. Educ. Training, 2005*, vol. 2005, no. August 2005, 2005, doi: 10.1109/ITHET.2005.1560310.

[24] A. Freiberg-Hoffmann, J. B. Stover, and N. Donis, "Influence of learning strategies on learning styles: Their impact on academic achievement of college students from Buenos aires," *Probl. Educ. 21st Century*, vol. 75, no. 1, pp. 6–18, 2017.

[25] H. L. Sari, D. Suranti, and L. N. Zulita, "Implementation of k-means clustering method for electronic learning model," *J. Phys. Conf. Ser.*, vol. 930, no. 1, pp. 0–6, 2017, doi: 10.1088/1742-6596/930/1/012021.

[26] S. Samudi, S. Widodo, and H. Brawijaya, "The K-Medoids Clustering Method for Learning Applications during the COVID-19 Pandemic," *SinkrOn*, vol. 5, no. 1, p. 116, 2020, doi: 10.33395/sinkron.v5i1.10649.